Daniel Silva

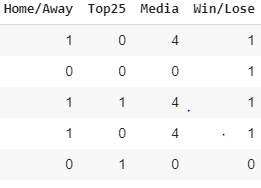
CAP 5610

Homework 4

2/11/2020

The code for this assignment can be found here:  
<https://github.com/danielzgsilva/CAP5610-Machine-Learning/tree/master/HW4>

In this assignment, we were given a university’s football game data and asked to construct a Naïve Bayes Classifier which predicts the outcome of a given game. I leveraged Scikit-learn’s implementation of a Naïve Bayes Classifier for categorical data. Prior to fitting a classifier to the data, I first had to map each category of each feature to an integer in the range [0, num\_categories-1]. This essentially converts each categorical feature into an ordinal feature, where each integer in a column represents a distinct category. The dataset is now made up of solely integers:



Once the dataset consists of solely integers, it can be fed to the Naïve Bayes Classifier. Constructing and training Scikit-learn’s classifier is done like so:



Note that we are leveraging sklearn’s Naïve Bayes Classifier for categorical data, which is suitable for classification with discrete features that are categorically distributed. At this point, the classifier has been fitted to our data and we are ready to generate predictions on the test set. Generating predictions is done like so:



This results in the following predictions:

