

Lab 2 Report

Elise McEllhiney

February 16, 2017

Naïve Bayes Classifier

My Naïve Bayes Classifier code inputs two sets of data, a training set that is used to train the algorithm and a test set that is used to test whether or not it has managed to classify the test data correctly. In `calcConditionalProb(test,train)`, given a single class, I calculate the mean and standard deviation of each attribute and then use the normal distribution to calculate the probability that the test data belongs in the same class. Then, I take the product of the probabilities of the attributes to get the overall probability. This is then weighted by multiplying it by the class probability, `probTarget`. These are put into the probability dataframe, `prob_df`, and then I generate my prediction by outputting the class with the highest probability for each test case. Lastly, I generate the confusion matrix which shows the values predicted compared to the expected values. The average is generated by summing up all the correct predictions, `prediction==groundTruth`, and dividing that by the total number of predictions made.

The console output includes the raw training data, raw test data, the probability dataframe, prediction, confusion matrix, and accuracy. The probability dataframe has each test case being it's own row and the columns are associated with the classes. The prediction contains the class with a greater calculated probability of the associated test case, and thus our prediction for the test case. The confusion matrix has the ground truth values along the y-axis and my predicted values along the x-axis.