Ben Trey
Project 1 Milestone Report
Data Science Career Track
10/16/2019

Problem Statement:
Given simulated data of a student's scores, is it possible to create a curriculum that maximizes their growth?

Meeting the needs of each student exceeds the resources available. In the traditional classroom a teacher will try to create a curriculum that adapts to each student. The domain of education calls concept of modifying curriculum to meet the needs of each student differentiation. A teacher could break the class into small groups to differentiate instruction, but more than a few groups exceeds the abilities of most teachers. As classroom size grows the opportunity for appropriate differentiation decreases.

In an automated system the opportunity to meet every student's needs greatly increases. The ability to measure and use hundreds of parameters to model a student's success falls well within the capabilities of modern servers and databases. Our current progress in the field allows for differentiation on a level well beyond the traditional classroom.

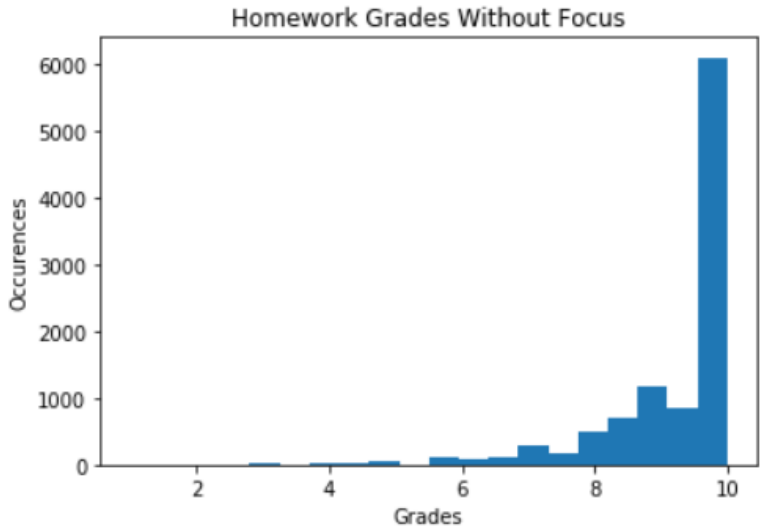
The opportunity to deliver a streamlined individualized education will attract the interests of several parties. Administrators will see the opportunity to reduce the requirements on teachers and improve the quality of graduates. It may also reduce the cost of education by improving the efficiency of assessment. Teachers will see an opportunity to develop all students at the optimal level of individual growth. As the system can adapt using data taken from the student, students will get problems at the appropriate level. They will not waste their time struggling with problems that are too difficult or completing their problems set with unchallenging problems.

Description of Dataset:
The dataset comes a simulation of students completing homework sets. A csv file containing the results of each problem exists for each of 50 problem sets. An additional csv file contains the topic, topic retaught, and difficulty of each problem. The dataset also includes the measured probability of getting a problem correct. The dataset did not need to be cleaned or wrangled.

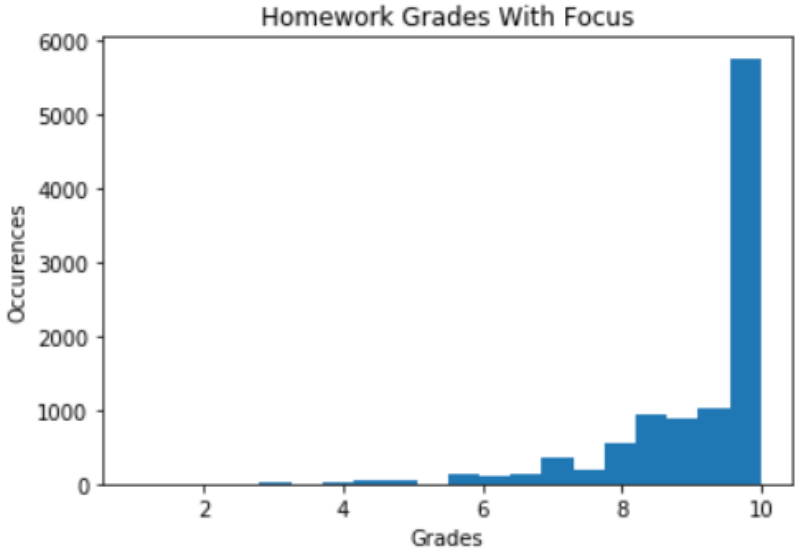
Initial Findings:

The data revealed a relationship between the problems given to each student and their homework scores. The initial data came from random subsets of homework problems. A second simulation focused on selecting random subsets of problems containing problems with topics a student recently answered incorrectly. The second simulation showed a statistically significant reduction in the performance of students in their mid course and end of course assessments. This finding suggests an algorithm can not choose homework sets for a student based only on the topics a student struggles with.

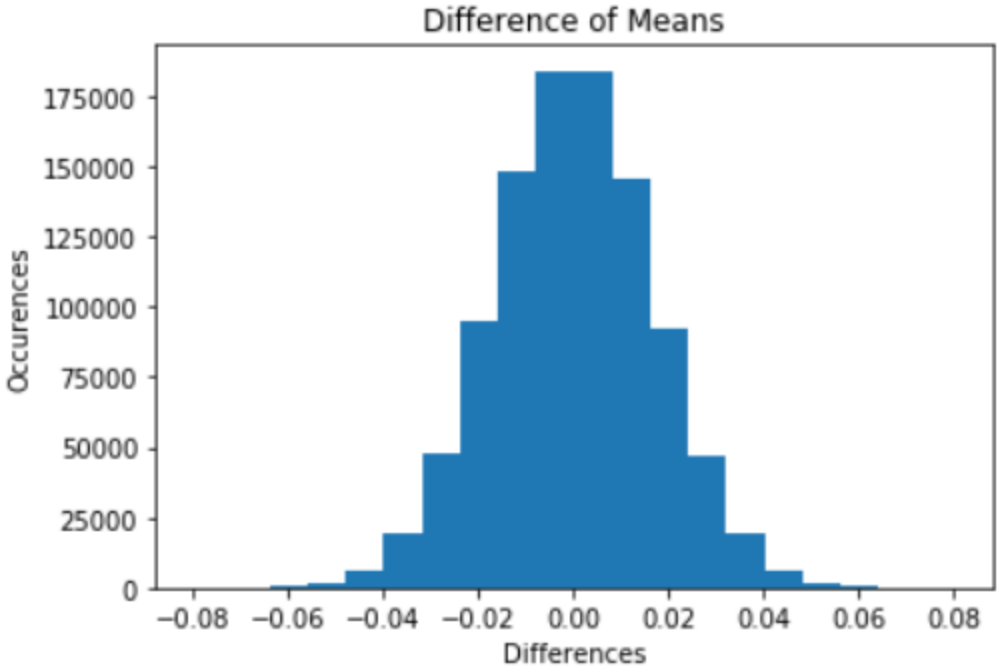
mean without a focus on the student's weaknesses 9.280552654785483



mean with a focus on the student's weaknesses 9.184037833782416



As in the two graphs above show the general distribution of grades appears the same while the means differ. Using the bootstrap method as shown in the graph below the distribution of the differences in means shows the likelihood of the previous measurement being a random event.



The graph above shows the distribution of the difference in means. The difference in the previous two methods is a low probability event and suggests, students using the second technique learned at a slower rate and received lower grades.