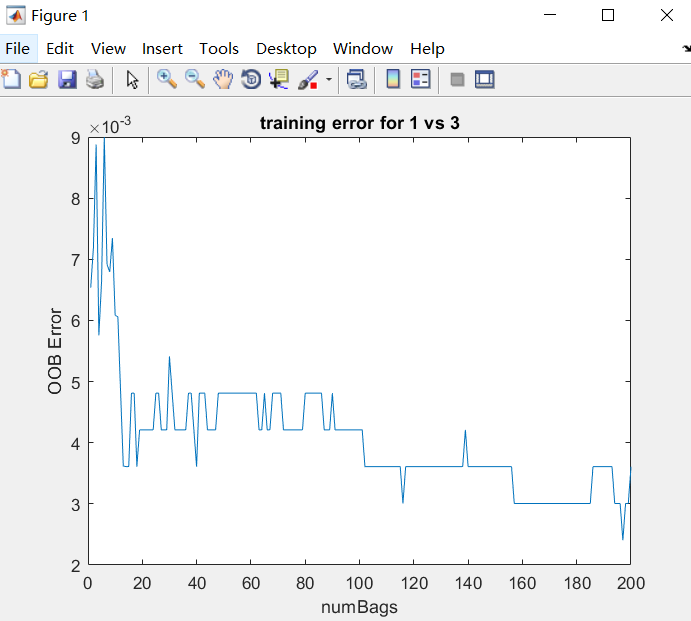
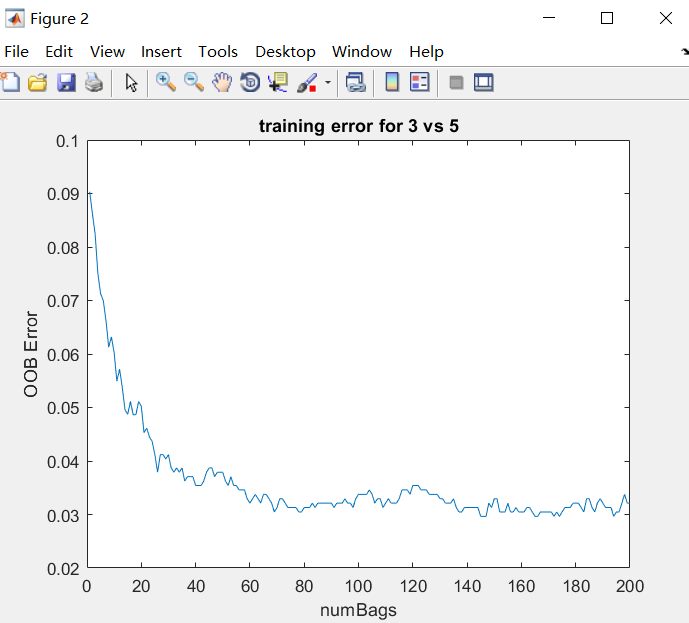
1.

a.





b.

|  |  |  |
| --- | --- | --- |
| Training Error | CV Error | OOB Error |
| 1 vs 3 | 0.0090 | 0.0036 |
| 3 vs 5 | 0.0684 | 0.0321 |

c.

|  |  |  |
| --- | --- | --- |
| Test Error | Single Tree | Ensemble of 200 Trees |
| 1 vs 3 | 0.1196 | 0.0116 |
| 3 vs 5 | 0.0163 | 0.0859 |

d.

For figure 1 (training error of 1 vs 3), the OOB error decreases dramatically as the number of bags goes up from 1 to roughly 15. Then, the OOB error decreases slightly when number of bags goes up from 15 to 200. Increasing size of ensemble can lead to better generalization. The cross validation error (training of 1 vs 3) is 0.0090. The OOB error is 0.0036 when the number of bags is 200. The OOB error performs better than cross validation error.

For figure 2 (training error of 3 vs 5), the OOB error decreases dramatically as the number of bags goes up from 1 to 80. Then, the OOB error almost remains stable along with the increment of number of bags. The cross validation error is 0.0684 and the OOB error (number of bags =200) is 0.0321. The OOB error also performs better than cross validation error. The cross validation error and OOB error of 1 vs 3 are lower than 3 vs 5.

In terms of test error, a single tree of 1 vs 3 doesn’t perform well and the test error is higher than single tree of 3 vs 5. However, the test error of ensemble of 1 vs 3 is lower than 3 vs 5. The test error of ensemble of 200 trees can lead to better generalization.