UNDERGRADUATE MATH SEMINAR

The next math seminar on the term will be

Professor Julius Barbanel

DATE: **THURSDAY, May 2**

Time &   **12:30 –** Refreshments in **Bailey 204**

Location: **12:50 – 1:45** Seminar in **Bailey 207**

In this seminar, **Professor Emeritus Julius Barbanel** of the Union College math department will present the following talk.

Title: A Fun Probability Problem

Abstract: Consider the following: 100 prisoners are numbered 1, 2, …, 100; 100 boxes are labeled B1, B2, …, B100; 100 pieces of paper have the numbers P1, P2, …, P100 written on them.

The pieces of paper are randomly distributed in the boxes, one piece in each box. Each box is closed so no one can see which paper it contains. One by one, each of the prisoners gets to pick 50 boxes. Each prisoner's goal is to find the box that has their number in it. The prisoners all go free if and only if <u>every</u> prisoner finds their number. If even one prisoner fails to find their number, then no prisoners go free.

The prisoners go one at a time. After each prisoner's turn, the boxes and papers are returned to their original state. The prisoners may discuss strategy before the experiment starts, but no communication is allowed once the experiment begins.

If each prisoner picks 50 boxes randomly, the probability that the players all succeed is approximately zero! We will show how this radically improves if the prisoners all agree to follow a certain strategy.

Pieces from Thesis – by Amanda Sgueglia

Amanda wrote her senior thesis this past winter term, supervised by Professor Roger Hoerl.

Although I am a math major and have always been interested in the math theory space, I have also always had a love for statistics. I enjoyed expanding on my statistical knowledge in all the classes I took with Professor Hoerl, especially in the Big Data Analytics course. After learning about machine learning methods and working with large data sets in this class, I decided this was the type of work that I wanted to pursue for my senior thesis.

Throughout the winter term of this year, I worked with Professor Hoerl and completed my thesis titled "Exploring the Limitations and Robustness of Big Data Analytics: Analyzing Salaries in Data Science." For this project, I used R to analyze three different data sets (all with roughly 3,000 data points) related to jobs in data science, with a dependent variable of salary. I used four different machine learning methods (linear regression, ANN, SVM, and Random Forests) and evaluated them when predicting salary in terms of their "robustness." Their robustness related to their ability to produce accurate predictions, even with data issues such as extrapolation beyond the current range of x variables, outliers, and missing x variables (as well as their two and three factor interactions). To measure accuracy, I looked at how these models predicted salaries in and out of sample, as well as how prone they were to overfitting. This process allowed me to assess the benefits and limitations of using the four different machine learning methods when applying them to data sets of my choice, and as a result understand the limitations of these techniques with big data as a whole.

I really enjoyed working with Professor Hoerl and expanding on my knowledge of machine learning methods. One main piece of advice I would give to underclassmen who will be writing their thesis is to pick a topic that truly interests them. Since I would like to work in the data analytics field after I graduate, my thesis process was an incredibly interesting and beneficial experience. While a senior thesis can seem daunting in the beginning, it is truly an exciting and rewarding opportunity that I am grateful for.