Pattern Recognition HW3 report

PART 1

1.

After running the code, I get:

Gini of data is 0.4628099173553719

Entropy of data is 0.9456603046006402

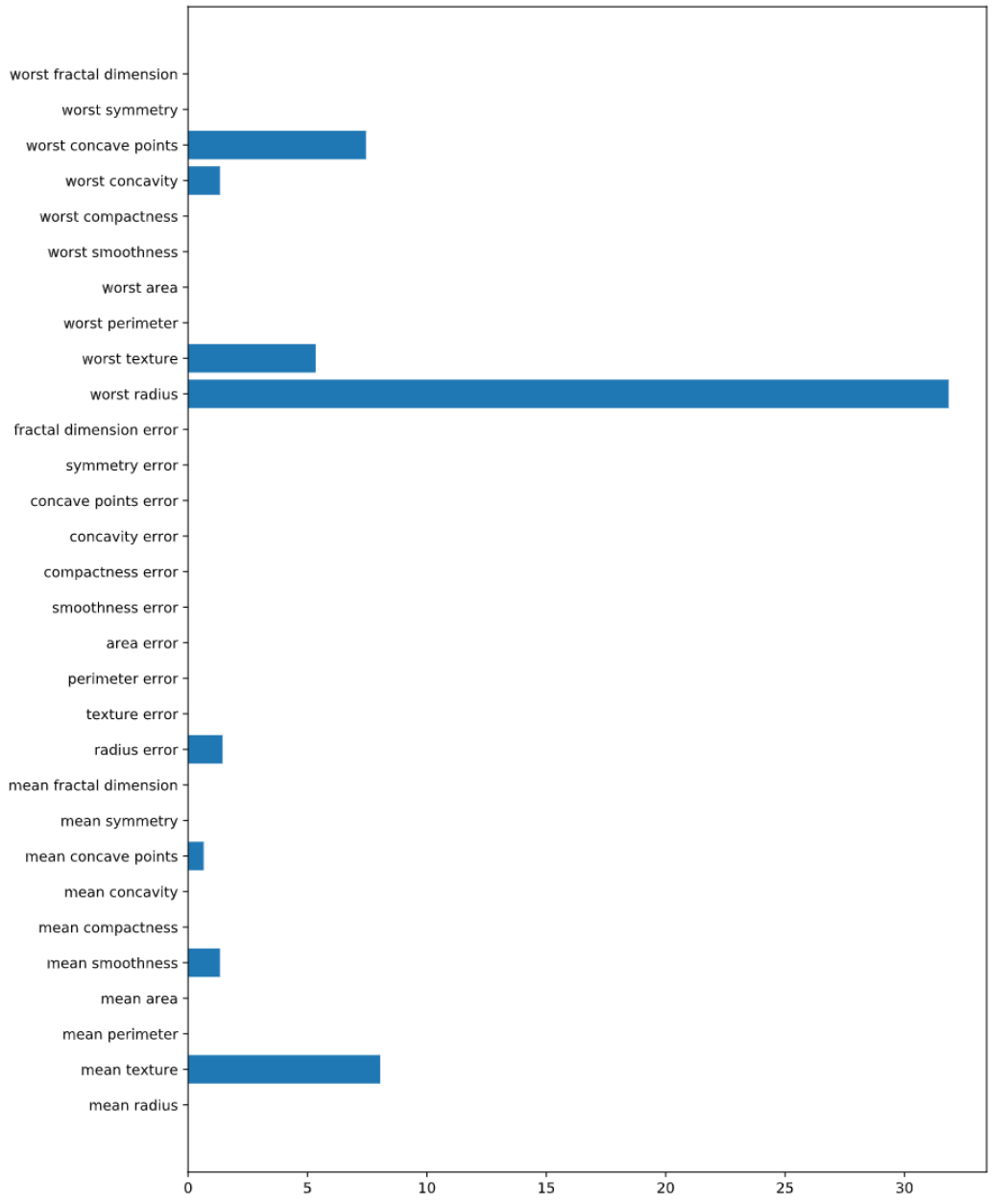
2.

1) score for Max_depth=3: 0.9440559440559441
   score for Max_depth=10: 0.9370629370629371
   The reason Max_depth=10 gives a lower score might be because of overfitting

2) score for gini criterion: 0.9440559440559441
   score for entropy criterion: 0.958041958041958

3.

Here is the graph showing feature importance for a tree of gini criterion and max_depth of 10



4.

I haven't finished the algorithm as I am missing the sampling on the features but it still creates a forest from different data samples and can predict the class with a majority vote

PART 2

For model A, left node will predict C1 and right node will predict C2. The number of misclassified points is 100 at the left node and 100 at the right node. We have a to misclassification rate of 200/800 = 0.25 for A

For model B, left node will predict C2 and right node will predict C1 therefore we have 200 misclassified point for left node but 0 for the right node. The misclassification rate 200/800=0.25 is the same as model A.

Using the formulas we get gini = 0,7778 and entropy = 0,8113 for model A and for model B, gini = 0,3333 and entropy = 0.6887. We can see that both the entropy and the gini of tree B are lower than tree A.