1.

a) Plots of the out-of-bag error as a function of the number of bags

|  |  |
| --- | --- |
| **One VS. Three** | **Three VS. Five** |
|  |  |

b) Cross-validation error for single decision tree and the OOB error for bagging

|  |  |  |
| --- | --- | --- |
|  | One VS. Three | Three VS. Five |
| Cross-validation error for single decision trees | 0.0078 | 0.0585 |
| the OOB error for bagging | 0.0024 | 0.0338 |

c) Test error for single decision tree and the 200 bagged decision trees

|  |  |  |
| --- | --- | --- |
|  | One VS. Three | Three VS. Five |
| Test error for single decision tree | 0.0163 | 0.1196 |
| Test error for 200 bagged decision trees | 0.0116 | 0.0859 |

d)

The above graphs show the relation of out-of-bag error with number of bags. We can observe in both graphs above that the error decreases sharply as number of bags beginning to grow from zero. There is good partial prediction of the dataset when we train the whole dataset as a bag iteratively. And when we collect all the votes and apply the majority rule to get our fit for the aggregated hypotheses, which predicts the entire dataset fairly well.

Bagging algorithm can generate smaller errors than a single decision since bagging can generate hypotheses with low variance. Additionally, we always find the hypothesis with lowest error for each

randomly chosen bag, so our bias is fixed. Low variance of the aggregated hypothesis leads to better performance.

We can see that one-vs-three classification problem have better results than the three-vs-five problem. We reason that ‘1’ and ‘3’ are less likely to be misidentified due to more differentiable features while ‘3’ and ‘5’ have much closer assemblance.

2.

Plots of training set error and the test set error against number of weak learners

|  |  |
| --- | --- |
| One VS. Three | Three VS. Five |
| C:\Users\77529\AppData\Local\Temp\WeChat Files\f088dde0a7b3e3de77c26cd0d1d4c26.jpg | C:\Users\77529\AppData\Local\Temp\WeChat Files\a09445d53bc6eefe6e45bc58674fa34.jpg |