Student Intervention System

Udacity Machine Learning Engineer
Nanodegree Program: Project 2

Anderson Daniel Trimm

February 5, 2016

1 Classification vs Regression

Our model predicts which students will pass or fail their high school final exam, and therefore need an intervention. Since we are predicting *discrete* labels (in this case, binary), this is a *classification* problem, as opposed to a *regression* problem, in which we would be predicting *continuous* labels.

2 Exploring the Data

In this section, we use the code in the accompanying ipython notebook to identify important characteristics of the dataset which will influence our prediction. Running the code, we find

```
Total number of students: 395
Number of students who passed: 265
Number of students who failed: 130
Number of features: 31
Graduation rate of the class: 0.67%
```

3 Training and Evaluating the Models