

Project Report: Student Intervention System

Aravind Battaje

March 5, 2016

1 Project Steps

TODO: Write later

2 Classification vs Regression

A machine learning algorithm can be classified into two types, based on its nature of outputs, viz., classification and regression. Classification supports outputs of discrete values and regression outputs continuous values. This project entails a classification type of problem because the output desired from the *intervention system* is discrete in nature, i.e., a student graduates or not from his/her current characteristics. Regression would be more suitable for, say an algorithm that predicts the final exam score from a student's current academic records.

3 Dataset

Several qualities of students such as their family background, social characteristics, extra-curricular activities, etc., along with the information if they graduated or not, are given along with the project (`student-data.csv`). The dataset possess following characteristics:

Total number of students	395
Number of students who passed	265
Number of students who failed	130
Graduation rate of the class	67.09%
Number of features of dataset	30

4 Notes

Use of Neural Networks was considered, but scikit-learn (stable) doesn't directly support multi-layer perceptron currently. Although other libraries (Theano, scikit-neuralnetwork) could be used, exploration has been pushed forward both because it was suggested to "choose 3 supervised learning models that are available in scikit-learn", and neural networks will be encountered during *Deep Learning*.