CS6510: Assignment 2
ES16BTECH11009

All the codes are written in python3.

A4a) Test Accuracy: 0.9787735849056604

Number of support vectors: 28

A4b) for 50 points:

Test accuracy: 0.9811320754716981

Number of support vectors: 2

for 100 points:

Test accuracy: 0.9811320754716981

Number of support vectors: 4

for 200 points:

Test accuracy: 0.9811320754716981

Number of support vectors: 8

for 800 points:

Test accuracy: 0.9811320754716981

Number of support vectors: 14

A4c) 1) False

for q = 2 c = 0.0001

Training error 0.008968609865470878

number of support_vectors 236

for q = 5 c = 0.0001

Training error 0.004484304932735439

number of support_vectors 26

2) True

for q = 2 c = 0.001

number of support_vectors 76

for q = 5 c = 0.001

number of support_vectors 25

3) False

for q = 2 c = 0.01

Training error 0.004484304932735439

number of support_vectors 34

for q = 5 c = 0.001

Training error 0.0038436899423446302

number of support_vectors 23

4) False

for q = 2 c = 1

Test error 0.018867924528301883

number of support_vectors 24

for q = 5 c = 1

Test error 0.021226415094339646

number of support_vectors 21

A4d)

C= 0.01 , train error = 0.0038436899423446302

C= 1 , train error = 0.004484304932735439

C= 100 , train error = 0.0032030749519538215

C= 10000 , train error = 0.002562459961563124

C= 1000000 , train error = 0.0006406149903908087

For C= 1000000 train error is lowest

C= 0.01 , test error 0.02358490566037741

C= 1 , test error 0.021226415094339646

C= 100 , test error 0.018867924528301883

C= 10000 , test error 0.02358490566037741

C= 1000000 , test error 0.02358490566037741

For C= 100 test error is lowest

A5a)

Train error :0.0

Test error: 0.02400000000000002

Number of support vectors:1084

A5b) RBF kernel:

Train error :0.0

Test error:0.5

Number of support vectors: 6000

Poly kernel :

Train error : 0.0

Test error:0.02100000000000002

Number of support vectors:1755

Linear and Poly kernel have the least train error

A6a) For parameters :

numberOfForests(k) = 21 , numberOfAttribuites(m) = 8

Sklearn:: accuracy = 0.94710, Time = 2.017 s

Mine :: accuracy = 0.94347 , Time = 8.781s

k = 15, m = 12

Sklearn:: accuracy =0.94565, Time =0.609s

Mine :: accuracy = 0.95072, Time = 9.385s

K = 25 , m = 18

Sklearn:: accuracy =0.94637, Time = 0.761s

Mine :: accuracy =0.94420, Time = 20.446s

Sklearn's Random Forest is much faster than mine random forest for all the configurations I tried. In terms of accuracy both the classifiers give similar output .

A6 b) I ran random forest with parameter k = 21 and with m changing I got the following values :

m	test error
1	0.1188405797
3	0.06739130435
5	0.05652173913
7	0.05434782609
8	0.04637681159
9	0.04710144928
11	0.05797101449
13	0.0384057971
15	0.04202898551
17	0.04637681159
19	0.04637681159
21	0.04927536232
23	0.05362318841
25	0.05144927536
27	0.05507246377
29	0.05144927536
31	0.05579710145
33	0.05434782609

35	0.05434782609
37	0.06376811594
39	0.05869565217
41	0.06086956522
43	0.06449275362
45	0.05724637681
47	0.06304347826
49	0.06231884058
51	0.0615942029
53	0.04927536232
55	0.05507246377

Minimum error came at m = 13

A6 c) Values for oob and test error:

m	test error	oob error
1	0.1188405797	0.1935068967
3	0.06739130435	0.1328779882
5	0.05652173913	0.118167975
7	0.05434782609	0.1132892772
8	0.04637681159	0.117739241
9	0.04710144928	0.1194837451
11	0.05797101449	0.108721042
13	0.0384057971	0.1072722166
15	0.04202898551	0.1138362827
17	0.04637681159	0.1080261971
19	0.04637681159	0.1111751748
21	0.04927536232	0.1130675182
23	0.05362318841	0.1068286986
25	0.05144927536	0.1091202082
27	0.05507246377	0.1127718396
29	0.05144927536	0.1107464408
31	0.05579710145	0.1178131607
33	0.05434782609	0.1109829837
35	0.05434782609	0.1067991307

37	0.06376811594	0.1083810115
39	0.05869565217	0.1166156621
41	0.06086956522	0.1137475791
43	0.06449275362	0.1049215712
45	0.05724637681	0.1106577372
47	0.06304347826	0.1082036043
49	0.06231884058	0.1109829837
51	0.0615942029	0.1082479561
53	0.04927536232	0.1138067149
55	0.05507246377	0.1128753271

Graph: