Assignment 2: Mushroom Classification Report

CPSC-371

Artificial Intelligence

Due: March 17, 2023

Andrew Hunter-Owega	Daniel Strickland	Nicholas Hirt
ahunter@unbc.ca	dstrickla@unbc.ca	nhirt@unbc.ca
230 147 039	230 146 357	230 127 295

Test Data Division:

To train and test the various models for predicting the edibility of mushrooms, a data set with eight thousand mushroom records was utilized. For the purpose of this exercise, we decided to use an 80-20 split in favor of training data. In other words, 6400 mushroom records were used for the purpose of training, and 1600 mushroom records were used as testing data. The first 6400 mushroom records in MushroomData_8000.txt are used to train the model. It is these training records for which the Nearest Neighbor searches for closest neighbors, where statistics are computed over in Bayes, and the data used to train the perceptron model exist. The final 1600 entries of MushroomData_8000.txt are utilized for the purpose of testing the accuracy of the three applications. This is accomplished by allowing the trained models to predict the results of the 1600 values, and comparing with the actual answer. Every incorrect prediction is determined and counted, to determine a final error rate percentage.

Analysis of Methods:

Nearest Neighbours:

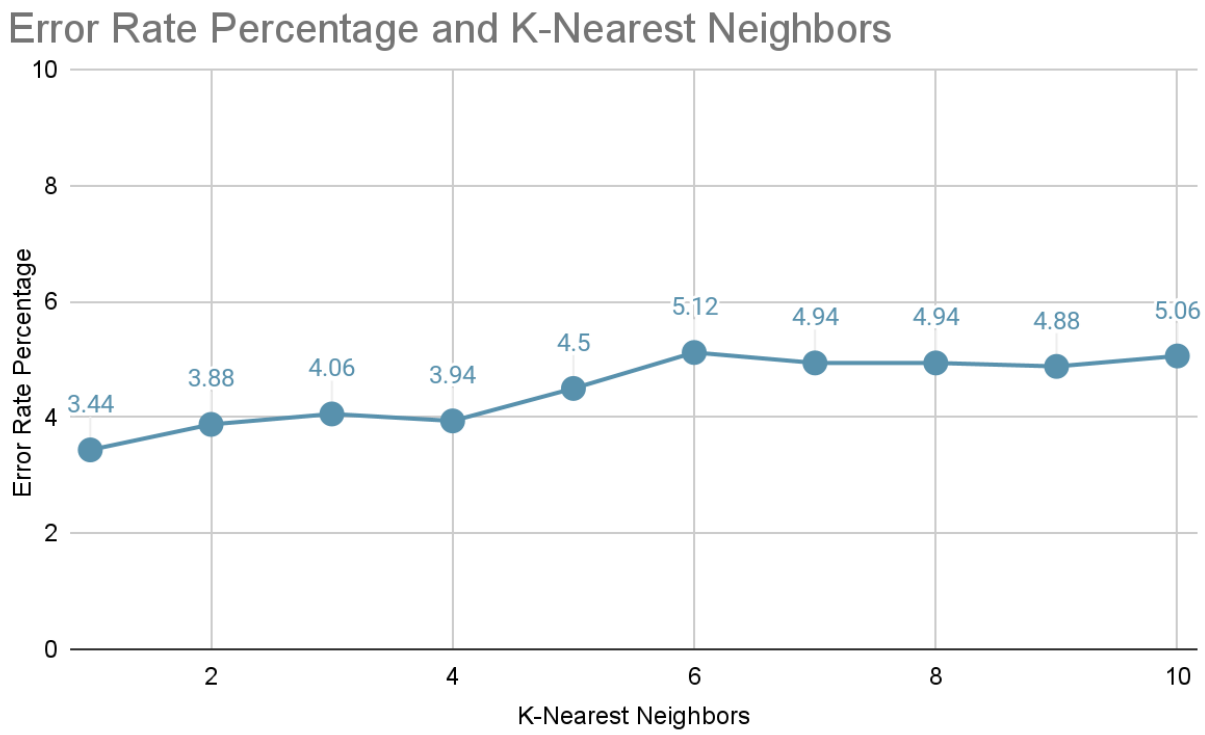
Settings:

k - nearest neighbors: 1

Best Achieved Accuracy Rate: 96.56%

- Depending on the number of nearest neighbors to be analyzed an error rate between 3.44% and 5.06% are calculated.

Outcome Chart:3



Bayes:

Settings: N/A

Best Achieved Accuracy Rate:

- Due to the repeatable and relatively non-variable nature of the Bayes method for prediction, the edibility of mushroom records does not typically vary. The determined percentage for accuracy was determined to be 93%.

Outcome Chart: N/A

Perceptron:

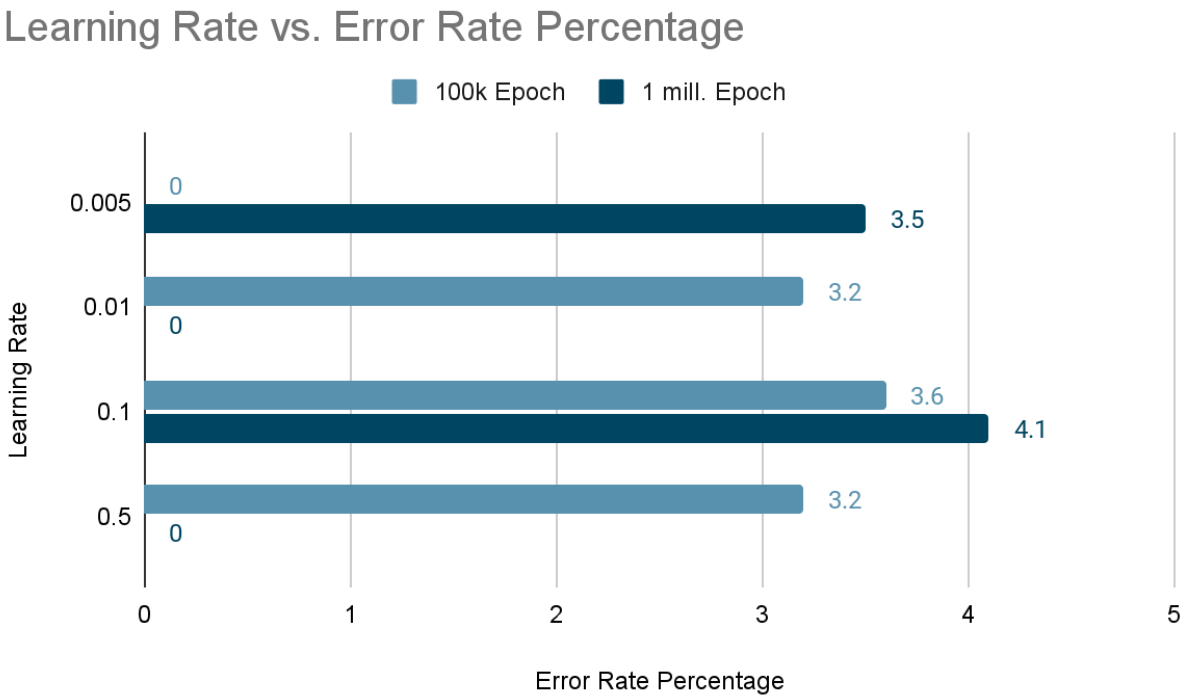
Settings:

- **Epoch Count:** 1 million
  Learning Rate: 0.005
  Starting Weights: 0

Best Achieved Accuracy Rate: 96.8%

- Depending upon the settings of the Perceptron an error rate between 3.2% and 4.1% was seen in the 1600 test records.

Outcome Chart:



Results Analysis:

The results achieved above showcase that for the purpose of classifying mushrooms all of the three methods are roughly equal in accuracy. The K-NN nearest neighbor is able to achieve an accuracy rating of 96.56%, the bayes prediction is able to achieve an accuracy rating of 93.00%, and the perceptron based model was able to achieve the best accuracy rating of 96.8%. While each of the methods can eventually be brought to approximately the same accuracy level, their needs for computation differ significantly. The nearest neighbor for example takes a very long time during its runtime, as it must perform thousands of euclidean distance calculations. The perceptron model however, has a heavy upfront cost to train the model, and then is computationally cheap (at least compared to nearest neighbor) to make predictions. The Bayes prediction is somewhere in between as it needs to compute expensive statistics upfront during runtime, but is then able to make relatively quick predictions. This relatively better performance of Bayes, comes at the cost of some accuracy as seen in the results. Regardless, any of the three artificial intelligence methods achieve roughly the same error rate.