Programming Homework 7

CS540

August 18, 2019

**Introduction**

1. After 20 years of data on Kobe Bryant’s makes and misses, it is possible to predict if certain shots went in or not. This program is based off of a Kaggle competition to predict if certain shots are a make or a miss. The data can be found on kaggle.com. This program uses a decision tree with kfold cross validation similarly to P3.
2. Instead of reusing the TA’s P3 code I found it easier to use the machine learning library scikit-learn.
3. It says on piazza that using a completely new data set would warrant 8 points, so hopefully this counts for full credit.

**Details**

1. Download the Kobe Bryant Shot Selection data. ([www.kaggle.com/c/kobe-bryant-shot-selection/data)](http://www.kaggle.com/c/kobe-bryant-shot-selection/data))
2. Clean the data to separate test set and train set. This can be done by setting any row in the ‘shot-made-flag’ column that contains null to a separate csv
3. Convert string data to categorical binary data.
4. Use basic classification to determine over fitting
5. Utilize k-fold cross validation and grid search CV to predict if the shot is made
6. Print results to CSV

**Results:**

1. The output CSV contains the index for the shot and if it is a miss, 0, or a make, 1.