Jake Stephens

Assignment #5

August 10, 2021

**Question #1:**

1. Done
2. Done

**Question #2:**

1. Done
2. Accuracy: 72.62%
3. Confusion Matrix:

[[160 84]

[207 612]]

**Question #3:**

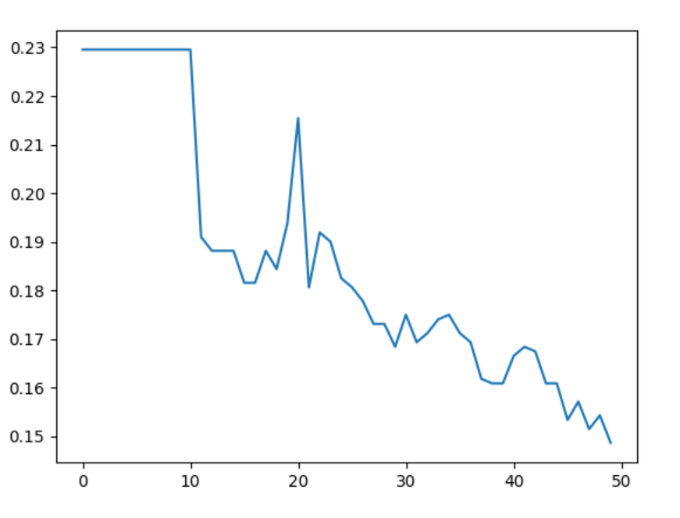
1. Done
2. Accuracy: 78.93%
3. Confusion Matrix:

[[135 109]

[115 704]]

**Question #4:**

1. **Done**
2. For each 10 values on the x axis, (0-10, 11-20, 21-30, 31-40, 41-50) this is d=1,2,3,4,5. The value of d increases by 1 for each 10 values on the x axis. The values of N increment by 1, all the way to 10 for each value of d.



1. Accuracy: 85.14%

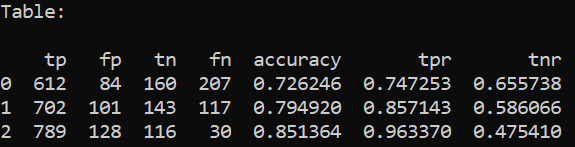
With the best combination being **d=5** and **N=10**. The maximum values for each. Which is interesting. I am curious to know if I had increased d and N even further, the accuracy would get better and better. As you can see from the graph, the error rate is trending downward as the values increase.

1. Confusion Matrix:

[[116 128]

[ 30 789]]

**Question #5:**



The table above shows the results of the 3 classifiers used. Row 0 is Naïve Bayesian, row 1 is Decision Tree, and row 2 is Random Forest. It is surprising to me how well the accuracy of Random Forest was. However, it was the worst at predicting false positives. Which is bad because in this case, we are talking about the status of a human fetus, and to predict that the fetus is in normal status, when it really is abnormal, is not good at all.

The Naïve Bayesian method had the best true negative rate of 0.655738, which is significantly higher than the other 2 methods. Which is very interesting to me because it had the lowest accuracy overall, but it is the best method for predicting negatives.

The Decision Tree is right in the middle of the other 2 methods it seems. For every value in the table, it is between Naïve Bayesian and Random Forest. But if I were to pick a method for making these predictions, I would choose Naïve Bayesian because of the high true negative rate. In my opinion for this situation it is better to be right about a negative, than wrong about a negative, regardless of the poor accuracy compared to the other 2 methods.