USING BAYESIAN LEARNING TO CLASSIFY COLLEGE ALGEBRA STUDENTS BY UNDERSTANDING IN REAL-TIME

by

ANDREW COUSINO

B.A., Knox College, 2004
M.S., Kansas State University, 2007

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Mathematics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2013

Abstract

The goal of this work is to provide instructors with detailed information about their classes at each assignment during the term. The information is both on an individual level and at the aggregate level. We used the large number of grades, which are available online these days, along with data-mining techniques to build our models. This enabled us to profile each student so that we might individualize our approach. From these profiles, we began to investigate what can be done in order to get students to do better, or at least be less frustrated. Regardless, the interactions with our undergraduates will improve as our knowledge about them increases.

We start with a categorization of Studio College Algebra students into groups, or clusters, at some point in time during the semester. In our case, we used the grouping just after the first exam, as described by Dr. Rachel Manspeaker in her PhD. dissertation. From this we built a naive Bayesian model which extends these student clusters from one point in the semester, to a classification at every assignment, attendance score, and exam in the course. A hidden Markov model was then constructed with the transition probabilities being derived from the Bayesian model. With this HMM, we were able to compute the most likely path that students take through the various categories over the semester. We observed that a majority of students settle into a group within the first two weeks of the term.

USING BAYESIAN LEARNING TO CLASSIFY COLLEGE ALGEBRA STUDENTS BY UNDERSTANDING IN REAL-TIME

by

ANDREW COUSINO

B.A., Knox College, 2004
M.S., Kansas State University, 2007

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Mathematics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2013

Approved by:

Major Professor
Andrew G. Bennett

Abstract

The goal of this work is to provide instructors with detailed information about their classes at each assignment during the term. The information is both on an individual level and at the aggregate level. We used the large number of grades, which are available online these days, along with data-mining techniques to build our models. This enabled us to profile each student so that we might individualize our approach. From these profiles, we began to investigate what can be done in order to get students to do better, or at least be less frustrated. Regardless, the interactions with our undergraduates will improve as our knowledge about them increases.

We start with a categorization of Studio College Algebra students into groups, or clusters, at some point in time during the semester. In our case, we used the grouping just after the first exam, as described by Dr. Rachel Manspeaker in her PhD. dissertation. From this we built a naive Bayesian model which extends these student clusters from one point in the semester, to a classification at every assignment, attendance score, and exam in the course. A hidden Markov model was then constructed with the transition probabilities being derived from the Bayesian model. With this HMM, we were able to compute the most likely path that students take through the various categories over the semester. We observed that a majority of students settle into a group within the first two weeks of the term.

Table of Contents

List of Figures

List of Tables

List of Algorithms

Acknowledgments

I would like to thank Dr. Robby and Dr. William Hsu for serving on my graduate committee. Your patience and time has been greatly appreciated. Dr. Robert Burckel and Dr. Todd Cochrane have provided me with a significant amount of support and guidance. For this, I am deeply grateful. Dr. Andrew Arana showed me the joys of logic, without which I would not have had the motivation to get past a Masters degree in math.

Dr. Andrew Bennett has had an influence upon me which cannot be overestimated. Among the many things that he has taught me is how to deal with academic bureaucracy, how to be an effective professor, how to be a good teacher, and how to to be a good researcher. For the remainder of my life, Dr. Bennett will be a model against which I will compare myself and to which I will turn whenever I am unsure of how act in my academic life and in all other aspects of my life as well. It will be very difficult for me to leave Kansas State University because it will mean leaving Dr. Bennett. Although, I suspect he will be eager and delighted to see me leave.

I should also thank the professors of my almae matres Knox College and Johnson County Community College for their support in helping me get to graduate school and instilling within me a deep love of mathematics. In particular, Mike Martin, Dr. Mary Armon, and Dr. Dennis Schneider deserve individual mention for their contributions to my life.

Finally, I'd like to thank my family, Marsha Cousino, Mark Cousino, and Ron Cousino. I will never be able to pay back the monetary and emotional debt which I owe to you. I would not be able to do this without you now, and without the help you have given in the past.

Chapter 1

Introduction

The goal of this work is to provide instructors with detailed information on each of their students at every assignment during the term. The information will be on an individual level and may be aggregated to provide broader perspective. This will enable instructors to profile students so that they may tailor assignments and instruction to each student. They may begin to break free from the one-size-fits-all style of instruction without treading blindly into unfamiliar territory. With this new knowledge, the effects of customized instruction can be observed in the models described herein, and refined in order to better serve our students.

We start with a categorization of Studio College Algebra[i] students into groups, or clusters. In our case, we will use the grouping just after the first exam as described by Dr. Rachel Manspeaker in her PhD. dissertation, "Using Data Mining to Differentiate Instruction in College Algebra". Then we build a naive Bayesian model using the classifications and the course grades as a data set. The Bayesian model extends Dr. Manspeaker's classification from being at one point in the semester to classifications after every assignment, attendance score, and exam in the course. Dr. Manspeaker divided the class into five groups. This was, in effect, a snapshot of the class at the time of the first exam. The model which we build divides the class into the exact same five groups,

[i] In the United States, college algebra courses typically introduces functions, exponentials, logarithms, and basic matrix operations.

but does so at every time a student receives a grade during the term.

Next, a Hidden Markov Model (HMM) is constructed with the transition probabilities which are derived from the Bayesian model and represent the chance of moving in time from one classification group to another. With this HMM, we are able to compute the most likely path that students take through the various categories over the semester. In other words, we computed, for each student, the path through the categories with the largest probability. Every student has been assigned to a group at each point in the term, and together these assignments form the student's path between groups over the course of the semester. And this particular path is the most likely path which the student could have taken during the course. The instances when these paths stayed constant for a sufficiently long time turned into the object of study. We interpreted these long runs in the path to be times when a student had settled into a particular group, and observed that a majority of students settled into a group within the first two weeks of the term.

1.1 Motivation

Kansas State University, like many universities across the country, is focusing on student retention in an effort to better serve its students and to increase its revenue. Consequently, new ways of instruction are being encouraged. Underlying many of these ideas is the philosophy of differentiated instruction. The core mechanic is that in the same class, students are offered various methods of learning the course content. By doing this, students are able to learn according to their preference, rather than the traditional one-size-fits-all approach. Differentiated instruction can be strengthened by demanding that each student should be limited to a small subset of choices in learning, rather than being able to pick from the entire selection. But to do so requires a knowledge about everyone in the class on an individual basis. Data mining is being used to better understand students at this individual level. The data analysis informs us about which instruction methods should be applied to which students. Essentially, we try to cater to the students in a way which is effective

and results in the student staying enrolled.

Dr. Manspeaker's work set the stage for differentiated instruction in our math courses by establishing a systematic way of placing students into groups. In her work, she used grades, up to and including the first exam, to group students into clusters. She established that these groups were characterized by behaviors and attitudes about math class. With this technique, we can now classify each student just after the first exam. This tells us how each student is behaving around the time of this exam, but does not indicate behavior at any other time in the semester. This leads one to wonder about whether students changed their attitudes or behaviors.

A question of significant interest was whether a student's attitudes and behaviors could be predicted before the semester even began. This would be a great aid in the task of placing incoming first-year students into the appropriate math course. For those new students that placed into either version of College Algebra, we could recommend they take either the traditional or studio version, provided we knew about their attitudes. In order to test whether student attitudes can be predicted before the semester, Dr. Manspeaker added a survey to the mathematics placement exam, which first-year students are required to take prior to the start of the academic year. However, from the survey responses she was unable to predict the students' groups.

There are a number of explanations for this difficulty in placing students before the start of the term. But one particular interpretation stood out, namely that student attitudes and behaviors change from summer to their first month of college. Personal experience and anecdotes about students adjusting to college suggests that this interpretation is quite reasonable. It is also common for instructors to observe students change their behavior over the term. However, the research literature only covers behavioral changes across the typical four-year college experience. This is the gap which the current work is trying to fill.

1.2 Dr. Manspeaker's Research

Dr. Rachel Manspeaker wanted to develop a way to classify students in Studio College Algebra. Her focus was to identify early in the semester students who may need help. So she examined student grades up to and including the first exam. This data set admitted clustering into five groups. She then took prototypical students in each group and interviewed them, in order to identify what distinguished each cluster from the others. These interviews showed that student attitudes about math and behaviors within the class strongly placed them into specific clusters.

These groups became labeled as smart slackers (SS), over-achievers (OA), Intervention (I), employees (E), and rote memorizers (RM). The smart slacker group is aptly named, for its members are those who are intelligent and knowledgeable but lacking motivation; over-achievers are the students who work hard and do well; intervention group members work very hard and do well on assignments, but poorly on exams; employees treat college like a menial job by doing the least amount of work possible while still being compliant; and the rote memorizers get through math classes by mimicking the proper motions and resist any deeper meaning. It should be noted, however, that the names "smart slacker" and "intervention" do not appear in Dr. Manspeaker's dissertation. The reason for the name change will be given in the subsequent chapter. With the groups established for one semester, Dr. Manspeaker confirmed the result by performing the same clustering algorithms as in previous terms and showed that the individual clusters exhibited the same attitudes and behaviors as they had in previous semesters.

As mentioned earlier, Dr. Manspeaker tried to see if the behavioral group of a student could be predicted prior to arriving on campus. She added an attitude survey to the math placement exam, which is required of every incoming first-year student. She then ran her classification algorithms to group those first-year students who took Studio College Algebra their first semester. Unfortunately, the responses to the attitude survey did not significantly correlate to behavioral groups. This led to the belief that students changed their attitudes and beliefs by the first month of

class. This became the significant motivator for the present work.

1.3 Goals

This work builds a model of student behaviors and attitudes at each assignment, exam question, and attendance score for each student in Studio College Algebra. Our objectives are to see how quickly students settle into a group and whether we can identify behavioral patterns in individuals solely from the student's grades. The first goal, about how quickly students settle into a group, extends Dr. Manspeaker's work by enabling us to provide evidence in support of the claim that students change their behaviors quickly upon entering college. We know that their attitudes before entering their first year do not correlate well with their assigned group at the first exam. So if we can demonstrate that a typical student forms consistent and predictable behaviors prior to this first exam, then the proposition about student behaviors being formed early in their college career but not prior to their college career gains support. Alternatively, our analysis had the potential to demonstrate that (1) students enter higher education with consistent behaviors, (2) students do not typically exhibit stable behavior through their College Algebra course, or (3) students begin to exhibit stable behavior after the first exam. Either of these cases would have been interesting to observe as it would have weakened the claim about behaviors forming early.

The second goal, identifying patterns of behaviors within students, would help to further our work to improve instruction. This is where differentiated instruction would guide us to a particular method of teaching for each undergraduate based upon that individual's behavioral patterns. Such an ability to prescribe instruction allows teachers to experiment with their techniques. Even within the same semesters, similar students could be taught with various techniques. Their results could be meaningfully compared, provided we trust that the students' behavioral types are comparable. This would be an important advance for educational research.

1.4 Limitations

One of the important qualifiers regarding this work is that we have only produced models of student behavior over time. We have not verified empirically the validity or accuracy of these models. One way to validate these models is through student interviews. For example, in order to validate that most students settle into a group after the second week of class, we can perform interviews with a random selection of students as soon as the third week of class to see whether they have an established pattern of behavior in the class. Then once the semester is over, the models presented in this paper can be computed. In particular the most likely path each student took over the term through Dr. Manspeaker's groups can be calculated. These most likely paths can be compared to the results of the interviews to see how accurately the models predicted student behaviors. If one of the interviewed students exhibited stable behavior, and if the type of behavior demonstrated matched the predicted behavior of the model, then this observation would be evidence in favor of the model's accuracy. And if a sufficiently large majority of those interviewed had their behaviors correctly predicted by the model, then we could consider the models accurate, and say that there is no longer a lack of empirical evidence for these models.

Next, we should note the clear focus on the Studio College Algebra course at Kansas State University. We did not want to randomly sample college algebra students from various schools, as we needed a common course structure over which we could build our models. Our models vary with course, course schedule, and teaching style. These are further limitations for this work. It is safe to assume that the course and course schedules did not alter the models significantly because this particular course is always lead by the same instructor, Dr. Rekha Natarajan, who directs the course in an almost identical manner from one semester to the next. However, the differences due to the effects from the numerous teaching assistants was not taken into account, and must therefore be noted among the limitations and weaknesses of this work. While this flaw in the model tarnishes its accuracy, it does not prevent us from demonstrating its value and

potential. Rather, these criticisms of model accuracy are the motivation for future research and not the present work.

Chapter 2

Literature Survey

2.1 Dr. Rachel Manspeaker's Dissertation

In order to track Studio College Algebra students, we need to have a means of labeling or categorizing them. With this, we can now attempt to observe changes in a student's category. We chose to build off of the work our colleague, Dr. Rachel Manspeaker, who was a fellow PhD. student at the time this work began.

The work done by Dr. Rachel Manspeaker in her PhD. dissertation[5] is integral to this work, as it provides us with a means of classifying students into categories. She clustered students of the Studio College Algebra course by grades up to and including the first exam of the course. This data was clustered using standard techniques — Singular Value Decomposition (SVD), AGglomerative NESting (AGNES), and Partitioning About Medoids (PAM). The SVD analysis was used to reduce the dimensions of the data, from over thirty scores for attendance, participation, assignments, and the questions on the first exam, down to five orthogonal dimensions which were the largest in the spectral decomposition of the data. From this reduced data set, she used AGNES to find an appropriate number of clusters in the data, which was determined to be five. Then PAM was used to obtain the stable clustering of the data into five subsets. Dr. Manspeaker followed this

	Prop. of class	Avg. final grade	Avg. ACT comp.	Avg. ACT math
OA	33%	3.24	22.38	21.44
SS	13.7%	2.13	23.29	22.34
E	24.7%	2.62	22.6	21.4
RM	10.6%	1.24	21	20.6
I	17.6%	1.53	20.92	19.27

Table 2.1: Statistics for each cluster

quantitative analysis with blind interviews of prototypical students from each cluster to learn their behaviors as students and attitudes towards the class and towards mathematics in general. The following are the names given to the five clusters.

- Over-achievers (OA)
- Intervention (I)
- Employees (E)
- Rote Memorizers (RM)
- Smart Slackers (SS)

2.1.1 Profile of Over-achievers

These are the excellent students. They see mathematics as a useful subject, both in general and in their own futures. However, they do not see how the class specifically applies to their lives. This cluster has the best scores in all facets of the class. As for the first exam, such students only struggle with the nonstandard problems, i.e., problems which were not found in previous exams and not worked by the instructors in class, as well as the graphical questions.

2.1.2 Profile of Smart Slackers

Smart slackers are intelligent and well prepared, but bored and frustrated with the course. Ideally, such a person would be placed in a higher-level course such as trigonometry. On the first exam, they do well on application problems and questions involving graphs, but do poorly on the procedural and nonstandard parts. These students tend to have the second lowest scores in all aspects of the course, except on the first exam, where their performance varies from moderate to excellent. Eventually, they drop the course or end up performing poorly. In her dissertation, Dr. Manspeaker called this group "under-achievers", as she felt that her former name "smart slackers" was disparaging. However, I prefer clear, descriptive labels and have no heart .

2.1.3 Profile of Employees

Employees view the class as a menial job for which they are "paid" with a passing grade. They dislike math, and learn through memorization, not comprehension. On exams, this group does well on procedural problems which do not require independent thought or creativity. These undergraduates tend to have the second highest scores in the course for all things except attendance, in which they place third. This is because their attendance starts off strong but eventually drops. In general, if these students aren't learning, they blame the teacher.

2.1.4 Profile of Rote Memorizers

Rote memorizers rely on memorizing rather than comprehension to pull them through the course. They have negative views about math which progressively get worse after the course. They believe that talent in mathematics is an inherited trait rather than one which is developed through work and understanding. Students in this cluster had the lowest scores in all aspects of the course, with only a few exceptions, which still put them near the bottom of the class. They typically stop working after the first assignment and stop attending after the first class. These students strongly dislike

the studio version of College Algebra, and would prefer to be in the traditional version. Advisers have been advised to place those whom they presume to be rote memorizers in traditional College Algebra, as the traditional course more closely fits the rote memorizer's idea of how a math class is supposed to run.

2.1.5 Profile of Intervention Group

This type of student is the one whose great efforts in learning go unrewarded. They have excellent attendance, do well on written homework, and yet do poorly on online homework. On exams, these students have the most trouble with nonstandard questions and applied problems. Despite their poor test score, they demonstrated a better knowledge of the material in interviews. Quantitatively, this group is very similar to the rote memorizers except that this cluster has good written homework and attendance. Originally, Dr. Manspeaker called this group "Sisyphean strivers" in her thesis. I am using the name employed within our research group, rather than that from Dr. Manspeaker's dissertation. When writing her work, she decided to change the group names in order to avoid and prevent the use of the derogatory label "smart slackers". And "Sisyphean strivers" was chosen due to it having the same abbreviation as "smart slacker", thus discouraging its use for future researchers, and because these students seemed to be unable to succeed in college algebra despite how hard they tried. The name "intervention group" was used because Dr. Manspeaker focused on trying to intervene and help this group overcome their poor test performance. In short, "intervention group" is our name for Dr. Manspeaker's "Sisyphean strivers", and our "smart slackers" is Dr. Manspeaker's "under-achievers".

2.2 Bayesian Inference

With a method of categorizing students, we can now develop a means of extending these categories over the entire semester. This is accomplished by utilizing naive Bayesian models built from

student grades and classifications. In this work, we used Dr. Manspeaker's groups, but these methods are independent of the method of categorization. Any other system of grouping students can be likewise extrapolated over the entire course.

Using Dr. Manspeaker's data along with the course records, we set out to build a model of how students changed groups throughout the semester. The way we built the model was rather straightforward. In order to predict a student's behavioral group given data about their assignments, we needed to compute probabilities of the form $P(\text{IN BEHAVIORAL GROUP } c \mid \text{SCORE } s \text{ ON ASSIGNMENT } a)$. We could then compute $P(\text{IN BEHAVIORAL GROUP } c)$ once we had $P(\text{SCORE } s \text{ ON ASSIGNMENT } a)$. After grades have been recorded, we can use the proportion of students who got a score of s on assignment a to represent $P(\text{SCORE ON ASSIGNMENT } a \text{ WAS } s)$. Strictly speaking, these probabilities are not the appropriate ones. What is in fact used in the actual computations are not of the form "score on assignment 1 was 7 out of 10" but are instead "score on assignment 1 was in the 70^{th} to 80^{th} percentile". Rather than have the variable "score" be the numerical score, we chose to make it a category. This helps reduce the noise from grading idiosyncrasies as well as performance variations — a student having a bad day which ends up affecting their score on an assignment or exam. But the following description of how we created the model works whether we think of assignment scores as being numerical or categorical quantities. For brevity and simplicity, we will just treat them as discrete and categorical.

The trouble is that $P(\text{IN BEHAVIORAL GROUP } c \mid \text{SCORE ON ASSIGNMENT } a \text{ WAS } s)$ is difficult to obtain directly. For the sake of the reader, we will use the random variables S_a to be the score on assignment a and C to be the behavioral group of the student. But by using Bayes' theorem, we can get an estimate, provided we know $P(S_a \mid C)$ for each a and $P(C)$. At this point, it is important to introduce the difference between the training and testing data sets. With a few exceptions, the 2010 Fall (F10) semester was the source of the testing data, and the 2009 Fall (F09) semester was the training data. In all other instances, we will note which semester is training and which is the testing. The training data set is used to calculate the needed parameters $P(S_a \mid C)$ and $P(C)$. The

conditional probability of assignment scores given their classification in Dr. Manspeaker's groups is simply the proportion of students who got that score out of all the students from 2009 in that classification. And it would be very easy to take the probability of being in a category as the proportion of all F09 students in that category. It is important to note here that we are assuming that a student's behavioral group is constant throughout the term when computing the parameters from the training data. The student's group is the one assigned from Dr. Manspeaker's work. With the necessary parameters set, we can now apply our Bayesian inference to the testing data to generate predictions about the behavioral groups of the testing data. It is worth noting that while we knew the groups to which Dr. Manspeaker assigned the students of both the training and testing data, we do not in practice need to know anything about the groups of the testing data. We only need to know the assigned groups of the training data.

To compute predictions from the testing data, we need to allow for the groups to change from assignment to assignment. We do so by estimating the group probabilities recursively.

$$\begin{aligned} P^{(1)}(C_0) &:= \frac{1}{\#\text{ OF GROUPS}} \\ P^{(1)}(C_m) &:= P^{(1)}(C_m \mid S_m)P(S_m), \text{for each } m \\ P^{(1)}(C_{m+1} \mid S_{m+1}) &:= \frac{P(S_{m+1} \mid C = C_{m+1})P^{(1)}(C_m = C_{m+1})}{\sum P(S_{m+1} \mid C = C_{m+1})P^{(1)}(C_m = C_{m+1})}, \text{for each } m \end{aligned}$$

There is a lot which is introduced in the above definition. First, we are using $P^{(1)}$ to represent estimates. These estimates are computed using the definitions listed above. Whereas P denotes actual values and are computed by simply counting the data. For example, $P(S_m)$ is found by looking at assignment m, counting the number of students whose grade was S_m, and dividing by the number of students in the class. Notice that we are assuming that each student has an equal chance of being in each class at the beginning of the term. This is a choice which we made. For given the lack of information instructors have about any of the students at the beginning of the

term, the uniform distribution is a reasonable choice for the initial distribution.

Next, the variables C_m represent the class of the student at the m^{th} assignment. The expression $P(S_{m+1} \mid C = C_{m+1})$ denotes the probability of a student, whose assigned behavioral group (C) is C_{m+1}, receiving a grade of S_{m+1} on assignment $m+1$. It can be easy to confuse the different notions of "group" here, and so we shall try to address the matter. With the expression $C = C_{m+1}$, we are not trying to capture the subset of students whose assigned group, given by Dr. Manspeaker, is equal to their Bayesian group at assignment $m + 1$. Instead, C_{m+1} is simply a variable in the expression $C = C_{m+1}$ whose value ranges over the five behavioral groups. So in the definition for $P^{(1)}(C_{m+1} \mid S_{m+1})$, the subexpression $P(S_{m+1} \mid C = C_{m+1})$ is interpreted as the probability that a randomly selected student will get a grade equal to the value of S_{m+1} out of all the students whose group, as assigned by Dr. Manspeaker, is C_{m+1}. The reason we chose to regard these values as random variables, namely by the use capital letters, is that on the left-hand side of the definition, their role is in fact that of a random variable. This overloaded notation is used here because it is standard notation in statistics and because, after having written down some statistical ideas and derivations, this ambiguous and overloaded notation seemed preferable to a more explicit and exact notation. Finally, the sum $\sum P(S_{m+1} \mid C = C_{m+1})P^{(1)}(C_m = C_{m+1})$ is taken over all possible values for S_{m+1} and C_{m+1}.

2.3 Kullback-Leibler Divergence

With each student assigned a probability distribution over Dr. Manspeaker's groups, we then wanted to try to make sense of this new data. We had two data sets, 2009 Fall (F09) and 2010 Fall (F10), which each can be labeled as the training or testing data set. Each of these four choices would produce a model of behavior, that being a set of probability distributions over the behavioral groups. We chose to focus on two models in particular. The first model, called the "forecasting model", used the F09 data as the training data and F10 as the testing. This model forecasts the

groups in the F10 semester based upon parameters derived from the F09 data. Our hopes are that in the future historical data will likewise be used to help predict current data, or in this case simply more recent data. The other model, called the "F10 model", used F10 as both the training and the testing data sets. In this case, we are using F10 to predict itself. We wanted to compare the forecasting and F10 models against each other, where we saw the F10 model as the "truer" model and the forecasting model as the approximation.

To compare these two models, we turned to information theory. The Kullback-Leibler divergence fit our needs. It is used when trying to examine a stream of characters whose frequency is governed by the random variable X, but another random variable Y is being used to encode this stream. The divergence measures the expected number of extra bits the Y encoding will need in order to correctly represent the original stream. In such a scenario, X is commonly referred to as the true, observed, and/or exact distribution, and Y as the theory, model, or approximation of X. For us, X ranged over the distributions from the F10 model, and Y ranged over the F10 model distributions. The Kullback-Leibler divergence of X and Y, denoted $D_{KL}(X||Y)$, satisfies $D_{KL}(X||Y) := -\int_{\Omega} \log \frac{dY}{dX} dX$ for general probability measures with Y being absolutely continuous with respect to X. Hence the formula for discrete distributions is $D_{KL}(P||Q) = -\sum_x P(x) \log \frac{Q(x)}{P(x)}$.

In our case, we used the KL divergence, the abbreviation which we will be using, to compare two sets of prediction models about student groups. Both models had probability distributions for each student at each assignment over the F10 semester. This allowed us to compare the corresponding distributions from the two models with the KL divergence, and then aggregate the results to get an idea of how well the forecast model predicted the F10 model

After the forecasting model did not compare favorably with the F10 model, we used hill-climbing, which is a standard artificial intelligence technique, to find locally optimal forecasting models. We will provide more detail in the next chapter. But the basic idea is that we searched through the forecasting models built from subsets of the assignments in order to find the models which were best predictors of the corresponding F10 model.

2.4 Hidden Markov Models

In another attempt to find a way to cluster the models by student, we turned to another method of inference. Hidden Markov models, or HMMs, deal with systems with "hidden variables" and "observable variables". With a proper sample, hidden Markov models predict the hidden (or difficult to observe) variables given values for the observed variables in the sample data. Take two finite sequences of random variables, a sequence of observable variables $(Y_t)_{t=1}^T$ and a sequence of unobserved or hidden variables $(S_t)_{t=1}^T$, which are also referred to as states. We assume that the states satisfy the Markov condition that S_{t+1} only depends upon S_t for each t, and that each observation Y_t only depends upon the current state S_t. Said more rigorously, each state is conditionally independent of any previous observations and any previous state, given the last state — $P(S_{t+1} \mid Y_\star, S_{t-r}, S_t) = P(S_{t+1} \mid S_t)$ for any choice of $t, r, \star > 0$; and each observation is conditionally independent of any previous observations and any previous state given the current state — $P(Y_t, S_{t+r_1}, Y_{t+r_2} \mid S_t) = P(Y_t \mid S_t)P(S_{t+r_1}, Y_{t+r_2} \mid S_t)$ for all t, r_1, and r_2 (and r_1 and r_2 may be negative). The only other restriction is that the conditional distributions for $Y_t \mid S_t$ and $S_{t+1} \mid S_t$ are independent of t. We will use $A(s, s')$ to denote $P(S_{t+1} = s' \mid S_t = s)$ and $b(s, y)$ to denote $P(Y_t = y \mid S_t = s)$ for any t. Any HMM can be described completely by the parameters A, b, and ν, where $\nu(s) = P(S_1 = s)$.

In our context, the observables, $Y_\star$, are the assignments and the hidden variables, $S_\star$ are the classifications. So then $A(s, s')$ is the probability of going from behavioral group s on one assignment to the group s' on the next assignment, $b(s, y)$ is the probability of a student in group s on some assignment getting a score of y on that same assignment, and finally, $\nu(s)$ is the probability of a student being in behavior group s at the beginning of the course. It is clear that a student's classification over the semester is not likely Markov, as S_{t+1} would depend upon S_t for many values of t. And observations may depend on past classifications. In fact, none of the stated requirements about the distributions hold. So this can at best be seen as a weak model for our

case. Still, if anything can be gleaned from the analysis, we have a clear starting point when examining more complicated Bayesian networks.

Finally, to represent a contiguous subsequence, we will use the notation $y_{s\ldots t}$ to mean the sequence $y_s, y_{s+1}, \ldots, y_{t-1}, y_t$, and for convenience will just assume that $s \leq t$. With these conventions established, we will now go on to describe the relevant algorithm for this work, the Viterbi algorithm which calculates the most likely state sequence taken, given a sequence of observations. We shall give the discrete version of this algorithm, as it is most applicable. However, it can be given in continuous form. This is done in the text by Fraser[2]. For the remainder of the sections, we shall work with some fixed HMM and will therefore take the parameters (A, b, ν) as known.

2.4.1 Viterbi Algorithm

The Viterbi algorithm finds the state sequence $\hat{s}_{1\ldots T}$ which maximizes $P(S_{1\ldots T}, Y_{1\ldots T})$, where the sequence of observations $y_{1\ldots T}$ has been given. In our context, this means that the Viterbi algorithm will return, for each student, a sequence of behavioral states which is the most likely behavioral sequence to have occurred given their scores on all the homework. This result differs from those of Bayesian inference because the Viterbi algorithm returns one behavioral group for each student at each assignment rather than returning a probability distribution over the behavioral groups. Our analysis could now associate a sequence, or string, of behavioral groups to each student, which is simpler than a sequence of probability distributions.

One technical note is that the probabilities are incredibly small, it is actually advantageous to work with logarithms to avoid floating-point arithmetic errors like underflows.

$$
\begin{aligned}
\delta(s,t) &:= \max_{s_{1\ldots t-1}} \log P(y_{1\ldots t}, s_{1\ldots t-1}, S_t = s) \\
\omega(s,s',t) &:= \max_{s_{1\ldots t-1}} \log P(y_{1\ldots t+1}, s_{1\ldots t-1}, S_t = s, S_{t+1} = s') \\
\psi(s',t) &:= \operatorname*{argmax}_{s} \omega(s,s',t)
\end{aligned}
$$

Both δ and ω take the maximum over sequences of states, and ψ returns the state s which maximizes $\omega(s, s', t)$, where s' and t are given. If $s_{1\ldots t-1}$ is the sequence of states which maximizes the probability of $\delta(s, t)$, for some state s and time t, then $s_{1\ldots t-1}, s$ is the most likely sequence of all the sequences of length t which end in s, given $y_{1\ldots t}$, a sequence of observations. A similar statement can be made for $\omega(s, s', t)$, namely that the sequence which achieves the maximum probability will be the most likely sequence of all those that have length $t + 1$ and end with s, s'. An astute reader may guess that the Viterbi algorithm will construct the most likely sequence starting with the last element and ending with the first. This is indeed the case. But we still need to derive a few equations from the above definitions in order to fully understand the algorithm.

The following derivation justifies the correctness of the final equation, that will be employed in the algorithm. This computation utilizes the independence assumptions of HMMs, takes loga-

rithms of both sides, and finally takes the maximum over $s_{1\ldots t-1}$.

$$\begin{aligned}
P(y_{1\ldots t+1}, s_{1\ldots t+1}) &= P(y_{1\ldots t}, s_{1\ldots t})P(y_{t+1}, s_{t+1} \mid y_{1\ldots t}, s_{1\ldots t}) \\
&= P(y_{1\ldots t}, s_{1\ldots t})P(y_{t+1}, s_{t+1} \mid y_t, s_t) \\
&= P(y_{1\ldots t}, s_{1\ldots t})P(s_{t+1} \mid s_t)P(y_t \mid s_t) \\
\log P(y_{1\ldots t+1}, s_{1\ldots t+1}) &= \log P(y_{1\ldots t}, s_{1\ldots t}) + \log A(s_t, s_{t+1}) + \log b(s_t, y_t) \\
\omega(s_t, s_{t+1}, t) &= \delta(s_t, t) + \log A(s_t, s_{t+1}) + \log b(s_t, y_t)
\end{aligned}$$

Recall that the matrices A and b are parameters of the hidden Markov model, with $A(s_t, s_{t+1})$ being the transition probability of moving from the state s_t to s_{t+1} and $b(s_t, y_t)$ being the observation probability of witnessing y_t while being in state s_t. The next equation follows from the definitions of δ, ω, and ψ, and is used to calculate values for δ by iteration.

$$\delta(s_{t+1}, t+1) = \omega(\psi(s_{t+1}, t), s_{t+1}, t)$$

This algorithm itself has two phases. In the first stage, the algorithm constructs values for ψ, ω, and δ. The algorithm builds these values iteratively over increasing values of t. Further, we only need to know ω for the current value of t, and δ for the current and previous values of t. So we store $\omega(\cdot, \cdot, t)$ as a two-dimensional array for the current value of t, and store δ in two one-dimensional arrays $\delta_{\text{old}}(\cdot)$ and $\delta_{\text{new}}(\cdot)$ for the previous and current values of t, respectively. But there is no optimization for ψ, and so it is stored as a two-dimensional array $\psi(\cdot, \cdot)$. The second stage of the algorithm determines the most likely end-state using δ, then steps backwards over values of t to get the remaining states from ψ.

Algorithm 1 Viterbi Algorithm

for all $s \in \mathcal{S}$ **do**
  $\delta_{\text{new}} \leftarrow \log P(Y_1 = y_1, S_1 = s)$
end for
for $t = 2 \rightarrow T - 1$ **do**
  $\delta_{\text{old}} \leftarrow \delta_{\text{new}}$
  for all $s_{\text{new}} \in \mathcal{S}$ **do**
    for all $s_{\text{old}} \in \mathcal{S}$ **do**
      $\omega(s_{\text{old}}, s_{\text{new}}) \leftarrow \delta(s_{\text{old}}, t) + \log A(s_{\text{old}}, s_{\text{new}}) + \log b(s_{new}, y_{t+1})$
    end for
    $\psi(s_{\text{new}}, t + 1) \leftarrow \arg\max_{s_{\text{old}}} \omega(s_{\text{old}}, s_{\text{new}})$
    $\delta_{\text{new}}(s_{\text{new}}) \leftarrow \omega(\psi(s_{\text{new}}, t + 1), s_{\text{new}})$
  end for
end for
$\hat{s}_T \leftarrow \arg\max_s \delta_{\text{new}}(s)$
for $t = T - 1 \rightarrow 1$ **do**
  $\hat{s}_t \leftarrow \psi(\hat{s}_{t+1}, t + 1)$
end for
return $(\delta_{\text{new}}(\hat{s}_T), \hat{s}_\cdot)$

2.5 Importance of a Student's Early College Experience

Our findings with HMMs suggest that students settle into their behavioral groups very early in the semester. There is a great deal of research about the importance of a student's first year, or even first semester, of college. However, there does not seem to be much work involving student behavior over any shorter time intervals. There is a fair bit in the literature about how syllabi affect student attitudes. In addition, the author found one article talking about attitudes and activities done in the first week of class. So the literature does offer some support for one of the major findings of this work, that students' attitudes and behaviors change in the beginning of the semester. But this work is unique in the literature because we predict that student behaviors settle by the second week of class.

Another reason why the literature about affecting student behavior at the beginning of the term is important to this work is that we found student attitudes and behaviors begin to stabilize after the second week of class. As there is little instructor interaction with the students in the first two

weeks, any means of changing student behavior for the better is important. Further, it is likely that the ways in which we might alter their attitudes is through subtle means, such as body language, tone of language and voice, and gestures. Reviewing the literature helps to reveal some of the subtle means of influence, and allows us to see how effective each technique may be.

2.5.1 Tone of the Syllabus

The articles about course syllabi investigated the connections between the language and content of the syllabus and the students' attitudes towards the instructor and the course. In one study by Harnish and Bridges[3], the authors sought to find a connection between the tone of language used in the syllabus and the students' attitudes regarding the instructor and course difficulty. Students were told that they were evaluating a candidate for an adjunct position in the department, but due to "scheduling difficulties" the candidate was unable to present the lecture. To grade the phony candidate, students were given a syllabus, ostensibly written by the instructor, and a questionnaire. The syllabus was either the friendly or unfriendly version, and the two versions were distributed randomly. A basic skeleton syllabus was used, so that the two versions did not differ in content or layout. Below are some examples of how the versions differed. Testing 172 students, the authors found that there was a statistically significant effect of increasing instructors' approachability and their perceived motivation when the syllabus was warmer in tone. Also, the students who received the friendly syllabus preceived the course to be less difficult than those who received the unfriendly version.

	Unfriendly	Friendly
Office Hours	If you need to contact me outside of office hours, you may email me, call my office, or contact the department and leave a message.	I welcome you to contact me outside of class and office hours. You may email me, call my office, or contact the department and leave me a message.
Class Participation	Come prepared to actively participate in this course. This is the best way to engage yourself in learning the material (and it makes the lectures more interesting...).	I hope you actively participate in this course. I say this because I found it is the best way to engage you in learning the material (and it makes the lectures more fun)...

2.5.2 Offers of Help in Syllabus

With evidence that the tone of the syllabus affects students' attitudes, Perrine, Lisle, and Tucker[6] wrote an article suggesting that when an instructor includes an offer of help in the syllabus, it may affect students' beliefs. In particular, they showed that the students will be more willing to seek help from the instructor when an offer of help is included in the syllabus. In a similar experimental design as the previous paper, 104 students were given a brief description of a course along with a statement from the instructor's syllabus. The two varieties of statements follow. The results showed that there was a significant effect on students' responses to a survey about how willing they were to seek help from the instructor. Concerning a number of problems (e.g., trouble understanding textbook, low grade on first exam, and hard to hear instructor's lectures), the students who saw the supportive syllabus felt more willing to seek help.

Supportive	The course is enjoyable, but demanding. There is a large amount of material and it can be overwhelming at times. If you find yourself doing poorly in the course, please come talk to me. Any time during the semester that you have problems in this course, I want to know about it. Together we can try to pinpoint the problem and get you off to a better start.
Neutral	The course is enjoyable, but demanding. There is a large amount of material and it can be overwhelming at times. Please do not let yourself fall behind. It is very important that you keep up with the readings.

2.5.3 Syllabus Detail

The last article about syllabi, authored by Saville, Zinn, Brown, and Marchuk[8], focused on the level of detail in the syllabus and students' perceptions of the effectiveness of the teacher. Half of the 97 student participants were given a brief syllabus while the other half were given a more detailed version. The brief version was two pages long while the long version was six whole pages. Both versions contained the course objectives, information about the textbook, basic description of the assignments, grading scale, brief statement of course policies, and a schedule listing assignment due-dates. But the detailed version expounded upon items whenever possible. In one area, the detailed version went beyond saying that there were six exams by stating the type of problems which would appear on them, e.g., multiple-choice, essay, etc. Also the detailed document had a course schedule that included due-dates as well as which chapter would be covered each class and which ones would be covered in each exam. The researchers tested whether the amount of information in the syllabus would affect beliefs such as how personable, creative, encouraging, enthusiastic, knowledgeable, prepared, and fair the instructor was. In addition, the students'

comprehension of the syllabus was queried. The findings were that comprehension did not significantly differ between versions and had little correlation to the aforementioned beliefs. Further, these beliefs were more likely to be stronger with the students who received the more descriptive syllabus.

These papers about syllabi affecting student beliefs do reinforce the idea that a syllabus is the first introduction students have for the course. It is our first tool, as instructors, to sell the course to the class, and to start them off with a good first impression.

2.5.4 Icebreakers

This last paper which we will discuss deals with icebreakers and introduction games. Although such shenanigans may seem more appropriate for a humanities class, they are included because author's criterion for teaching techniques is "by any means necessary." So if such activities can affect student attitudes, and ultimately result in improved student learning, then these practices are worth consideration at least, if not worthy of being put into practice.

Hermann, Foster, and Hardin[4] looked at how icebreakers might affect students' feelings of support and satisfaction in the course. Before the start of a large, introductory, lecture class of 377 students, the instructors were divided into two groups: one which was only told to conduct a typical first day and another which was additionally told to do a reciprocal interview (this term is described below), where the students would ask questions of the teacher and then the teacher would ask questions of the students. This icebreaker involves the class dividing into small groups, in which they elect a representative and choose questions for the representative to ask the instructor. After the student representatives ask the instructor questions, the instructor asks the representatives questions. Finally, the groups again choose another student to ask the instructor a second round of questions. This process of back and forth interviews is what is meant by the phrase "reciprocal interview". At the end of the semester, students were asked to fill out a survey about how supportive and how clear the instructor was. In the sections which performed the re-

ciprocal interviews, the students reported perceiving the instructor to have been more supportive and clearer. The authors do admit, however, that the findings may have occurred only with the instructors for that term. And so this result may have had less to do with the icebreaker activities and more to do with the particular instructors who used icebreakers.

2.5.5 Summary of the Literature

All these studies indicate that student attitudes can be influenced by instructor choices as early as the first day of class. This is encouraging for us and our finding that attitudes and behaviors can start stabilizing for most students within the first few weeks of class. Also, these articles give us ideas as to how we can affect attitudes early in the semester. But recall that Dr. Manspeaker originally clustered her groups based upon behaviors up to and including exam 1. Granted the above articles do not mention anything about student behaviors, but if attitudes can change so early in the semester, it seems plausible that students' behaviors can also change early.

Chapter 3

Data-Mining Narrative

3.1 Data Collection

We collected data on the Math 100 Studio College Algebra course. This course was chosen primarily because it has one of the largest enrollments in the department. Assignment data were already being collected through the university course management system. More importantly, the Studio College Algebra course remains nearly identical from one term to the next. With a single coordinator (currently Dr. Rekha Natarajan), assignments are similar between semesters, the schedule remains almost unaltered, and exam questions are equivalent but for handful of exceptions. We wanted to compensate for the rare times when exam questions were not congruent in order for the comparison of the two classes to be more accurate. These exceptions may be due to a reordering of the questions or to the addition or removal of a question. In case of reordering, adjustments were made so that like questions are aligned in the data sets. Otherwise, the questions with no analogue were just dropped from the data sets. Finally, we also included Dr. Rachel Manspeaker's groups, for which she additionally used student interviews to learn about her classification groups[5].

We used data from Fall 2009 ("F09") and Fall 2010 ("F10") terms. We chose to use two Fall terms rather than consecutive Fall and Spring terms, as the Fall and Spring consist of different

	I	OA	E	SS	RM	Total
Fall 2009	36 (15.3%)	98 (41.7%)	80 (34.0%)	15 (6.4%)	6 (2.6%)	235
Fall 2010	62 (23.0%)	122 (45.2%)	15 (5.6%)	61 (22.6%)	10 (3.7%)	270

Table 3.1: Size of Dr. Manspeaker's groups

types of students. In the Fall term, the class is made up mainly of freshmen. The Spring term however consists primarily of older students, many of whom are retaking the course. Even Dr. Manspeaker noted this difference in her analysis. She found that the behavioral groups were the same for Fall and Spring terms. However, the proportion of students in the groups differed between Fall and Spring terms. Still these proportions stayed roughly similar, with only a few large shifts, between one Fall term and the next Fall term or between consecutive Spring terms (Table 3.1). Although, in this case, there is a substantial difference in the proportion of employees and smart slackers. This discrepancy remains unexplained, but is assumed to be atypical. But for this work, this difference is unimportant, for our goal is to be able to detect such shifts in the data.

As our goal for this work is identifying trends in models of student groups, we wanted to have data sets which were as homogeneous as was reasonable. So we focused our attention upon the comparing Fall terms with each other. While we could have also compared two Spring terms, our intuition was that first-year students would exhibit the greatest change in behavior as they are adapting from secondary school and life with their parents to college life, where parents are absent. Older students undoubtedly continue to change their behaviors, but we suspected that these adjustments are relatively small compared to those in their younger counterparts.

With these two data sets, we built two models. The first was the experimental model and was trained on F09 data to predict F10 groups. The other model, called the observed model, was trained on F10 data and predicted F10 groups.

3.2 Bayesian Learning

Initially, we were classifying students based on a conceptual exam, the Pre-calculus Conceptual Analysis or PCA, developed at Arizona State University[1]. Our plan was to use one Bayesian network for each of the two classifications, Dr. Manspeaker's behavioral and the PCA's conceptual classification. The conceptual classification was based upon the change in the PCA score at the beginning and end of the term. Unfortunately, there was generally little change in these scores. So we had difficulty classifying students based upon such small fluctuations. Because of these minimal changes, we moved away from Bayesian networks and instead focused more on general, repeated Bayesian inference. Our fear was that a full Bayesian network would over-fit this data and the weak PCA classification. So instead, we opted to use a simpler Bayesian inference model which would only look to the previous assignment when predicting the next, whereas, a Bayesian network would start by considering all past assignments and keeping just enough memory to be a good predictor. With Dr. Manspeaker's classifications however, we had no such trouble, as the classifications were already done. Further, her behavioral groups had clear prototypes which were simpler to distinguish from one another.

With Bayesian networks discarded for simple Bayesian inference, we ran the algorithms which generated probability distributions for each student at each assignment. These distributions assigned a probability to each group, which we interpreted as the probability of that particular student being in that particular classification group at the time of some fixed assignment. Such a distribution is graphed in figure 3.1. In this example, each set of vertically stacked bars represents one probability distribution at one point in the term. The height of the bars is the probability the student is in the group, which is designated by the color of the bar, at a particular point in the course. Notice from the header on the figure, that this particular student was classified by Dr. Manspeaker as an over-achiever. Despite an initial prominence of the intervention group, this student is quite clearly an over-achiever according to the distributions. A more astute observer will

notice that in the initial set of stacked bars all are of equal height. This is because we need to choose an initial distribution for every student, and we choose to use the uniform distribution.

These Bayesian models gave us a model of the students over time. In order to test the validity of this model, we wanted to get some predictions from the model which might then be confirmed or falsified with further study. In particular, we were interested in predictions about when and how an individual student might change classification. We may, for example, then test such claims by interviewing the student around such pivotal times to try to detect any transition between groups.

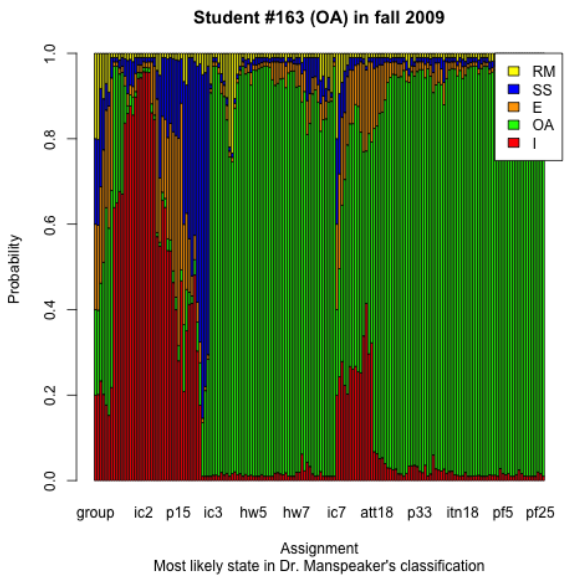


Figure 3.1: Individual's Behavioral Profile

One of the first decisions, aside from what tools to use in order to analyze the data, is how to identify when a student changes from one group to the next in the Bayesian inference analysis. Initially, we settled upon the idea that for the PCA classification, a student would "be in" one of the three groups at a fixed time if the associated probability was at least 50%. This required the introduction of a "no group" classification for the times when a student did not reach this threshold for any of the three groups. For Dr. Manspeaker's behavioral analysis, we simply choose to say that a student was in the group which was the mode of the distribution. The difference of requiring a majority for the PCA rather than taking the plurality as with Dr. Manspeaker's groups was motivated by the fact that there was little change between the post-test and pre-test scores on the PCA exam. So, to reflect such uncertainty in the PCA data, we demanded more of the statement "student X is in group Y on assignment Z".

With this mapping of students into classifications, we were then able to say when a student switched groups. Recall that one of our tasks was to identify when any individual student would

change groups. So the group assignments were run through various clustering algorithms. The hope was that there may be distinct types of students based upon how they moved through Dr. Manspeaker's or the PCA's classifications, a sort of meta-classification — a clustering of how students moved through the original clusters. Unfortunately, we could not find much, as the data did not admit obvious groupings.

We did find something interesting with the Bayesian group assignments for Dr. Manspeaker's groups. We checked these assignments against Dr. Manspeaker's original assignment. The results of these comparisons can be found in table 3.2, 3.3, 3.4, and 3.5. As there are Bayesian group mappings for every assignment in the course, we had to narrow our focus and narrowed it to the three exams and the final exam. In each of the four tables, the Bayesian group mappings are represented in the columns and Dr. Manspeaker's groups are in the rows. So the entries of these tables give the numbers of students (in the Fall 2010 course) whom Dr. Manspeaker assigned to group x and whom the Bayesian data assigned to group y. With this, we have a means of validating the Bayesian inference approach. Though these tables will not validate our models as correct, they remove the doubt that the Bayesian group assignments are unrelated to Dr. Manspeaker's.

One of the first thing to notice in these four tables is that the entries of the main diagonals are often, but not always, dominant in their respective row and column. These diagonal entries represents how many students were placed into the same groups by Dr. Manspeaker and the Bayesian model, and in some sense, how many students were "correctly" placed by the Bayesian model. In particular, we see that over-achievers are the easiest for the model to identify, followed by the intervention students. Whereas employees, smart slackers, and rote memorizers are difficult to detect. With the over-achievers, notice that in all four tables the diagonal entres, which represent those students correctly placed as over-achievers, are the maximum entry in both their row and column. The placement into the intervention group is successful in a similar way with all but the first table. As for the employees and rote memorizers, notice that Dr. Manspeaker only identified fifteen and ten students as respective members of these groups. So it is not surprising that the

Assigned\Bayes	**I**	**OA**	**E**	**SS**	**RM**	**Total**
I	12	6	17	8	19	62
OA	2	88	5	22	5	122
E	0	7	4	3	1	15
SS	8	19	11	16	7	61
RM	9	0	0	1	0	10
Total	31	120	37	50	32	270

Table 3.2: Dr. Manspeaker's vs. Bayesian groups at exam 1

Assigned\Bayes	**I**	**OA**	**E**	**SS**	**RM**	**Total**
I	22	16	13	6	5	62
OA	13	66	25	11	7	122
E	7	5	3	0	0	15
SS	19	22	10	7	3	61
RM	8	0	2	0	0	10
Total	69	109	53	24	15	270

Table 3.3: Dr. Manspeaker's groups vs. Bayesian groups at exam 2

model is having trouble placing students into these groups, as it does not have a good sample of such students. This leaves us with the smart slackers group. We will talk more about predicting this group with the Bayesian model in the next chapter. All we shall say for now is that the smart slacker group is unique among the behavioral groups because it is the only group which inherently has two very different ways of behaving: its typical member is, depending upon the time in the term, quite competent and, at other times, is a poor student. This may be a problem for the models presented in this study, and may potentially be a problem for even Dr. Manspeaker's model.

We looked carefully at table 3.2. While the other three tables are interesting, there are two fundamental problems with comparing Dr. Manspeaker's group assignments with the Bayesian group assignments. The first is that all three comparisons are between group assignments at different times in the semester. Dr. Manspeaker classified students just after the first exam, and these particular Bayesian classifications model the students at a later exam. So it is reasonable that the off diagonal cells in the later three tables will be larger than the corresponding cells in the first

Assigned\Bayes	**I**	**OA**	**E**	**SS**	**RM**	**Total**
I	27	12	15	3	5	62
OA	16	55	23	5	23	122
E	5	2	3	2	3	15
SS	9	14	22	6	10	61
RM	5	0	3	1	1	10
Total	62	83	66	17	42	270

Table 3.4: Dr. Manspeaker's groups vs. Bayesian groups at exam 3

Assigned\Bayes	**I**	**OA**	**E**	**SS**	**RM**	**Total**
I	19	22	14	3	4	62
OA	24	62	30	4	2	122
E	6	5	1	0	3	15
SS	13	25	15	2	6	61
RM	7	1	1	0	1	10
Total	69	115	61	9	16	270

Table 3.5: Dr. Manspeaker's groups vs. Bayesian groups at final exam

table. The second reason the first table deserves more scrutiny than the others is that we expect student behavior to change over time. Not only will group assignments at the first exam become less relevant as the term progresses, but the students themselves will change. Hence, the efficacy of comparing any two group assignments quite dependent upon the difference in time between them.

There are four large off-diagonal entries in the first table. Two of these entries are the assigned over-achiever and predicted smart slacker (22) and the converse, assigned smart slacker and predicted over-achiever (19). Recall that the smart slackers do poorly in most aspects of the course with the exception of the first exam, where they perform anywhere between moderate and excellent. With both groups doing reasonably well on the first exam, this helps to explain a bit of the confusion between the two groups up to this point in the class. It may also be the case that by the first exam, some of the students who started the term as over-achievers start to become smart slackers as the term progresses. With good anecdotal evidence for this transformation, we have

another possible source of confusion between over-achievers and smart slackers. The other two entries are those where the students who were assigned to the intervention group and predicted to be an employee (17) and those who were also assigned to the intervention group and predicted to be rote memorizers (19). Recall from Dr. Manspeaker's work that intervention students have similar grades as rote memorizers with the exception of attendance and written homework. As for mistaking employees for intervention students, it is important to remember that employees typically have the second highest grades in all categories except attendance. And in attendance, employees do well at the beginning of the term but eventually stop attending. So both employees have good attendance early in the term, do well on written homework, are both strong on procedural exam questions, and are weak on all other types of exam questions. It is therefore not surprising that these two groups, employee and intervention, are confused early in the class.

3.3 Walking through Model Space

So the Bayesian groups failed to cluster. But the Bayesian inference and group mapping was done with all of the assignments in the course. We could instead take a subset of the assignments, and run the inference algorithm using only the data from this subset of grades. We thought that it might be possible that some assignments were interfering with group predictions and clustering. For some assignments might be too easy or too hard, and thus mislead the inference algorithm. So some subset of course work might be better at predicting a student's group than others.

We cannot examine each model built of a subset of data as there are $2^{165} \approx 4.67680 \times 10^{49}$ models in total. So we decided to use hill-climbing, which is a standard exploration technique in artificial intelligence. Hill-climbing provides us with a technique of searching for a local extreme in the space of all models. Below is an outline of the hill-climbing algorithm.

1. Select a random set of assignments and generate the corresponding model.

2. With this as the current model, compute all neighboring models — models with either one

more or one less assignment than the current.

3. Compare the current model with all of its neighbors. If some of the neighbors are better, take the best one as the current one and repeat at step 2. Otherwise, we have found a local optimum.

We hoped that such a technique would winnow the assignment set down to the essential ones.

As to the question about how one model is better than another, we employed the Kullback-Leibler divergence (KL divergence) from information theory. When comparing the fitness of a subset of assignments in their ability to predict student groups, we would build two models to predict the groups of F10 data, one which is trained on F09 data and another trained on F10 data. The F09/F10 model was our "test" while the F10/F10 was the "control". By comparing the test model against the control, we aimed to learn how good the particular subset of assignments, upon which the models were built, was at predicting groups. This fitness would be determined by the KL divergence. The two models gave us two probability distributions for each assignment and each student from F10. Each corresponding distribution was compared via the KL divergence, and the median divergence across all the students and assignments was used as the metric for comparison.

Initially, models were compared using the mean of all the KL divergences. However, after numerous tests, the locally optimal models obtained from the walks typically used only one assignment and never more than ten assignments. Then the median was chosen as it would not be affected by a few poorly predictive distributions. It is standard when performing hill-climbing, to perform multiple "climbs" which start at random locations. The reason is that hill-climbing only returns a local optimum. And returning multiple local optima may provide insight into which assignments are more predictive by examining which assignments occur more frequently. Two sets of walks were performed. The first set allowed models to draw from all the assignments in the class, but the second set removed the questions from the first exam. We performed this sec-

ond walk after observing the prominence of the first exam and final exam questions in the locally optimal models. This was done to test whether the many occurences of questions from the first exam was due to some intrinsic importance. We discuss the results and implications more in the following chapter. For now, we shall merely acknowledge the two sets of walks. Each set of walks had ten walks which started randomly at a model built from 30 assignments, five walks started at random models of 60 assignments, four walks at random models of 90 assignments, and one walk at a random model of all the assignments. Additionally, with the extra assignments, we ran three walks in the first set at random models of 120 assignments. With the first exam removes, there weren't enough assignments to perform walks starting at models of 120 assignments. So in total, 43 walks were performed over the two sets — the first set comprised 23 walks, and the second set accounted for 20 walks.

Below is a table summarizing the results of the hill-climbing through the various Bayesian models. The tallies are divided up in two ways: by assignment category and by approximate time during the semester, in other words between which two subsequent exams the assignment occurs. The assignment categories have been abbreviated in order to fit the table into the margins. The categories are, in order, attendance (Attend), extra credit (EC), homework (HW), online homework (OHW) [which is itself subdivided], lecture (Lect), studio (Studio), and exam questions (Exam Qs). The online homework data used here is written as two numbers separted by a forward slash "/". Of these two, the first number represent the number of occurrences of the scores of these assignments during the particular time. And the second number represents the number of occurances for, what is known in our research group as, the inverse time to ninety of the assignments. With our in-house online homework system, the students are not limited in the number of times they may submit the homework for grading. All of the submissions from a given student, which is submitted prior to the deadline, are graded but only the best grade is recorded while the others are ignored. The phrase "inverse time to ninety" refers to the reciprocal of the number of attempts it took the student to achieve a grade of at least 90%. A student who does not reach this mark on

	Attend	EC	HW	OHW	Lect	Studio	Exam Qs	Total
$\star \leq$ **Exam 1**	86	21	5	28/11	13	3	72	239
Exam 1 $< \star \leq$ **Exam 2**	43	0	14	43/27	8	8	70	214
Exam 2 $< \star \leq$ **Exam 3**	49	0	5	31/14	34	6	32	171
Exam 3 $< \star$	39	23	14	30/12	36	11	102	268
Total	217	44	40	132/64	91	28	276	892

Table 3.6: Walks

	Attend	EC	HW	OHW	Lect	Studio	Exam Qs	Total
$\star <$ **Exam 1**	70	18	6	17/11	8	5	-	135
Exam 1 $< \star \leq$ **Exam 2**	36	0	9	32/18	16	0	64	175
Exam 2 $< \star \leq$ **Exam 3**	45	0	5	15/10	27	8	28	138
Exam 3 $< \star$	32	20	11	21/14	25	10	88	221
Total	183	38	31	85/53	76	23	180	669

Table 3.7: Walks with exam 1 dropped

an assignment is assigned a zero for their inverse time to ninety. This metric acts a velocity, how quickly does the student "move" through the assignment. The data are contained in Table 3.6.

We noticed that the questions of the first and final exam featured prominently in the hill-climbing results. This observation led to our next inquiry. The question we investigated was whether the first exam was intrinsically important itself or whether its significance derived from the fact that it impacted students first and overshadowed the effects of the other exams. If the first exam merely overshadowed the importance of the other exams, then one would expect the second exam to rise in importance with the first one removed. Or the third exam would be more significant if the second exam has little importance. So the hill-climbing was repeated but with all of the first exam data. The results are in Table 3.7.

While interesting, these results did not answer our questions about identifying changes in student behavior. The results of the walks through model space will be discussed in the next chapter. For now we will stick to the narrative of our data-mining journey.

3.4 A Hidden Markov Model in the Data

Around this point in time, we realized that we could return to the idea of using Bayesian networks, namely hidden Markov models (HMM). An astute reader may recall that our reason for choosing Bayesian inference over more general Bayesian networks was the worry that the networks would over-fit the data. Like our inference method, HMMs only depend upon the previous state and the current observation. So there wasn't a great deal of concern about over-fitting with a HMM. Also, we had decided to abandon the PCA data, which was the data with marginal variance and the significant possibility of being over-fit.

The idea is that Dr. Manspeaker's behavioral groups served as the hidden states, and the grades as the observable states. As for the parameters of the HMM, we could compute the state transition probability simply by using the distributions from Bayes inference, the probability of observing a particular grade given the student's group can be derived from the Bayes data and the grades for the semester, and finally the initial state distribution can be taken to be uniform or computed from training data. As we did with the Bayes models, we can build the HMM parameters with one training data set, in order to get a model of a testing data set. For example, we can build a HMM model with the 2009 data and use that for predictions of the 2010 data.

For us, the natural analysis was the most likely path through Dr. Manspeaker's groups for each student. Having done this, we immediately tried clustering the paths again. But as before, the group assignments didn't admit an obvious clustering. After a bit of thought, we came to the idea of focusing on the times when a student belongs to the same predicted group over a number of assignments. We called these sequences of a constant group membership a "run" for the student. For example, suppose that some student has been assigned to the Employee group by the Bayesian model for each of the sixteen questions on the second exam; then we would say that this student is in a run of the Employee group which starts at the first question of the second exam and has a length of 16. At first, we wanted to focus on each student's longest run, and then later analyzed all

Dr. Manspeaker \ Run	I	OA	E	SS	RM	Total
I	10	44	8	0	0	62
OA	5	110	7	0	0	122
E	3	8	4	0	0	15
SS	4	50	7	0	0	61
RM	5	4	1	0	0	10
Total	27	216	27	0	0	270

Table 3.8: Dr. Manspeaker's Fall 2010 group versus the group of longest run

runs for each student which were at least 10 long. The minimum length of a run was chosen to be ten because during the semester, ten assignments covers about a week and a half with, of course, the exception of the exams which are more than ten questions. This time frame is roughly a little more time than it takes for a student to have turned in a written assignment, and have it returned graded. The hope was that students might be prompted to a change upon seeing their work graded.

Table 3.8 shows the number from students of Fall 2010 in each of Dr. Manspeaker's groups versus the group of their longest run. Here the longest run is obtained from the model trained on 2009 data and tested on 2010. We can see the difficulty the model has with students which Dr. Manspeaker identified as smart slackers or rote memorizers. As there were only fifteen smart slackers and six rote memorizers out of 235 students in 2009, it is understandable how the HMM is struggling to correctly place these groups. The model also had trouble with employee students. This is reasonable, as they were the hardest for Dr. Manspeaker to classify as well. The model did well in placing the over-achievers. But for some unknown reason, the runs also had trouble with the smart slackers.

What was particularly interesting was how quickly people started their longest run. Table 3.9 shows the number of students who had started their longest run by the end of each exam and in which group the run was. Notice that by the first exam, about 63.0% of the students had already started their longest run. In fact, if we examine further, we can see that 164 students ($\approx$ 60.7%) have started their largest run by the second week of class. This warranted further study. For we

# of Students	**I**	**OA**	**E**	**SS**	**RM**	**Total**
Exam 1	18	140	19	0	0	177
Exam 2	25	173	23	0	0	221
Exam 3	27	215	27	0	0	269
Final	27	216	27	0	0	270

Table 3.9: Number of students who have started their longest run prior to the corresponding exam

now have an hypothesis as to how soon students start to settle into a pattern of behavior.

The drawback of the Viterbi analysis, however, is that we cannot determine student groups in real time. The way the algorithm works, the selection of the most likely path varies greatly as new data are added. So the paths obtained mid-semester would change drastically as data are added. We would need to wait until the end of the semester before we could run the algorithm to get the most likely path, and then analyze the runs of that path. But if these models of behavior have decent accuracy, then we have made significant progress. For while real-time data would be preferable, merely having the information is an improvement over our present situation.

We will conclude this chapter with the last bit of data analysis done for this project. We examined all the runs of a student which are at least 10 assignments long. By ignoring everything else, we seemed to be throwing out the "noise" of fluctuating behavior and focusing on the strong, consistent "signals". We could then look at how many students were in a run at any given time. This is shown in figure 3.2.

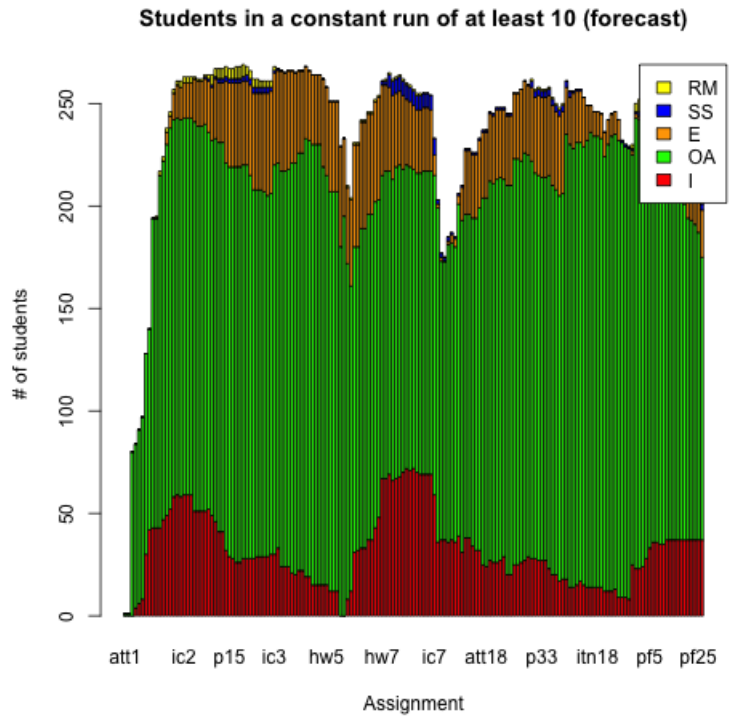


Figure 3.2: Long Runs

Chapter 4

Interpretation

Despite the sophistication of the analysis and models, we have yet to provide any empirical support for them. However, all of this work can still be useful, especially in the coming years. What we have is a basis on which we may formulate further claims for further investigation. This is the purpose of the current chapter: to pose hypotheses from the models, muse upon their possible implications, and suggest ways of testing such claims. We shall focus first upon the results from the walks through the space of naive Bayesian models and then focus upon the findings from the analysis of runs in the Viterbi paths.

4.1 Bayesian Walks

Recall that at the end of these walks, we have a collection of subsets of assignments which did a locally optimal job of predicting the Fall 2010 model from the Fall 2009 data. We will refer to an assignment one of these subsets as an assignment instance. Obviously, any particular assignment may have multiple instances within the collection. The amount of instances an assignment has in this collection is what we wish to examine. In Table 3.6, we see that exam questions are the most prominent measures, which account for 276 of the 892 instances ($\approx 30.9\%$). Next most frequent

is attendance in recitation and studio, which appeared 217 times ($\approx$ 24.3%). Together, these two assignment categories comprise over half of all instances found in the locally optimal models.

And within these categories we find more interesting patterns. One of the most striking is that questions from the third exam only occurred 32 times ($\approx$ 11.6% of the 276 exam question instances). This is less than half of the next smallest subcategory, that being the second exam with 70 instances. To provide a bit of context, exam two covers quadratics and transformations of functions; while exam three deals with the fundamental theorem of algebra. The first exam reviews basic algebraic techniques, and the final exam is cumulative and includes exponentials and logarithms. Within attendance scores, the most prominent subcategories are those which happen prior to the first exam, having 86 of the 217 appearances ($\approx$ 39.6%). The attendance between the second and third exams is the next most frequent, appearing 49 times. The Fall semester starts after mid-August, with the first exam around mid-September, second exam before mid-October, and third exam within the first two weeks of November.

In short, attendance and exams were most prominent in the Bayesian walks, and such data occur more frequently earlier rather than later in the term. These findings support our belief that students settle into behavioral groups early in the semester. We will now explore implications and explanations for these observations. First, we explore how these analyses relate back to Dr. Manspeaker's groups. Then we reflect on how attendance and exams are unique evaluations of students and why they may be better at predicting behavior. Finally, we focus upon how to improve student behavior using this information.

4.1.1 Reinforcing Dr. Manspeaker's Groups

One of the first interpretations we came to was that the results of the Bayesian walks were evidence which supported Dr. Manspeaker's thesis. The data suggest that attendance and exams were the more accurate predictors of membership in her conceptual groups from one term to the next. And our findings are consistent with hers. Recall that she found that over-achievers and intervention

		Attendance		
		Good	**Drops off**	**Poor**
Exam Scores	**Good**	Over-achievers	—	—
	Adequate	—	Employees	Smart Slackers
	Poor	Intervention	—	Rote Memorizers

Table 4.1: Qualifying Dr. Manspeaker's Groups by Attendance and Exam Scores

group members had good attendance, while the smart slackers and rote memorizers quickly stop attending. As for the employees, they start the semester with good attendance, but this good behavior eventually disappears. Whereas with exams, Dr. Manspeaker saw that smart slackers do well only on the first one, employees do alright overall, and the rote memorizers and intervention group tend to do the worst. Of course, the over-achievers excel on the exams. So one could infer just from her results that exams and attendance would be good variables to differentiate the various groups, as is illustrated in Table 4.1.

One of the problems with this analysis is that we are confirming Dr. Manspeaker's' results by using her results. At least on the surface, this criticism weakens when one realizes that the techniques employed in the two approaches are rather distinct. However, there may yet be some undiscovered connection between the two methods used which would validate this accusation of a circular analysis. Still, the observations present in this work give us new predictions to investigate.

One item to investigate is comparisons between individual sections of the course. The data used for this work did not include the section to which each student belonged. It would be interesting to see whether individual instructors had enough of an effect to impact the data, and whether one can determine the instructor based solely upon grade-book data. If it is possible to identify instructors by their students' grades, then it seems likely, based on previous attempts to judge instructors by teaching ability, that the data would only be able to do so in a few cases. This would reflect current wisdom, that it is difficult to quantify instructor ability for all but the most extreme cases. But even if we were unable to correlate grades with instructors, then this lack of correlation

would indicate that other changes may be more profitable in increasing student understanding.

Another research idea is to use the analysis techniques presented here to explore the differences between students in Fall and those in Spring terms. We have some data from Dr. Manspeaker's work. She found that while the behavioral groups are the same, their relative sizes vary between Fall and Spring. It would be very interesting to see whether students' behavioral patterns change as well. If this is the case, then a typical Spring semester employee would be different from a typical Fall semester employee. So not only would the proportions of the behavioral groups be different between semesters, but the groups themselves would be comprised of students whose conduct varied over the terms as well.

4.1.2 Stable Measures

Another proposition is that attendance and exams are more "stable" measures. While assignments like written homework can vary between terms and textbooks, attendance and exams are relatively constant in the Studio College Algebra course. This idea about the stability of attendance and exams quickly runs into trouble with online homework and studios, as these two are also very consistent across semesters. But what separates these two from attendance and exams is the possibility of collaborating on studios and of repeating the online homework as often as the student likes. This makes studios and online homeworks more formative assessments of the student's ability, rather than summative assessments. However, it is not unheard of for students to "collaborate" on exams. But we feel that it is safe to ignore this, as cheating, which is actively discouraged and monitored, is likely to be minimal in comparison to collaboration on homework assignments, which is encouraged.

As improving exam scores is almost the focus of the class, we feel it may be easier and more interesting to try to increase attendance in recitations and studios. It would be interesting to observe whether the group of over-achievers would grow with increased attendance. Furthermore, it is unclear how a change in attendance would change the frequency with which attendance mea-

sures occurred in the walks. The frequency of attendance measures is linked with the amount of information, on the behavioral groups, attendance carries. Information from attendance might increase if the over-achievers and intervention students came to class more and all other students kept the same pattern of attendance. This would mean that the disparity between good attendance and bad widened, which would allow attendance scores to discriminate more effectively between the groups. And if better attendance caused a positive change in student behavior, then information from attendance would increase even more. However, this idea of increasing attendance to better determine behavioral groups has diminishing returns. For if every student were to have perfect attendance, then attendance measures would yield no information. We should keep in mind though that this work to detect behavioral groups is to serve the higher purpose of improving instruction and student understanding. So we should not let the desire to more effectively determine a student's behavioral group prevent us from trying to improve attendance or any other positive behaviors.

4.1.3 Extending Research to Other Courses

With the suggestion that attendance may improve behavior, it is worthwhile to consider how to get numbers from this metric to increase. In comparing math with other departments (outside the science, math, and engineering disciplines), we find that these disciplines see attendance as much more essential than we do in math. For example, in English courses, attendance is integral to understanding. Students discuss works of literature and interpretations of them. Hearing other people's opinions is not something that can be done alone. Whereas in math, material is presented to the class, and in turn the students can ask questions of the instructor. If a student fully understands the lesson and has no questions, many instructors would assent, perhaps grudgingly, to the claim that the student does not need to attend that lecture. This just illustrates the simple fact that, in other fields, student engagement is more critical to success in class. While the educational literature is full of examples of people reinventing the mathematics classroom to be more interac-

tive, we instead wish to focus on how our research, including Dr. Manspeaker's, might be used to further elucidate the differences between various courses.

First, it would be very interesting to see these data-mining techniques applied to courses in other departments. Perhaps the most useful courses to examine are those which have the largest enrollment. Dr. Manspeaker's study would need to be replicated first, revealing what behavioral groups exist in such a class. Though behavior may not be a proper way to distinguish various clusters of students. If we compare these results to an introductory English class, the prominence of attendance as an indicator of a student's group would be most interesting.

And of course, tracking students throughout numerous classes would undoubtedly provide a greater wealth of information. This is a rather ambitious goal as it requires cooperation across campus and significant institutional support. Still, the idea of being able to track students throughout their careers is indeed promising. It would be interesting to see how the student clusters across departments compared. It is unlikely that one set of groups would describe every course. But there surely would be similarities within departments and even between departments. And if we track students throughout their time at Kansas State University, the data might lead to clusters which explain a given student's behavior in each class for every semester.

An easier task would be to compare the Studio College Algebra course with the traditional College Algebra. The studio class has a weekly schedule only one large lecture but two small sessions — recitation and studio, whereas the traditional has two large lectures and one small recitation each week. This may even show differences in the importance of attendance in predicting groups. We speculate that the groups will be the same, though their proportions will differ. And further, attendance will likely carry more information since the traditional sections are more passive in their teaching style than those of the studio sections.

4.2 Long Runs

Recall that long runs were long stretches of time in which the Viterbi path assigns a student to the same behavioral group. We examined each student's longest run in a behavioral group, and found that a majority of the students started their longest run no later than the second week of class. We then examined long runs, which we took to be runs that lasted over at least ten assignments. The minimum length of ten was chosen because that length lasted a little over a week. The hope is that this is a small enough amount of time for a student to start exhibiting change in behavior.

4.2.1 Early Behavioral Patterns

The main revelation from the long runs analysis is how quickly people started their longest run, with a majority having started by the second week. This seems to inspire hope that there may yet be a way to quickly identify a student's behavioral group. While the analyses presented here do not offer a means of such speedy associations, the work does suggest starting to look for these behaviors early. It is unreasonable to expect that one could determine every student's group shortly after the second week. But being able to pin down some students who can be identified early is a start and clearly improves our teaching.

The difficulty of course is designing tests and experiments to tease out student behavior so soon in the term. Classroom practices like taking attendance, learning names of the students, and the behaviors of the instructors are among the few which are significant and rather simple to implement this early in the course. The importance of learning names and taking attendance is in how they affect student perceptions. Establishing a habit of taking roll, and doing so noticeably, indicates to students that any absence will be noticed by the instructor. It is more costly for students to miss a class where the teacher observes they are gone, than to miss a class where the teacher is unaware. If the instructor knows the student's name, then this effect is multiplied. Further the suggestion is that the professor sees attending class as important rather than merely recommended

or even optional. As for the teachers' behaviors, the literature survey samples some of the work done on studying the effects of actions — such as having a positive tone in the syllabus[3] and encouraging students to ask for help outside of class[6] — taken at the beginning of the course.

4.2.2 Fluctuations in Stability of Behavior

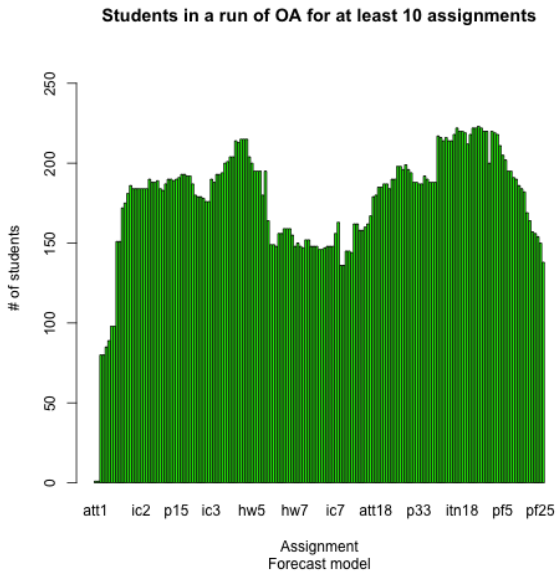


Figure 4.1: OA Runs

The other interesting observation comes from examining the over-achiever runs. The one with the biggest change in the number of people in a run occurs in Figure 4.1, just before the fifth homework. In a two week period between the sixth and eighth week of class, the number of students in over-achiever runs changes from around 175 to fewer than 100. Granted, there is a steep incline at the beginning and a steep decline at the end. The incline is simply due to the fact that there isn't enough information to classify many undergraduates at the start of the course. The decline at the end is the other significant change, which we will discuss in a moment.

The drop in over-achiever runs between the sixth and eighth week is rather interesting. The fifth homework covers inverse functions and composing functions; the sixth deals with translating graphs of functions; and the seventh tests knowledge of solving equations containing absolute values or radicals. One of the first interpretations is that this material is unfamiliar to the students, whereas the previous material on linear and quadratic functions is likely to be familiar to some of the students. So around this time in the semester, those students who were better prepared than the others no longer have the advantage of having seen the material. However, many students do not understand slopes and intercepts of lines, and are not able to interpret these concepts in a real-

world context. Evidence for this can be found in the early studios which deal with linear models. In particular, one studio has the students construct a linear model using data from the previous baseball season. They build a model of games won by each team from the difference between the number of runs the team scores against others and the number of runs the team allows against itself across the entire season. The meaning of the slope and of the y-intercept from the model, $\text{GAMES WON} = m(\text{RUNS SCORED} - \text{RUNS ALLOWED}) + b$, is opaque to the undergraduates. Students are unable to identify that the y-intercept for the model is roughly half the games in a season, and why this is reasonable. Nor are students able to understand that the slope is how many more games a team would win if the team scored one more run in the season. Students are often unable to go beyond the rote phrases, such as "slope is rise over run".

Around the time linear models are discussed, the students may also be exploring the aspects of college life outside the classroom. At this time in the semester, the world outside the classroom may yield some insights, as it is becoming even more significant. Students are becoming acquainted with the local bars, or reacquainted in some cases. First-year students have likely stumbled upon the idea that their parents are no longer nagging them about doing homework or getting to class on time. And those returning may be falling back into bad habits. Whatever the case, the easy-going attitude which accompanies a new term has gone. And concerns about life outside of class or just concerns about other courses cannot be ignored as influences on student behavior.

Still, the author believes that this is a good point in the course to start experimenting with teaching methodology. The events that happen outside College Algebra must be taken as more noise in our data. With large samples, we can safely ignore these variations to focus on whether or not a new presentation of quadratic graphs works better. The goal is to diminish the dramatic drop in the number of over-achiever runs. However the constraint is that our changes cannot radically affect the schedule or assignments of the course. For the data analysis techniques presented here assume that the training data and the testing data came from comparable semesters. Any drastic

changes to the course should be done incrementally, and major changes should be introduced one at a time. This will help to ensure a more reliable analysis of the data. Still, there are many experiments left to try and many more things to learn.

As for the final drop in number of over-achiever runs, it is harder to find new research questions here. One way to lose runs in Viterbi paths is for the transition probabilities to become more diffuse. In our case, it is more difficult for a student to remain an over-achiever from one final exam question to the next, than it would be for the same student to stay an over-achiever from homework 10 to the next assignment. Repeat this reasoning for each question on the final, and it becomes easy to see how we are losing runs. This might mean that our final exam is too difficult. An overly difficult assignment diminishes our ability to differentiate a student into one of Dr. Manspeaker's groups, and so makes transitions out of the over-achiever category more likely. It should be noted that the author is not suggesting that term exams should be easier for whatever reason, but is merely trying to suggest an interpretation of the analysis. There is also the possibility that the final has less power to distinguish students because the exam isn't challenging to the class. But this can be quickly discarded because it contradicts the plentiful ancedotes and personal experiences of instructors. Possibly if the semester could be exteneded so that there were homework and attendence points to be earned after the final, then the number of runs in over-achievers would not drop as much.

These claims of the final exam "spoiling" runs of over-achievers are difficult to verify. There aren't going to be students to interview after the term ends. It may be possible to interview some after they return from the break. But the lag introduces whole classes of new complications because the students will be in completely different emotional and mental states. One interpretation which might be testable is that homeworks after the exams "pull up" the number of runs. At first, this seems to violate the Markov property of these paths through a HMM, having a state depend upon any other state but the preceding state. But these paths are derived from the Viterbi algorithm, whose results vary greatly as the computation progresses through the model. The algorithm

builds up arrays of which is the most likely state transitioned into from each state at each time, and then starts to walk backwards from the most likely state at the end of the model given the final observation. If an extra time period is appended to the model, the new Viterbi path for this longer model will be completely different from the old path, provided the walk backwards in the new path moved to a different state than the old path ended on.

Chapter 5

Conclusion

Broadly, our intentions were to find how quickly we could identify to which behavioral group an arbitrary Studio College Algebra student belongs and to see if we could determine how that student's behavioral classification changed throughout the semester. We started by analyzing grades with a naive Bayesian approach. This analysis didn't bring us closer to achieving our two goals, but it did provide us with insight into which grades are predictive of Dr. Manspeaker's groups from one semester to the next. We obtained this by gauging how well various models predict the Fall 2010 group assignments, when build from Fall 2009 data. The quality of the prediction was determined by comparing them to the model forecasts when built from Fall 2010 data. The models varied in which assignments they included and excluded. As the number of models was too large, we used standard hill-climbing to find locally optimal models. These results showed that attendance and exam scores were found most often in the locally optimal models.

More importantly, we used the Bayesian models to generate transition probabilities for naive hidden Markov models. And the HMMs did provide us with clues to how our two goals might be achieved. The Viterbi algorithm produced, for each student, the most likely path through the behavioral states over the term. From these Viterbi paths, we focused on the times when a student was consistently in one group over a long period of time — what we called a "run" in that particular

group. We found that a majority of students started their longest run by the second week of class. And then by broadening our focus from longest run to long runs, we seem to have produced a data set which is much more amenable to analyzing changes between groups.

5.1 Identifying Behavior Fluctuation

Unfortunately, when we focused on long runs in Viterbi paths, i.e., a run whose length was at least ten, these filtered data did not cluster. So we were unable to easily pull out new groups of students based upon their filtered Viterbi paths. But if we aggregated the number of students in a run by the group in which the run occurred as in Figure 3.2, then we saw interesting changes over the semester. We have already examined the fluctuations in the over-achiever group in section 4.2.2. As for the rote memorizers and smart slackers, their numbers are too small to draw any reasonable conclusions from. But the intervention and employee groups are large enough for us make some educated guesses from.

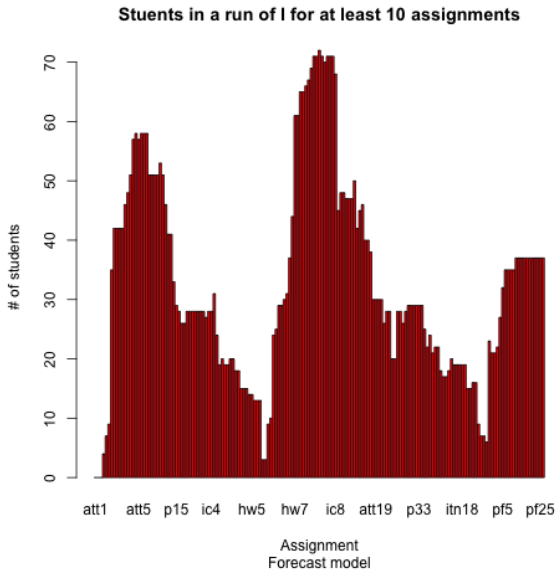


Figure 5.1: I Runs

The intervention group has two major peaks. The first one occurs about one week before the first exam and the size rises to slightly less than 60 students, and falls off right after the same exam to under 30. The second peak happens at one of the the second exam questions, and the group size falls from over 70 students to under 30 by the third exam. The first peak of intervention runs may be due to confusing this behavioral group with over-achievers. The reader may recall that both groups do well on homeworks and attendance,

but become distinguishable on exams, with over-achievers doing well on exams and intervention members doing poorly. So it is reasonable to see the initial intervention peak fall just before or at the beginning of an exam. The second peak occurs at one of the second exam questions and levels off by the third exam. One interpretation of this is that students who performed poorly on the second are incorrectly classified in the intervention group. Then after the exam, those who do not do well on the homework would be weeded out of the intervention group. Of course, an alternative interpretation is that student behaviors are indeed changing, and this change is reflected in the data.

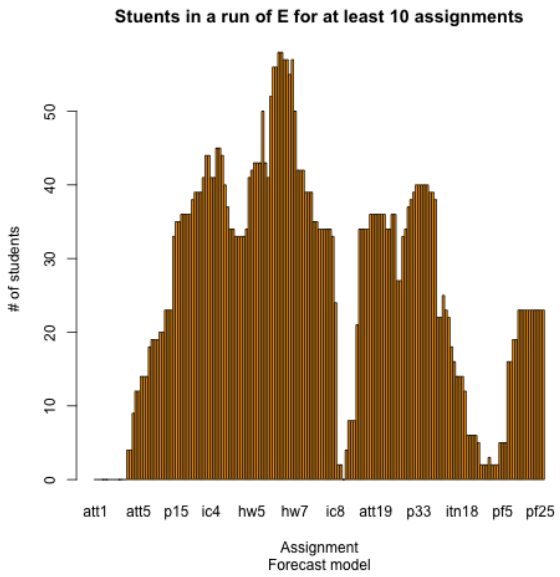


Figure 5.2: E Runs

The employee group has three crests. The first occurs at the end of the first exam and comprises 45 students. The number then falls slightly over the next week before it reaches the second crescendo about one week before the second exam. This falls slowly over the questions of the second exam and then crashes to zero students afterwards. Then the number of employees climbs quickly upward to reach its final maximum of roughly 40 students during the third exam. The initial rise in numbers of employees takes about one week to commence. It seems reasonable that it should take some time to identify employees, as it may take awhile to distinguish their persistent apathy from a troubled start of a semester. But the fall in employee numbers towards the end of the semester admits a more interesting interpretation. It may be possible that many employees start to work hard at the end of the term, in the hopes of improving their grades. To test this claim, we may try to interview employees around this time. However, Dr. Manspeaker found that employees, in addition to having poor class attendance, also

have poor attendance to interviews which they agreed to do. So this claim may be one of the more difficult ones to verify.

5.2 Further Research

In chapter 4, we talked about some ways that this research can be extended. The focus of this section is to emphasize, what we feel to be the more important suggestions and to discuss more direct extensions to this work. The first few suggestions are rather obvious: run the Bayesian and HMM analysis on the Spring semester data and do these analyses, as well as the behavioral group analysis, on the traditional College Algebra course. Dr. Manspeaker found that her groups from the Fall semester of Studio College Algebra also carried over into the Spring term as well. But there are important differences in these two populations of students: the Fall term has many more first-year students and, in general, more students taking College Algebra for the first time; whereas the Spring term has a larger number of students who are retaking College Algebra, and a smaller number of first-year students. Whether the long run analysis shows that Spring is a clearly different population from Fall or the two are almost interchangeable, either result would be fascinating.

As for the traditional College Algebra course, this will take more time, as Dr. Manspeaker's analysis needs to be repeated in full. It may be sufficient to skip the interviews and assume that the behavioral groups for traditional are the same as they are for studio. As this was the most time-consuming part of her work, its omission would speed up the process by years. In the traditional class, we would expect to see more rote memorizers, as Dr. Manspeaker has informed the advisers of her work and indicated that these types of undergraduates should be in traditional rather than Studio College Algebra. So we can expect to get more information about the fluctuation of students into and out of this group. The traditional course, with its different population distribution and altered focused, is likely to display new patterns of change in student behavior. By comparing

the results of these two courses, we may be able to uncover a bit of how students' behaviors differ in these two classes.

One of the ways which we could directly continue this research is to pay close attention to the beginning of the term. First, we might focus on student attendance early in the semester. We have seen that attendance is likely to be a good predictor of the Bayesian groups, and that attendance is an easy metric to watch during the critical first few weeks, before a majority of students start to settle into a long run. It seems that gathering the early attendance is a worthwhile endeavor. As this data is simple to gather and monitor, a lack of findings from the data would not be so costly. And it is possible that we might find a way to predict some of the students' early behaviors. Being able to identify students this early would be of great benefit towards trying to retain students and help them be more successful in College Algebra.

A more time-intensive task would be to interview students soon after the first few weeks of class. This is an attempt to answer the question of how soon do students form their beliefs and behaviors about math. While trying to answer the question herself, Dr. Manspeaker added questions to the university math placement exam which inquired about these attitudes. However, she found that she could not associate the students' answers to these questions with her groups. We came to believe that students do not have their behaviors set prior to arriving at college, but do have patterns of acting by the first exam. The work presented here using HMMs suggests that beliefs and behaviors stabilize within the first two weeks for a majority of students. By interviewing students early in the semester, we may confirm whether or not this prediction is correct.

Finally, it may be possible that further exploration of the data and the parameters of the analysis will turn up interesting results. For example, each problem from all the exams appears as an "assignment", or a measure in the language of HMMs. The totals for the exams are also measures. This means that a week's worth of assignments consists of a little fewer measures than the three regular exams, and the final exam is at least two if not three week's worth of measures. It could be that this configuration overestimates the importance of exams by allotting them too many measures

in the HMM analysis. Additionally, we could adjust the minimum length of the runs in the Viterbi paths. As mentioned earlier, we chose ten measures because it was roughly one week's worth of assignments. However, it may be that longer runs might do a better job of filtering out the group "signals" from the "noise" of moving between groups. In our testing, we could not find a different minimum length which would produce runs which would cluster more easily. And so we decided to stick with our initial choice of ten measures as a minimum length for a long run.

Bibliography

[1] Marilyn Carlson, Michael Oehrtman, and Nicole Engelke. The precalculus concept assessment: A tool for assessing students' reasoning abilities and understandings. *Cognition and Instruction*, 28(2):113–145, 2010.

[2] Andrew M. Fraser. *Hidden Markov Models and Dynamical Systems*. SIAM, 2008.

[3] R. J. Harnish and K. Bridges. Effect of syllabus tone: Students' perception of instructor and course. *Social Psychology of Education: An International Journal*, 14(3):319–330, 2011.

[4] A. D. Hermann, D. A. Foster, and E. E. Hardin. Does the first week of class matter? a quasi-experimental investigation of student satisfaction. *Teaching of Psychology*, 37(2):79–84, 2010.

[5] Rachel Manspeaker. *Using Data Mining to Differentiate Instruction in College Algebra*. PhD thesis, Kansas State University, 2011.

[6] R. M. Perrine, J. Lisle, and D. L. Tucker. Effects of a syllabus offer of help, student age, and class size, on college students' willingness to seek support from faculty. *Journal of Experimental Education*, 64(1):41–52, 1995.

[7] Lawrence R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, February 1989.

[8] B. K. Saville, T. E. Zinn, A. R. Brown, and K. A. Marchuk. Syllabus detail and students' perceptions of teacher effectiveness. *Teaching of Psychology*, 37(3):186–189, 2010.

Appendix A

Sample R Session

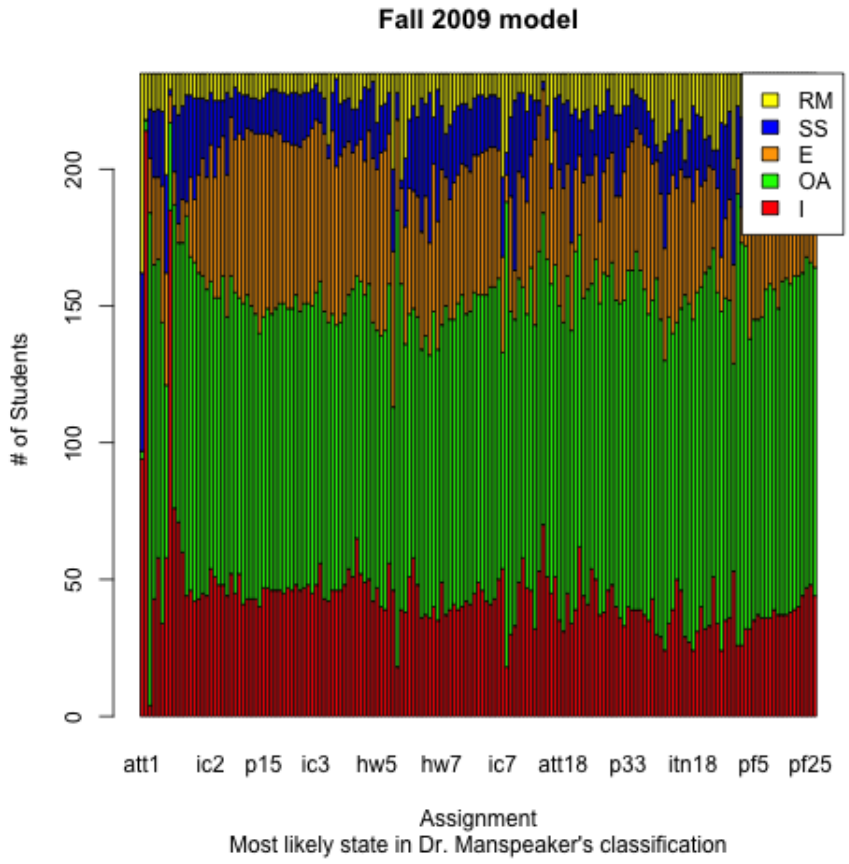
```
rachParamsF09 <- prepareBayesParams(m100f09["gp"], dataF09)
rachParamsF10 <- prepareBayesParams(m100f10["gp"], dataF10)
rachBayesF09 <- bayes(rachParamsF09, dataF09)
rachBayes <- bayes(rachParamsF09, dataF10)
rachBayesF10 <- bayes(rachParamsF10, dataF10)
rachGrpsF09 <- apply(rachBayesF09, c(1,2),
                     function(x) { max.col(matrix(x, nrow=1)) })
rachGrps <- apply(rachBayes, c(1,2),
                  function(x) { max.col(matrix(x, nrow=1)) })
rachGrpsF10 <- apply(rachBayesF10, c(1,2),
                     function(x) { max.col(matrix(x, nrow=1)) })
pi1F09 <- prop.table(table(m100f09["gp"]))
pi1F10 <- prop.table(table(m100f10["gp"]))
rachHmmF09 <- prepareHMMParams(rachParamsF09, dataF09)
rachHmmF10 <- prepareHMMParams(rachParamsF10, dataF10)
rachVitF09 <- viterbiAlgo(dataF09, pi1F09,
                          rachHmmF09[[1]], rachHmmF09[[2]])
rachVit <- viterbiAlgo(dataF10, pi1F09,
                       rachHmmF09[[1]], rachHmmF09[[2]])
rachVitF10 <- viterbiAlgo(dataF10, pi1F10,
                          rachHmmF10[[1]], rachHmmF10[[2]])
rachLRF09 <- filterRuns(rachVitF09[[2]], 10)
rachLR <- filterRuns(rachVit[[2]], 10)
rachLRF10 <- filterRuns(rachVitF10[[2]], 10)
```

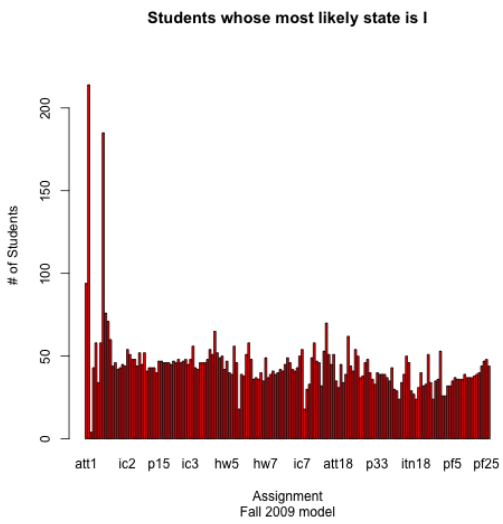
Appendix B

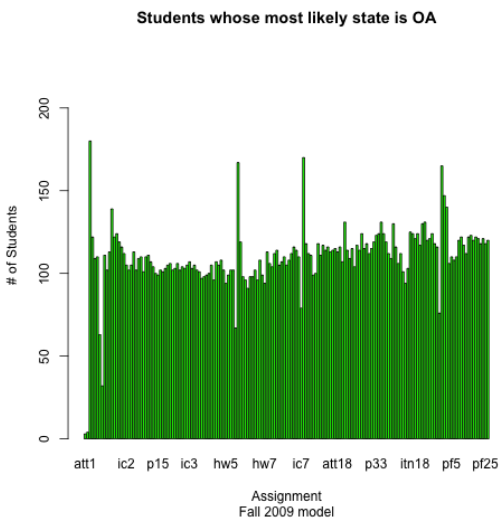
Graphs

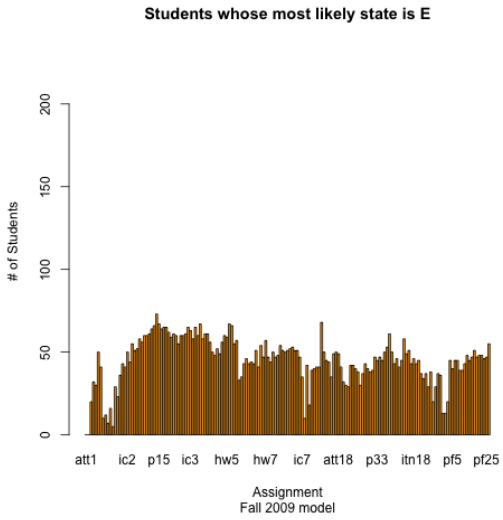
B.1 2009 Fall

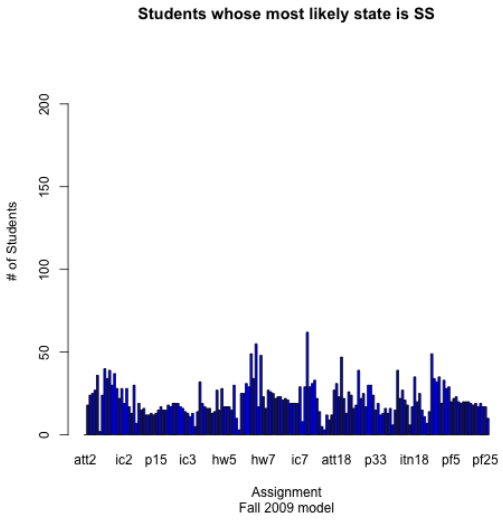
B.1.1 Bayesian Model

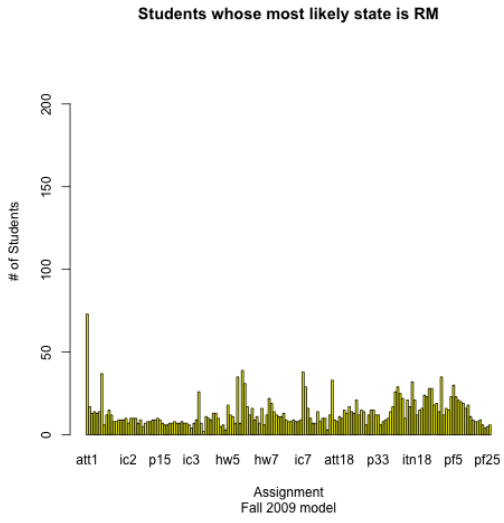



Students whose most likely state is I
# of Students
0
50
100
150
200
att1 ic2 p15 ic3 hw5 hw7 ic7 att18 p33 itn18 pf5 pf25
Assignment
Fall 2009 model

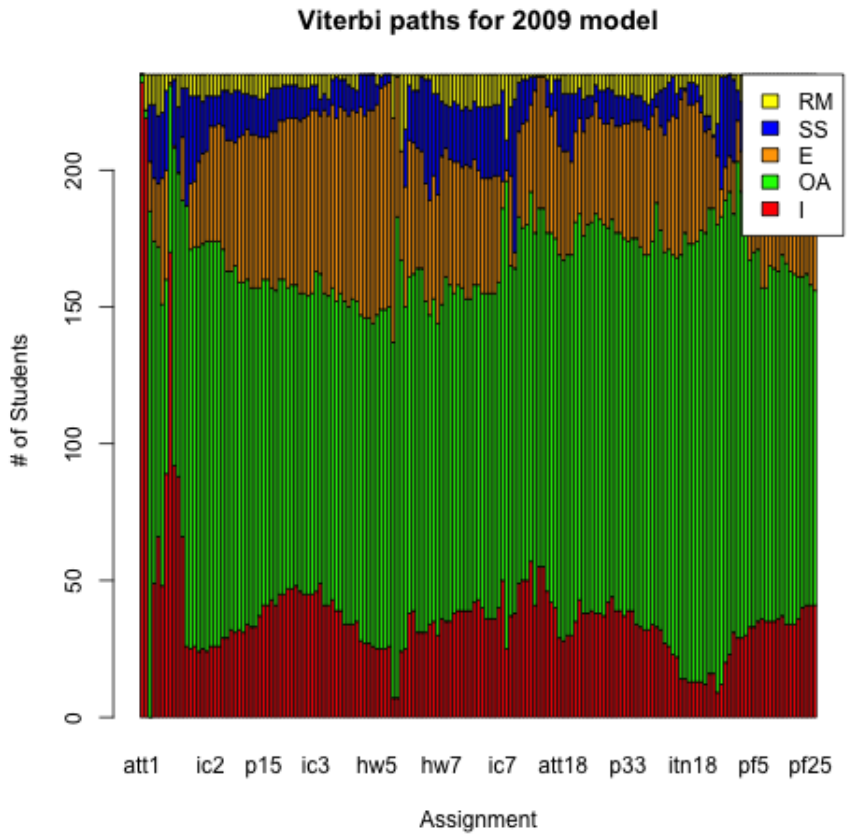

Students whose most likely state is OA
# of Students
0
50
100
150
200
att1 ic2 p15 ic3 hw5 hw7 ic7 att18 p33 itn18 pf5 pf25
Assignment
Fall 2009 model

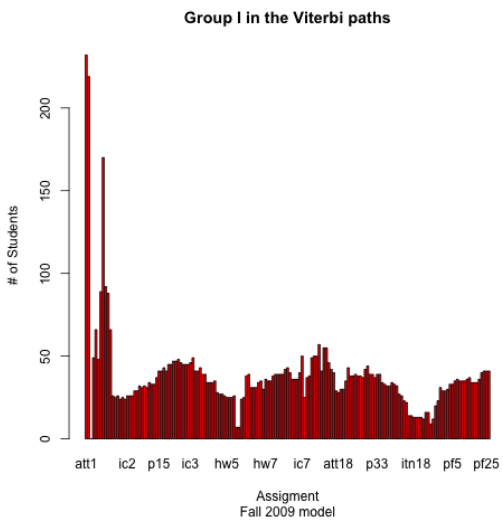

Students whose most likely state is E
# of Students
0
50
100
150
200
att1 ic2 p15 ic3 hw5 hw7 ic7 att18 p33 itn18 pf5 pf25
Assignment
Fall 2009 model

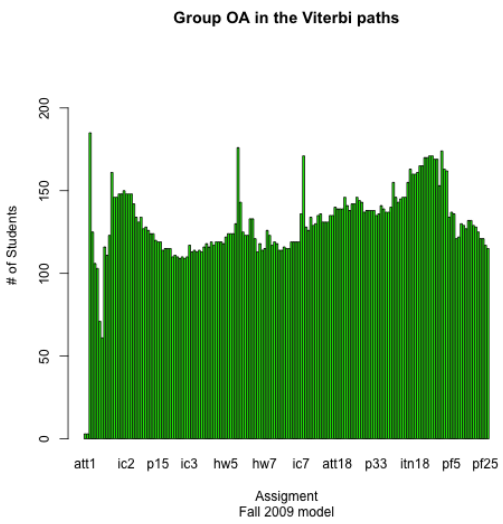

Students whose most likely state is SS
# of Students
0
50
100
150
200
att2 ic2 p15 ic3 hw5 hw7 ic7 att18 p33 itn18 pf5 pf25
Assignment
Fall 2009 model

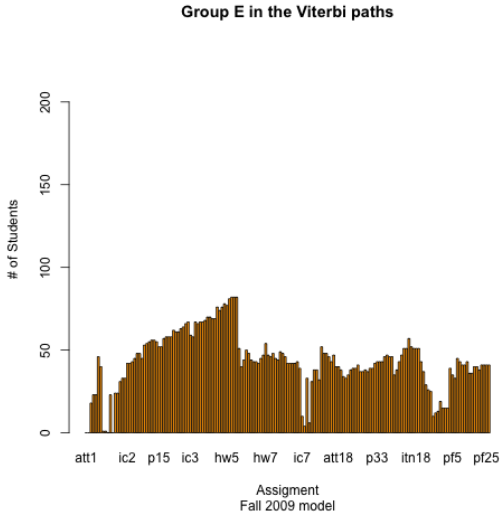

Students whose most likely state is RM
# of Students
0
50
100
150
200
att1 ic2 p15 ic3 hw5 hw7 ic7 att18 p33 itn18 pf5 pf25
Assignment
Fall 2009 model

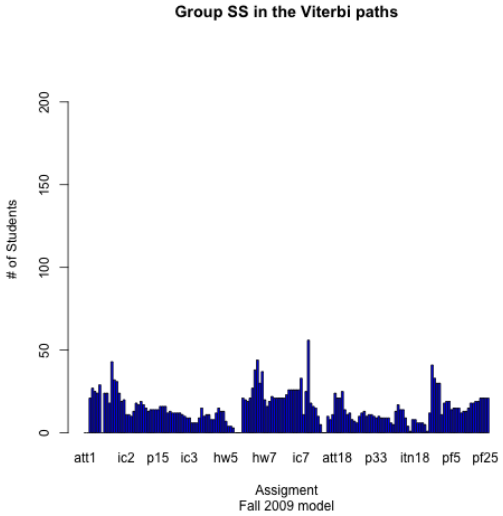
B.1.2 Viterbi Paths in Hidden Markov Model

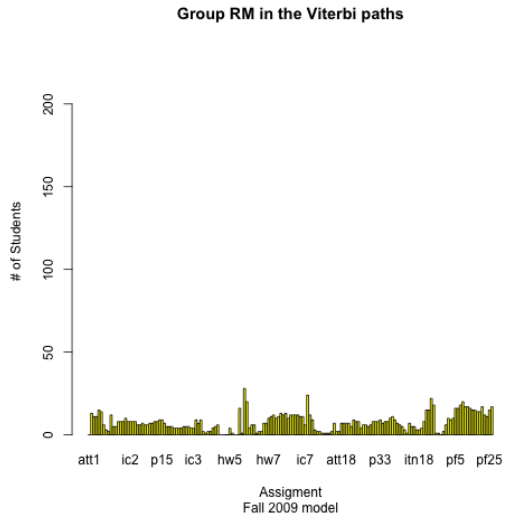



Group I in the Viterbi paths
# of Students
0
50
100
150
200
att1 ic2 p15 ic3 hw5 hw7 ic7 att18 p33 itn18 pf5 pf25
Assigment
Fall 2009 model

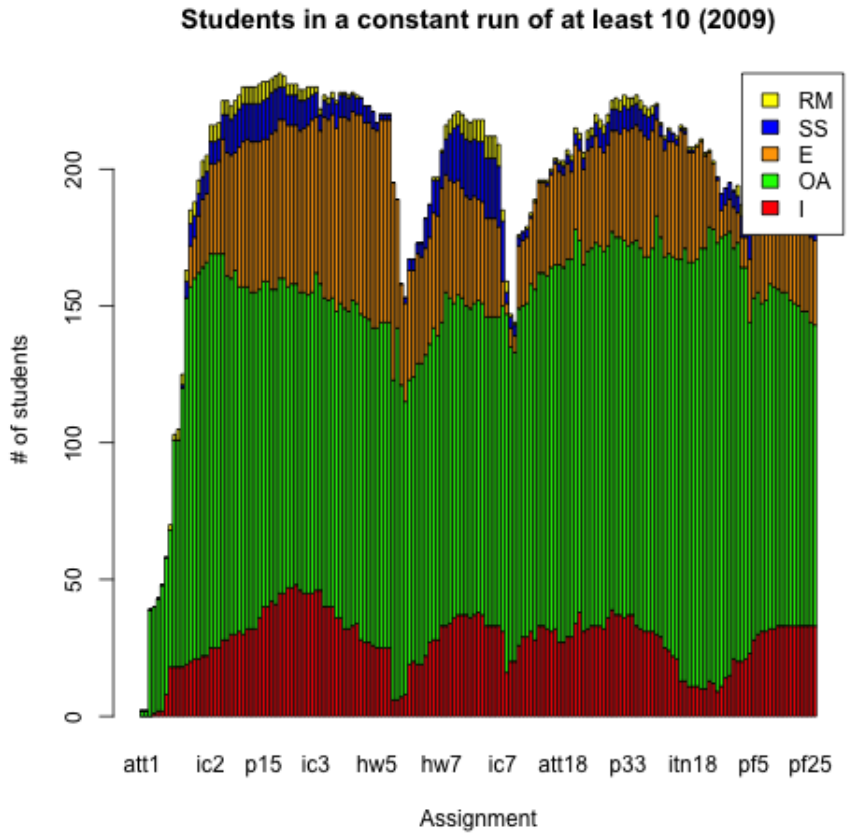

Group OA in the Viterbi paths
# of Students
0
50
100
150
200
att1 ic2 p15 ic3 hw5 hw7 ic7 att18 p33 itn18 pf5 pf25
Assigment
Fall 2009 model

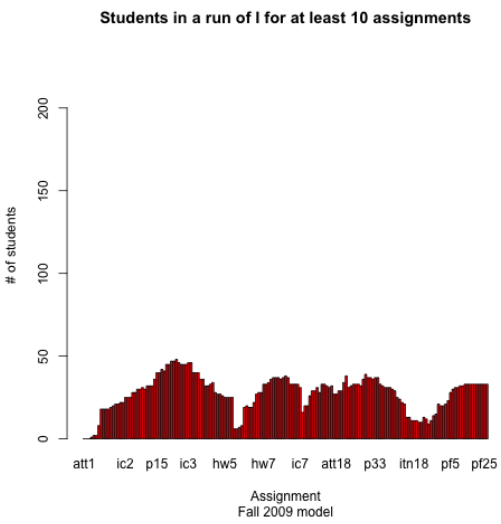

Group E in the Viterbi paths
# of Students
0
50
100
150
200
att1 ic2 p15 ic3 hw5 hw7 ic7 att18 p33 itn18 pf5 pf25
Assigment
Fall 2009 model

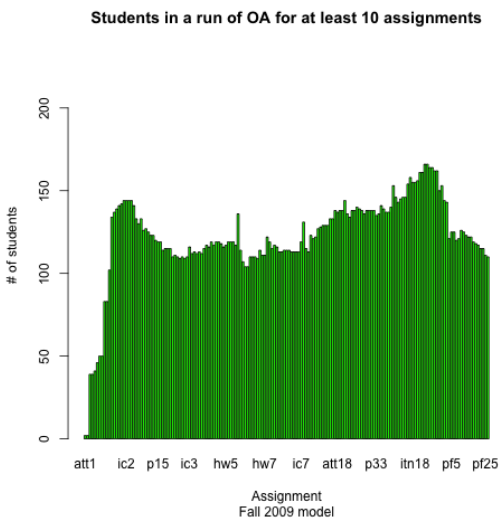

Group SS in the Viterbi paths
# of Students
0
50
100
150
200
att1 ic2 p15 ic3 hw5 hw7 ic7 att18 p33 itn18 pf5 pf25
Assigment
Fall 2009 model

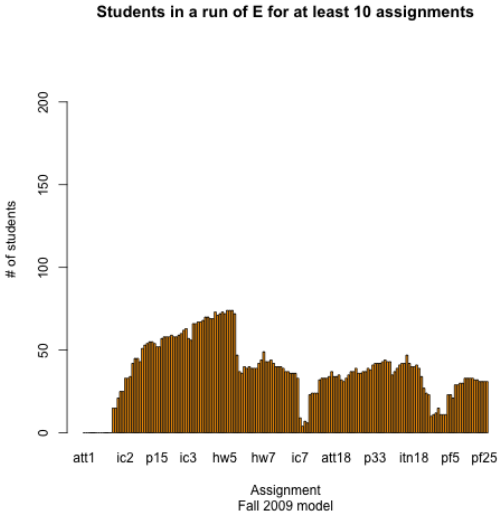

Group RM in the Viterbi paths
# of Students
0
50
100
150
200
att1 ic2 p15 ic3 hw5 hw7 ic7 att18 p33 itn18 pf5 pf25
Assigment
Fall 2009 model

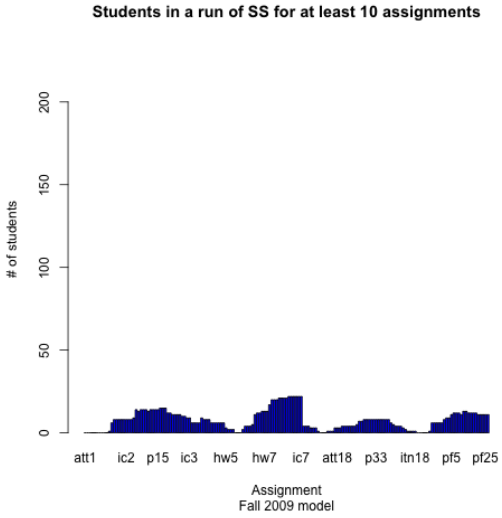
B.1.3 Long Runs in Viterbi Paths

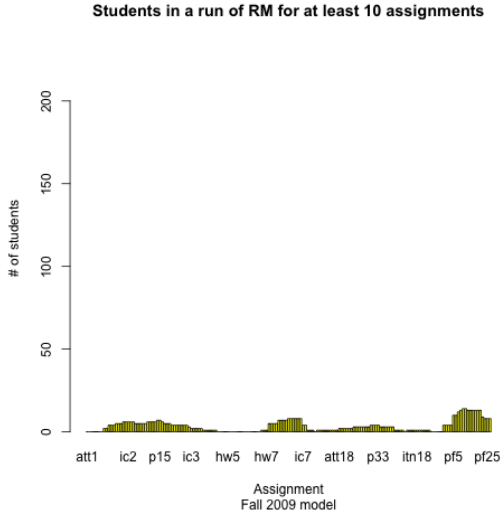



Students in a run of I for at least 10 assignments
# of students
0
50
100
150
200
att1 ic2 p15 ic3 hw5 hw7 ic7 att18 p33 itn18 pf5 pf25
Assignment
Fall 2009 model


Students in a run of OA for at least 10 assignments
# of students
0
50
100
150
200
att1 ic2 p15 ic3 hw5 hw7 ic7 att18 p33 itn18 pf5 pf25
Assignment
Fall 2009 model

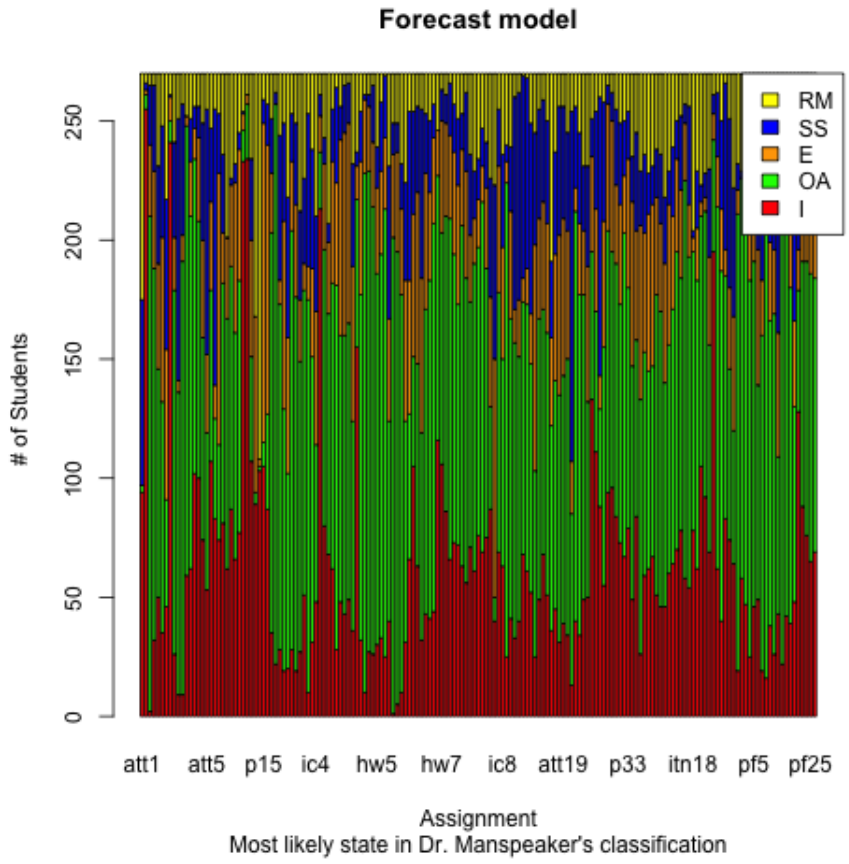

Students in a run of E for at least 10 assignments
# of students
0
50
100
150
200
att1 ic2 p15 ic3 hw5 hw7 ic7 att18 p33 itn18 pf5 pf25
Assignment
Fall 2009 model

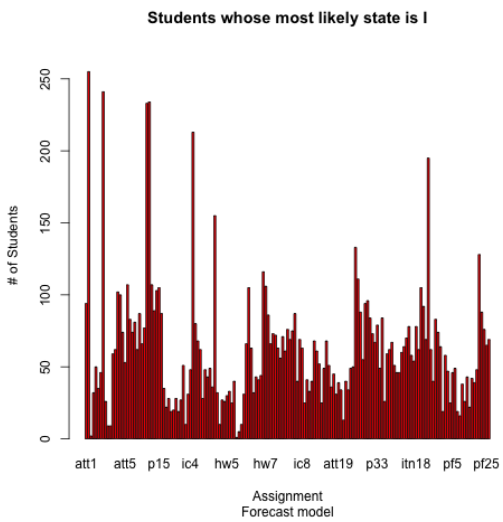

Students in a run of SS for at least 10 assignments
# of students
0
50
100
150
200
att1 ic2 p15 ic3 hw5 hw7 ic7 att18 p33 itn18 pf5 pf25
Assignment
Fall 2009 model

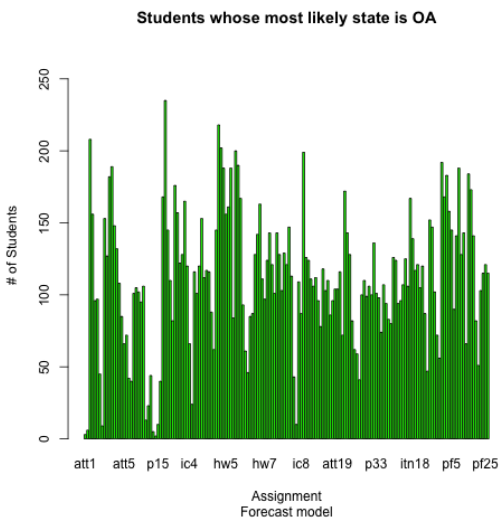

Students in a run of RM for at least 10 assignments
# of students
0
50
100
150
200
att1 ic2 p15 ic3 hw5 hw7 ic7 att18 p33 itn18 pf5 pf25
Assignment
Fall 2009 model

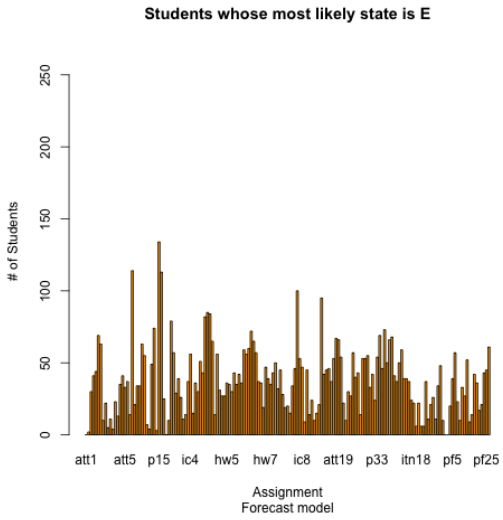
B.2 2009 Fall Predicting 2010 Fall

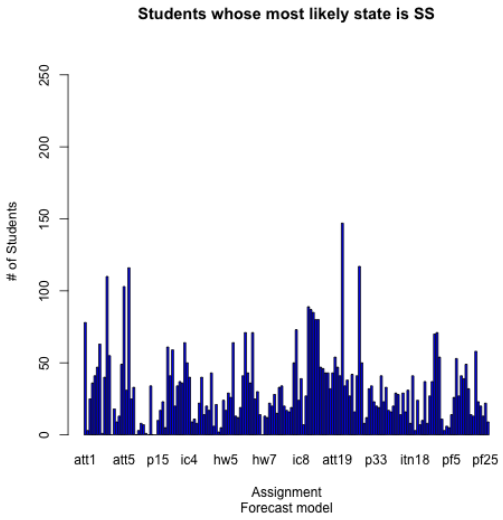
B.2.1 Bayesian Model

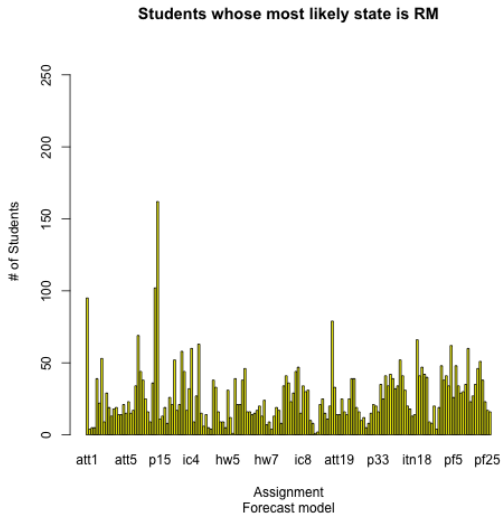



Students whose most likely state is I
# of Students
0
50
100
150
200
250
att1 att5 p15 ic4 hw5 hw7 ic8 att19 p33 itn18 pf5 pf25
Assignment
Forecast model

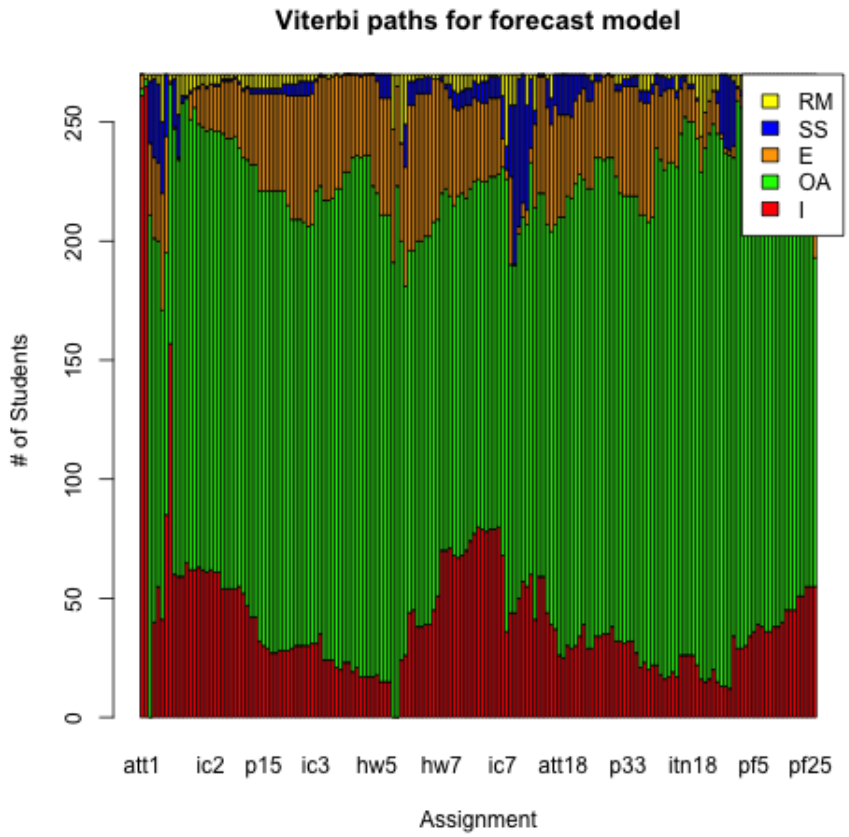

Students whose most likely state is OA
# of Students
0
50
100
150
200
250
att1 att5 p15 ic4 hw5 hw7 ic8 att19 p33 itn18 pf5 pf25
Assignment
Forecast model

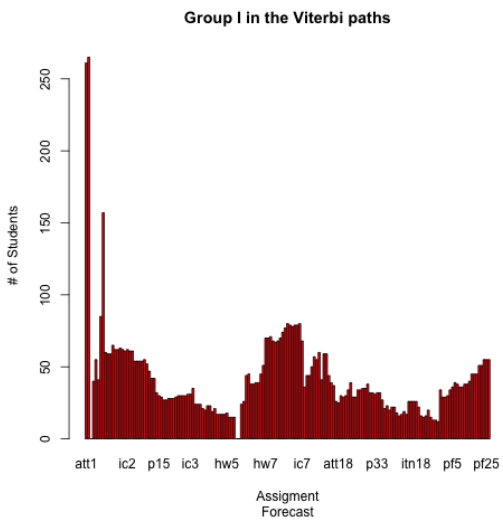

Students whose most likely state is E
# of Students
0
50
100
150
200
250
att1 att5 p15 ic4 hw5 hw7 ic8 att19 p33 itn18 pf5 pf25
Assignment
Forecast model

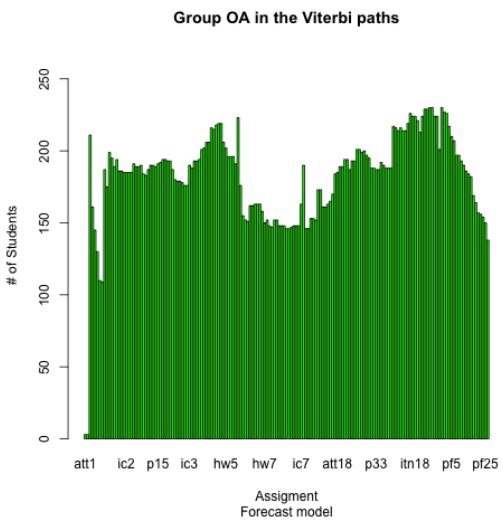

Students whose most likely state is SS
# of Students
0
50
100
150
200
250
att1 att5 p15 ic4 hw5 hw7 ic8 att19 p33 itn18 pf5 pf25
Assignment
Forecast model

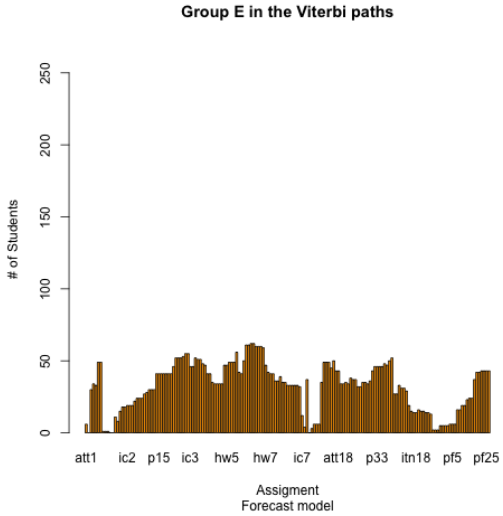

Students whose most likely state is RM
# of Students
0
50
100
150
200
250
att1 att5 p15 ic4 hw5 hw7 ic8 att19 p33 itn18 pf5 pf25
Assignment
Forecast model

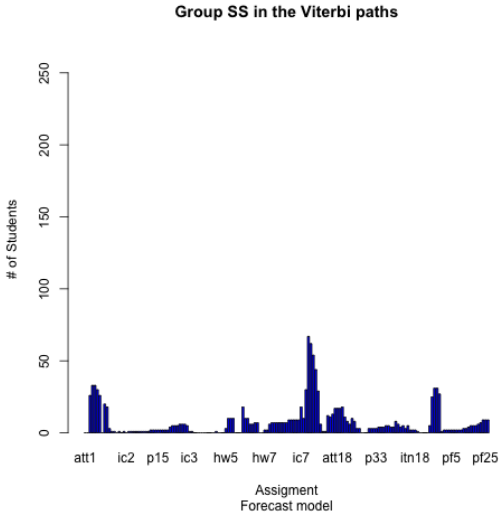
B.2.2 Viterbi Paths in Hidden Markov Model

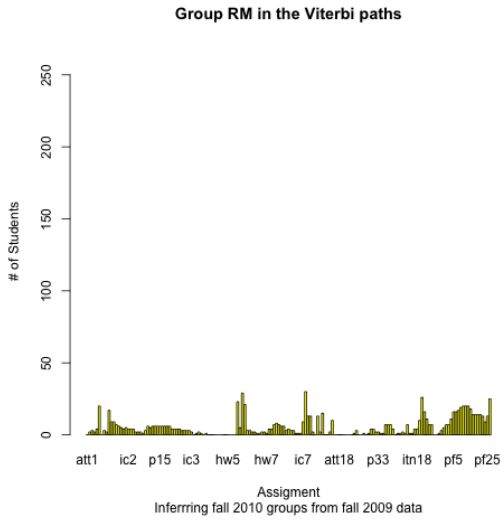



Group I in the Viterbi paths
# of Students
0
50
100
150
200
250
att1 ic2 p15 ic3 hw5 hw7 ic7 att18 p33 itn18 pf5 pf25
Assigment
Forecast

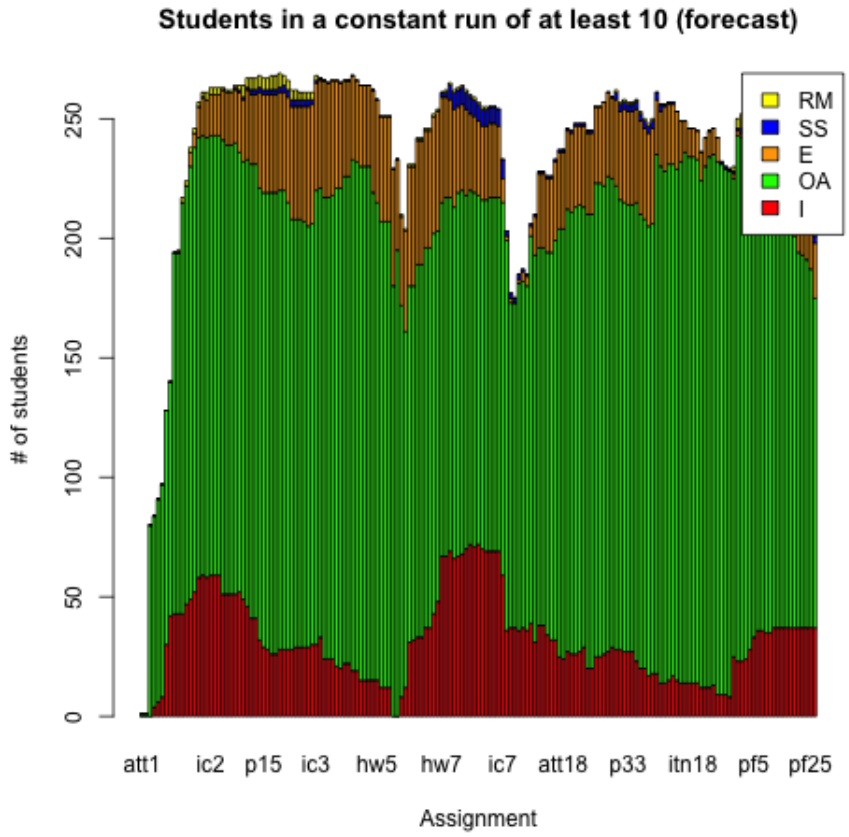

Group OA in the Viterbi paths
# of Students
0
50
100
150
200
250
att1 ic2 p15 ic3 hw5 hw7 ic7 att18 p33 itn18 pf5 pf25
Assigment
Forecast model

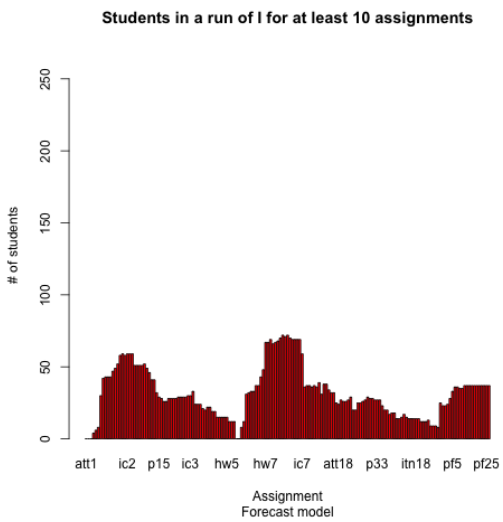

Group E in the Viterbi paths
# of Students
0
50
100
150
200
250
att1 ic2 p15 ic3 hw5 hw7 ic7 att18 p33 itn18 pf5 pf25
Assigment
Forecast model

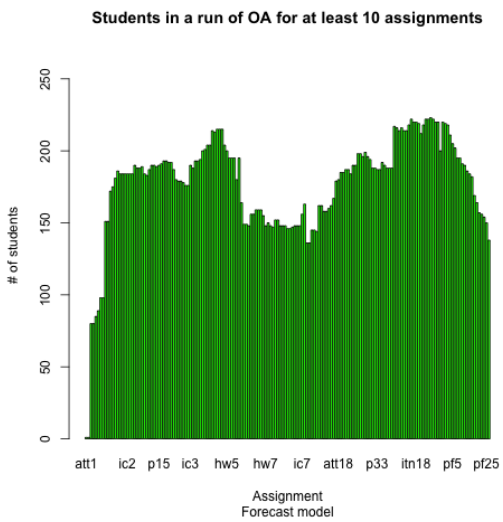

Group SS in the Viterbi paths
# of Students
0
50
100
150
200
250
att1 ic2 p15 ic3 hw5 hw7 ic7 att18 p33 itn18 pf5 pf25
Assigment
Forecast model

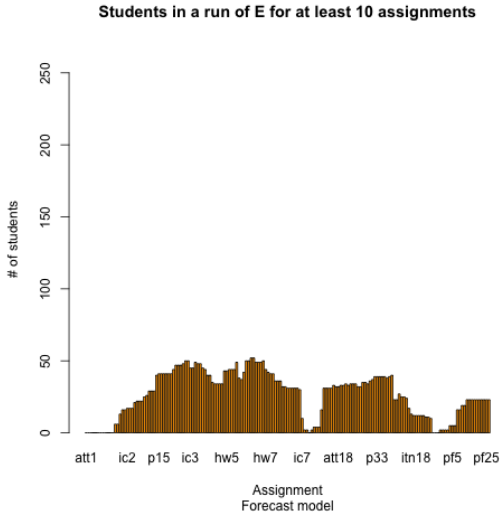

Group RM in the Viterbi paths
# of Students
0
50
100
150
200
250
att1 ic2 p15 ic3 hw5 hw7 ic7 att18 p33 itn18 pf5 pf25
Assigment
Inferrring fall 2010 groups from fall 2009 data

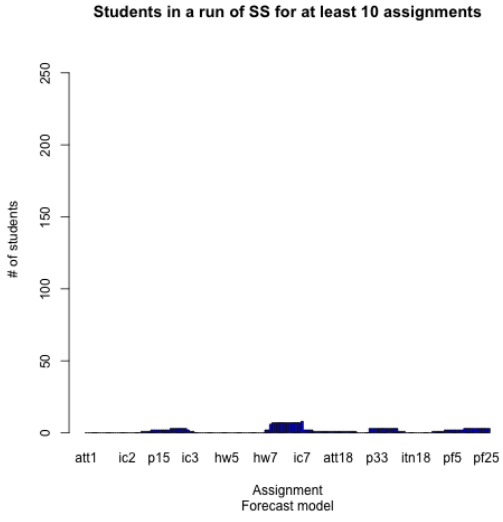
B.2.3 Long Runs in Viterbi Paths

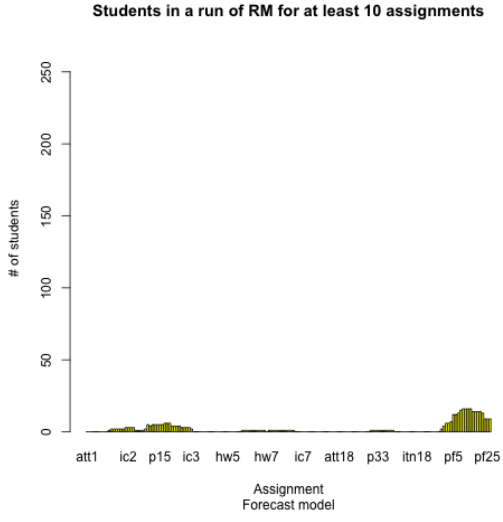



Students in a run of I for at least 10 assignments
# of students
0
50
100
150
200
250
att1 ic2 p15 ic3 hw5 hw7 ic7 att18 p33 itn18 pf5 pf25
Assignment
Forecast model


Students in a run of OA for at least 10 assignments
# of students
0
50
100
150
200
250
att1 ic2 p15 ic3 hw5 hw7 ic7 att18 p33 itn18 pf5 pf25
Assignment
Forecast model

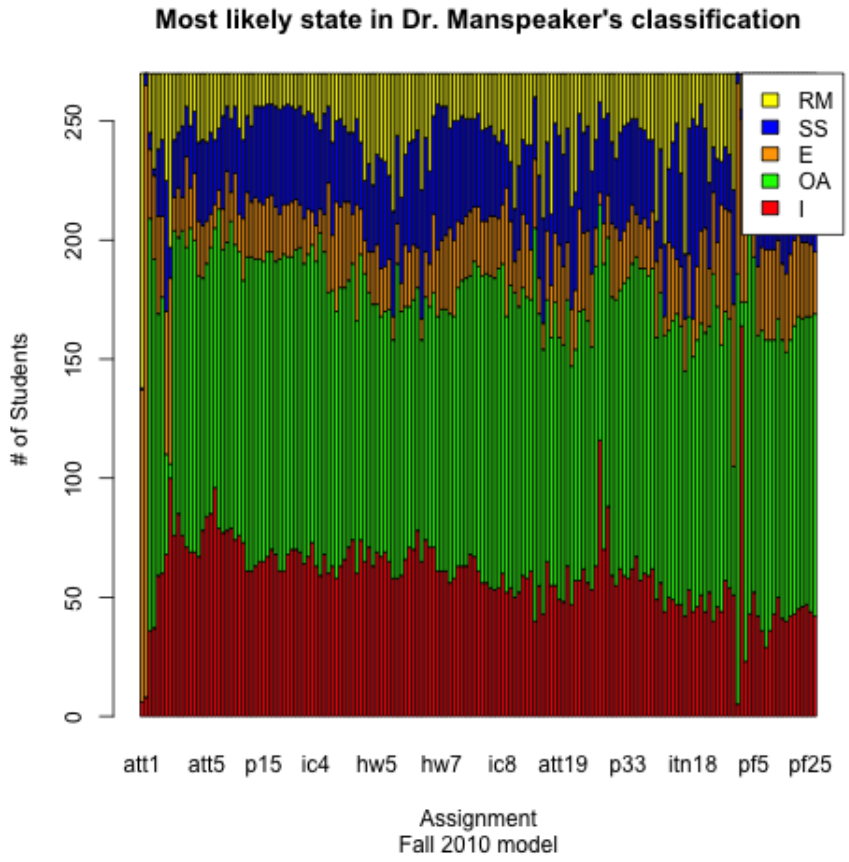

Students in a run of E for at least 10 assignments
# of students
0
50
100
150
200
250
att1 ic2 p15 ic3 hw5 hw7 ic7 att18 p33 itn18 pf5 pf25
Assignment
Forecast model

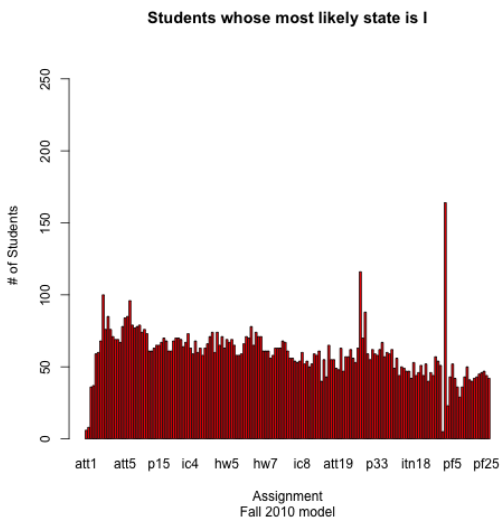

Students in a run of SS for at least 10 assignments
# of students
0
50
100
150
200
250
att1 ic2 p15 ic3 hw5 hw7 ic7 att18 p33 itn18 pf5 pf25
Assignment
Forecast model

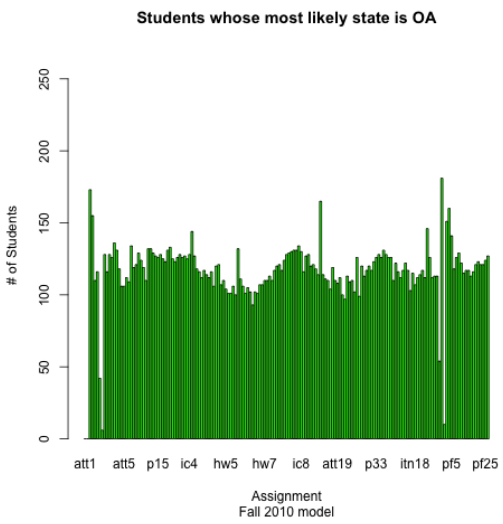

Students in a run of RM for at least 10 assignments
# of students
0
50
100
150
200
250
att1 ic2 p15 ic3 hw5 hw7 ic7 att18 p33 itn18 pf5 pf25
Assignment
Forecast model

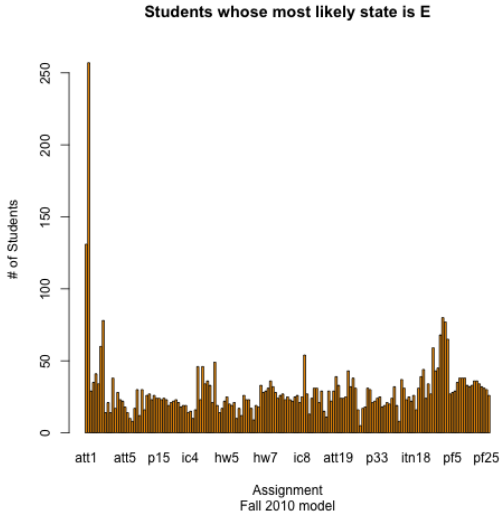
B.3 2010 Fall

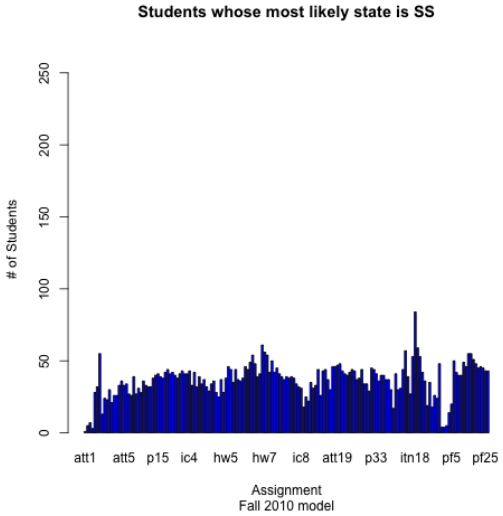
B.3.1 Bayesian Model

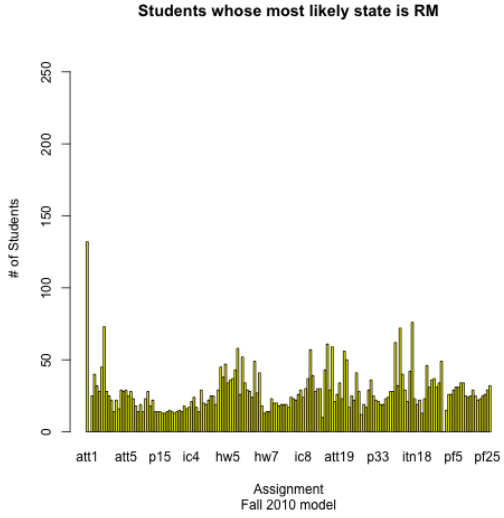



Students whose most likely state is I
# of Students
0
50
100
150
200
250
att1 att5 p15 ic4 hw5 hw7 ic8 att19 p33 itn18 pf5 pf25
Assignment
Fall 2010 model

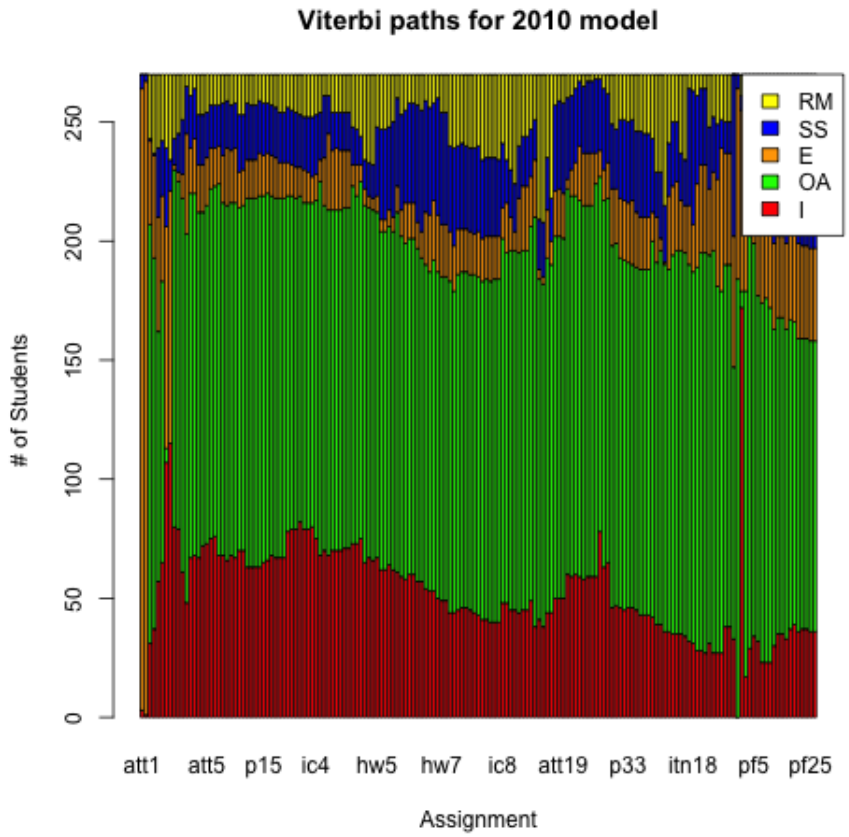

Students whose most likely state is OA
# of Students
0
50
100
150
200
250
att1 att5 p15 ic4 hw5 hw7 ic8 att19 p33 itn18 pf5 pf25
Assignment
Fall 2010 model

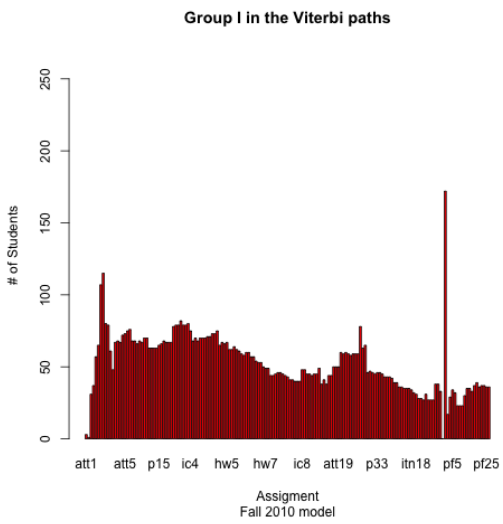

Students whose most likely state is E
# of Students
0
50
100
150
200
250
att1 att5 p15 ic4 hw5 hw7 ic8 att19 p33 itn18 pf5 pf25
Assignment
Fall 2010 model

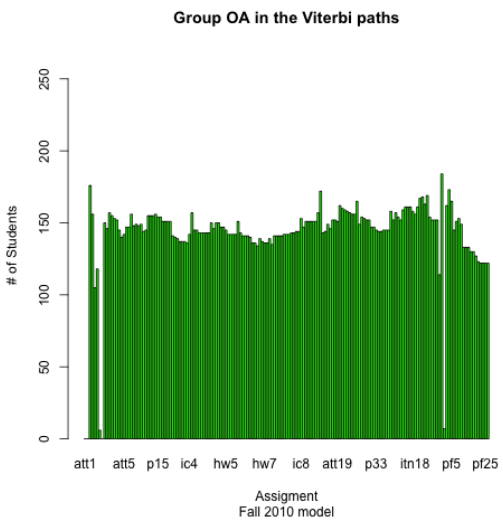

Students whose most likely state is SS
# of Students
0
50
100
150
200
250
att1 att5 p15 ic4 hw5 hw7 ic8 att19 p33 itn18 pf5 pf25
Assignment
Fall 2010 model

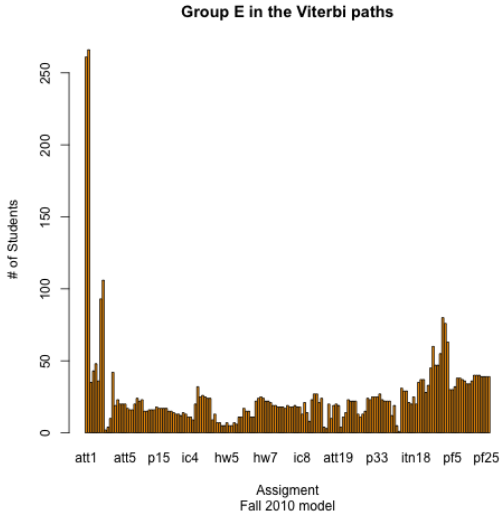

Students whose most likely state is RM
# of Students
0
50
100
150
200
250
att1 att5 p15 ic4 hw5 hw7 ic8 att19 p33 itn18 pf5 pf25
Assignment
Fall 2010 model

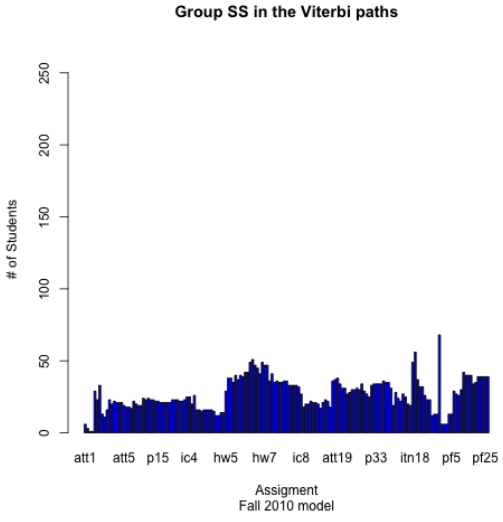
B.3.2 Viterbi Paths in Hidden Markov Model

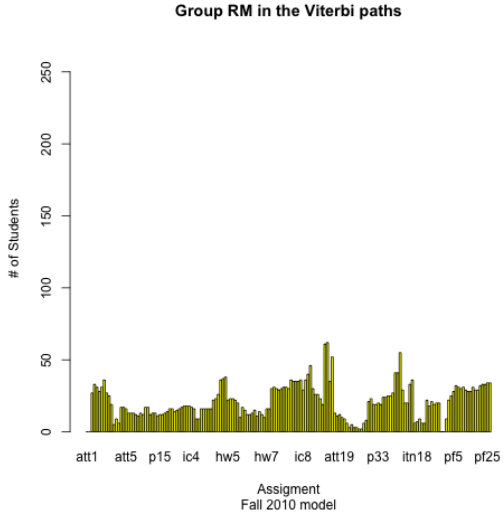



Group I in the Viterbi paths
# of Students
0
50
100
150
200
250
att1 att5 p15 ic4 hw5 hw7 ic8 att19 p33 itn18 pf5 pf25
Assigment
Fall 2010 model

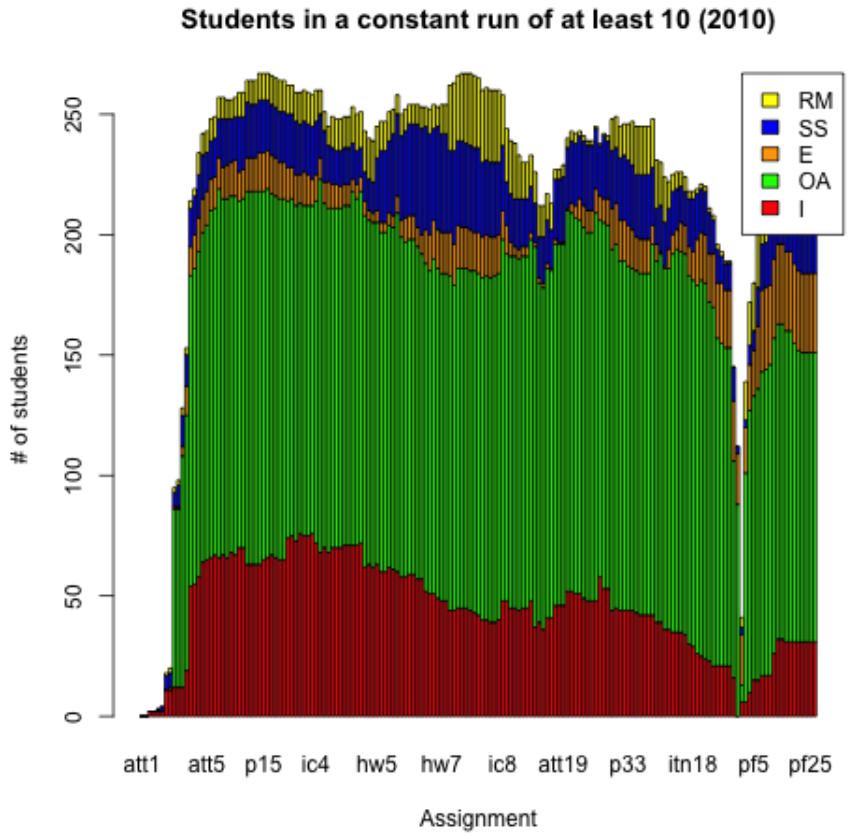

Group OA in the Viterbi paths
# of Students
0
50
100
150
200
250
att1 att5 p15 ic4 hw5 hw7 ic8 att19 p33 itn18 pf5 pf25
Assigment
Fall 2010 model

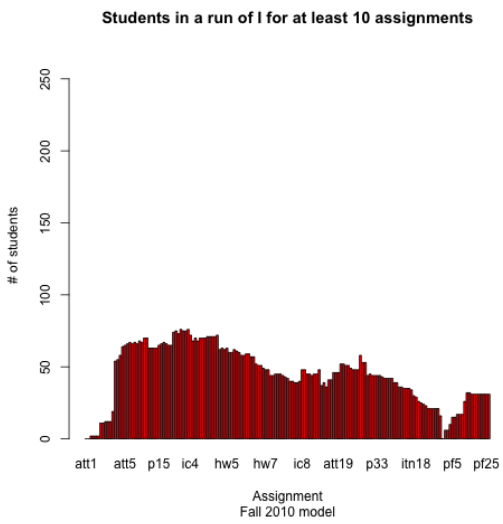

Group E in the Viterbi paths
# of Students
0
50
100
150
200
250
att1 att5 p15 ic4 hw5 hw7 ic8 att19 p33 itn18 pf5 pf25
Assigment
Fall 2010 model

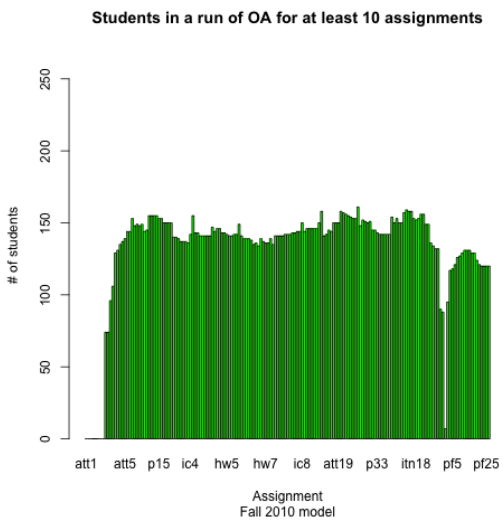

Group SS in the Viterbi paths
# of Students
0
50
100
150
200
250
att1 att5 p15 ic4 hw5 hw7 ic8 att19 p33 itn18 pf5 pf25
Assigment
Fall 2010 model

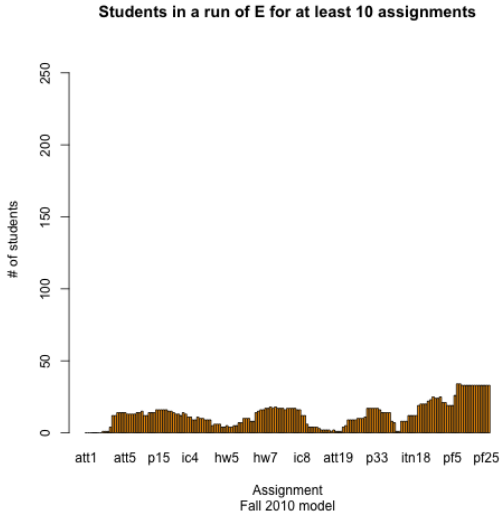

Group RM in the Viterbi paths
# of Students
0
50
100
150
200
250
att1 att5 p15 ic4 hw5 hw7 ic8 att19 p33 itn18 pf5 pf25
Assigment
Fall 2010 model

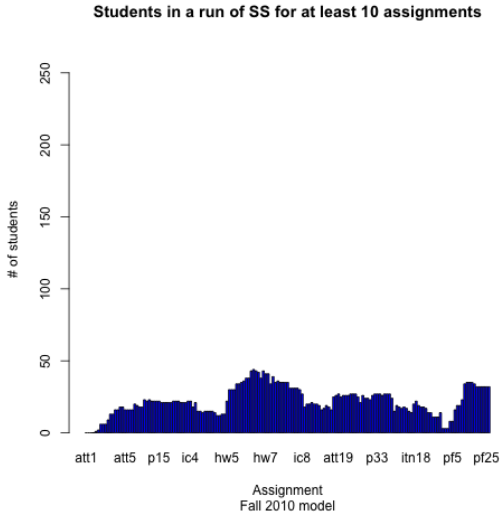
B.3.3 Long Runs in Viterbi Paths

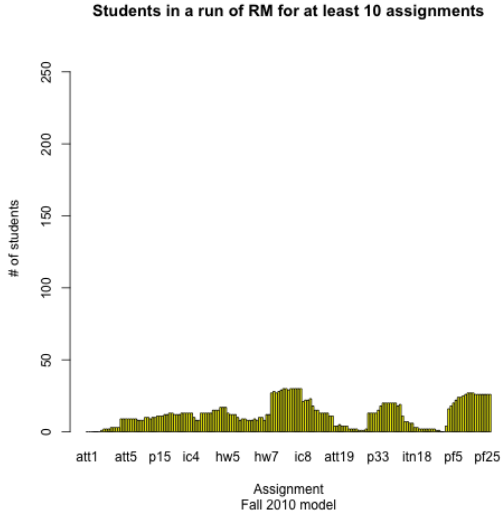



Students in a run of I for at least 10 assignments
# of students
0
50
100
150
200
250
att1 att5 p15 ic4 hw5 hw7 ic8 att19 p33 itn18 pf5 pf25
Assignment
Fall 2010 model


Students in a run of OA for at least 10 assignments
# of students
0
50
100
150
200
250
att1 att5 p15 ic4 hw5 hw7 ic8 att19 p33 itn18 pf5 pf25
Assignment
Fall 2010 model


Students in a run of E for at least 10 assignments
# of students
0
50
100
150
200
250
att1 att5 p15 ic4 hw5 hw7 ic8 att19 p33 itn18 pf5 pf25
Assignment
Fall 2010 model


Students in a run of SS for at least 10 assignments
# of students
0
50
100
150
200
250
att1 att5 p15 ic4 hw5 hw7 ic8 att19 p33 itn18 pf5 pf25
Assignment
Fall 2010 model


Students in a run of RM for at least 10 assignments
# of students
0
50
100
150
200
250
att1 att5 p15 ic4 hw5 hw7 ic8 att19 p33 itn18 pf5 pf25
Assignment
Fall 2010 model