

Machine Learning of the College Admissions Process

Sam Rush
Computer Systems Lab 2009-2010
Period 4

Abstract

The goal of this project is to analyze the various biases that exist in the college admissions system by attempting to predict college decisions. This project will attempt to reduce college admissions to pure numbers, excluding data that is inaccessible such as essays and teacher recommendations. Past user-submitted data from the 2007, 2008, and 2009 Senior Destinations websites will be used to train an artificial neural network in a process known as machine learning to perform a nonlinear least squares fit. Then, factors such as the gender bias and the race bias will not only be proven to exist but will be quantifiable based on their role in the least squares fit.

Introduction

- Students pad resumes, but can they do it more efficiently?
- Is there a gender bias? A race bias?
- How well can decisions be predicted?

Senior Destinations

Technical Details

- Languages used: PHP
- MySQL for database storage
- Hosted on my home server

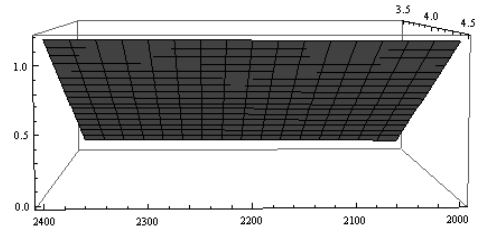
Incomplete Data

- Previous years lack race and gender data
- Only about 3/5 of each senior class submits data
- Due to FAIRGRADE, GPA data is inconsistent

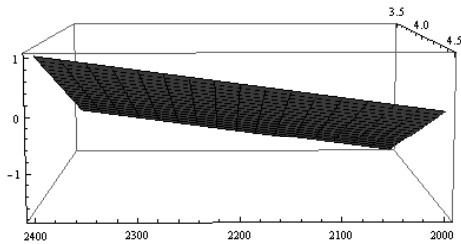
Procedure

- Linear least-squares fit using QR-decomposition
- $Ax=B$, A is the student matrix and B is the decision vector
- Row vector for a new student * x = result

UVa vs. GPA vs. SAT



Penn vs. GPA vs. SAT



Prediction Results

College	#Correct	Out of	Prediction Rate
Brown University	16	22	72.7%
Cornell University	28	37	75.7%
Duke University	16	20	80.0%
University of Pennsylvania	16	21	76.1%
University of Virginia	78	83	93.9%
Virginia Tech	40	40	100%

To Do List

- Add in the nonlinearity portion to the least squares, which follows quickly from the QR decomposition
- Add in other factors
- Compute the Biases

Works Cited

- Thiagarajan, Arvind. "TJHSST Class of 2009 Senior Destinations" <<http://www.kavitech.com/EduInfo/Destinations/Destinations.html>>
- Chen, Je. "TJHSST Class of 2008 Senior Destinations" <<http://www.tjhsst.edu/jchen2/college>>
- Wang, Jonathan and Zeng, Will. "TJHSST Class of 2007 Senior Destinations" <<http://www.tjhsst.edu/pwang/college/base.php>>
- Fairfax County School Board Votes to Change Grading Scale." Fairfax County Public Schools, 1/23/2009 <<http://commweb.fcps.edu/newsreleases/newsrelease.cfm?newsid=1058>>
- Chang, Lin "Applying Data Mining to Predict College Admissions Yield: A Case Study" New Directions for Institutional Research, n131 p53-68 Fall 2006
- Sauer, Timothy "Numerical Analysis", Addison Wesley, 2005, ISBN 03211268989
- Sedgewick, Robert and Wayne, Kevin, "GaussianElimination.java" 9/29/2009 <<http://www.cs.princeton.edu/introcs/95linear/GaussianElimination.java.html>>