Q4:

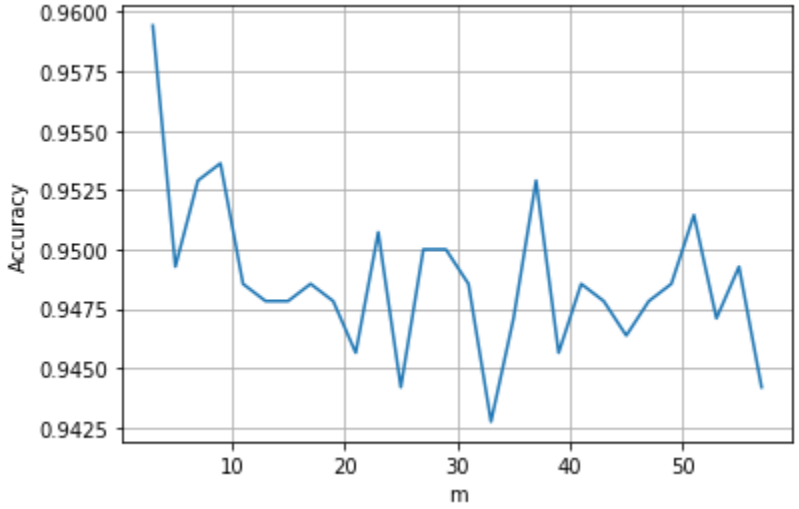
a) Random Forest from scratch

Accuracy around 0.93-0.95 is obtained. Takes about a minute to run

Random Forest from sklearn

Accuracy around 0.94-0.96 is obtained. Takes few seconds to run

b) Sensitivity to m

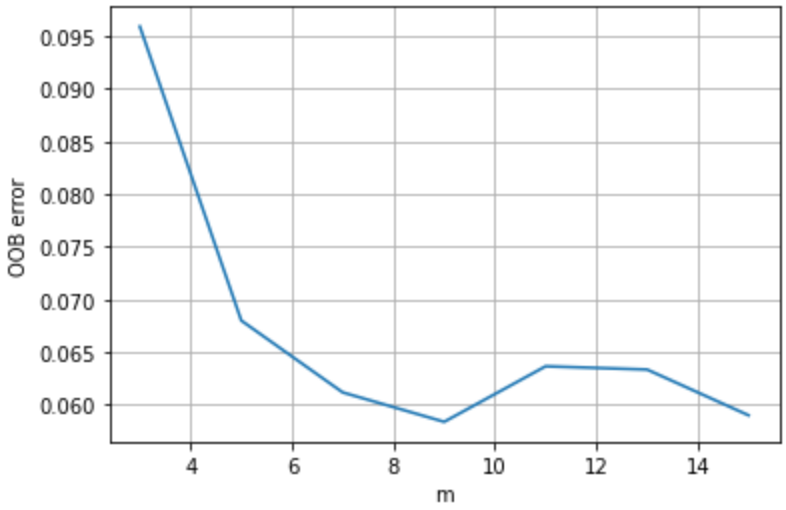


When m<<N, accuracies seem to be better than when m is comparable to N. We can also see that the accuracies seem to touch peaks when values of m are close to sqrt(N)/2, sqrt(N), 2sqrt(N), where N is total features (which is 57 here). These are the values which Breiman suggested.

c) Plot of OOB error

```
OOB for m= 3 : 0.09593294008072027
OOB for m= 5 : 0.06799130704750078
OOB for m= 7 : 0.0611611300838249
OOB for m= 9 : 0.058366966780502946
OOB for m= 11 : 0.06364483079788885
OOB for m= 13 : 0.06333436820863086
OOB for m= 15 : 0.05898789195901894
```

OOB error seems to decrease as m increases, and it ends up being around 0.06, in the range of m chosen. Plot on next page.


0.095
0.090
0.085
0.080
0.075
0.070
0.065
0.060
OOB error
4
6
8
10
12
14
m