Javon Kitson, Jordan Hettleman, Austin Pierce-Ptak
Dr. Simari
Computer Science 484
8th April 2019

Artificial Intelligence Project Part 1

Dataset

For this assignment we're using a database of graduate admissions. The dataset was created for those outside America and those who don't speak English as their first language which is why you see every student have a TOEFL score. The datasets parameters include:

1. Serial No. Equivalent to an id.
2. GRE Scores (0 to 340)
3. TOEFL Scores (0 to 120)
4. University Rating (1 to 5)
5. Statement of Purpose (1 to 5)
6. Letter of Recommendation Strength (1 to 5)
7. Undergraduate GPA (0 to 10)
8. Research Experience (0 or 1)
9. Chance of Admit (0% to 100%)

Proposed Methods

Problem type

The methods we decided to use are all unsupervised. Output will be outlined in the next paragraph.

Machine Learning Technique

We want to use K-mean to cluster the data into 10 clusters. Given how schools are ranked between 1 and 5 we thought to use two clusters per school category to show students information for those who were accepted into a specific category and those who weren't accepted into another category. This will only show those who applied to each so an individual who applied to a category 5 school and didn't get in wouldn't show in a yes cluster for a category 1 school. For MLP we want to predict whether an individual would get accepted into a specific category of school. With this we could use the aforementioned 10 clusters as test cases to evaluate individuals thus being able to predict whether any individual in a category 5 no cluster could get into a category 1 cluster.

We plan to use python as the coding language and import the libraries Sklearn, numpy, os, and pandas. Also, as outlined in the assignment we will compare the K-means technique against Naïve Bays.

Sources:

Acharya, Mohan S. "Graduate Admissions." Kaggle, Kaggle, 28 Dec. 2018, www.kaggle.com/mohansacharya/graduate-admissions#Admission_Predict_Ver1.1.csv.