Building Multi-Class Models with the Iris Dataset

1. Introduction

One of the classic example data sets in machine learning is the Iris dataset. This dataset contains three different species of iris, each entry with a measurement of sepal length and width and pedal length and width. For this project, I split the Iris dataset in to two, one for testing and one for training. Then, I trained two different models, one using a custom naive bayes algorithm and the other using vowpal wabbit. Later we compare the two models for accuracy and explain some flaws in our experiment.

2. Naive Bayes Analysis

2.1. Experimental Setup

For this experiment, I used a python script to train a model on various sized datasets. Then, using the rest of the iris model, I tested for accuracy. I also used basic "Add 1" smoothing to account for features that weren't seen in training. The naive bayes algorithm used, uses log space multiplication to simplify the expression

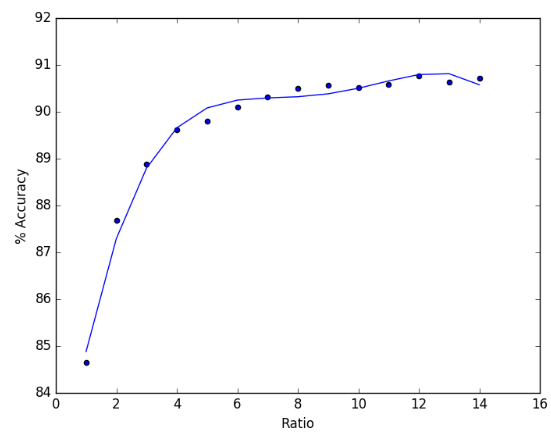
$$P(C)\prod_{i=0}^{n} P(X_i|C) = log(P(C)) + log(P(X_0|C)) + ... + log(P(X_n|C)$$

2.2. Results

Shuffling the dataset between runs, I computed the accuracy based on different ratios of training:test data. With ratios ranging from 1:1 to 15:1 , I ran 5000 tests on each ratio to calculate the average accuracy, producing figure 1.

As expected, the larger the training set, the more accurate the model. As we approach the full size of the dataset, the accuracy appears to approach somewhere around 91 percent, a decent result for naive bayes. With a larger dataset, the accuracy could likely be increased.

Figure 1. The blue line represents best fit, and the ratio is the number on the x-axis:1 (training:test)



3. Vowpal Wabbit Analysis

3.1. Experimental Setup

To set up VW, I shuffled the iris dataset and put it in two separate files, one for training and one for testing. Then, I used the scripts given to us to train and test the dataset using vowpal wabbit. Afterwards, I used a custom python script to calculate the accuracy of the model.

3.2. Results

Due to issues with VW, I wasn't able to get this working in time. My model seemed to not want to classify one of the classes, although I used scripts provided on the forum. This resulted in an accuracy of 66.66 percent. Other students were able to get an accuracy around 96 percent which is to be expected. A model like vowpal wabbit that uses logistic regression rather than naive bayes normally performs better.

4. Conclusion

For this project, I used the iris dataset to create two different multi-class models to predict flower type. Based on the experiments I ran, I determined that vowpal wabbit is a superior method to naive bayes for predicting class using this dataset. Overall, our predictions weren't very accurate. We did show that the more training examples, the better the model that resulted. In order to create a more robust system with either method, we would likely need a larger dataset to train on.

5. Software and Data

Vowpal Wabbit
https://github.com/JohnLangford/vowpalwabbit
Iris Dataset
https://archive.ics.uci.edu/ml/datasets/Iris