

1. PCA

a) see q1.py. Approximately 88.3% probability

b) see q1.py

8 components were needed to capture 95% variance. Although contribution of each features to different principal components vary, 1st, 3rd, and 7th components seems to be important as they have highest absolute values for the first 2~3 components that amounts to most of the variance.

c) see q1.py

2. Almost Random Forest

a) see rf.py

b) see rf.py

Setting parameters as high max_depth / high nest / low minleaf samples seems to lead to better accuracy Although the increase in accuracy is not drastic. The highest increase in accuracy was when increasing the maxdepth from 5 to 10, which led to 0.4% increase in accuracy from 85.8% to 86.2%. .

c) see rf.py

The accuracy on the test data is slightly higher than oob error but difference is minimal. My results for oof accuracy was 86.2% on optimal and test accuracy was 86.45% on optimal.