Capstone Machine Learning

Drew Matheson

10/24/2017

The Problem

Can we use the JAM assessment results from the end of 1 year at school to train a model to predict the National Standard judgement for a student at the End of Year 3?

This is a supervised classification problem with a binary outcome: Will a student achieve 'At' (1) or 'Below' (0) the National Standard after 3 years at school?

The independent variables are: * Gender * Ethnicity * Additive Strategies stage (from the JAM Assessement) * Numeral Identification (from the JAM Assessement) * Forwarwd Number Sequence (from the JAM Assessement) * Backwards Number Sequence (from the JAM Assessement) * Fraction Knowledge (from the JAM Assessement) * Place Value Knowledge (from the JAM Assessement) * Basic Facts Recall (from the JAM Assessement)

The dependent variable is: * At Standard (0 or 1)

Using Binary Logistic Regression

The first approach will be to create a binary logistic regression model that will predict if a student will achieve 'At' the standard using the independent variables. We will first try a model using all of the independent variables. We will look for variables in the model summary that have some level of significance and run a selection of models using different combinations of these independent variables. We will select a model that has significant coefficients, and the lowest AIC value.

Using CART

It is possible that this problem will require a CART approach. We will create a CART model and then run the model a number of times to try a range of the minbucket parameters. We will use Monte Carlo cross-validation to calculate the accuracy of each model and select the model with the greatest mean accuracy.