

Andrew Tran

CS1675

Assignment 10 Report

Due: 4/11/19

1a)

Attribute	Score
48	0.319181
25	0.214008
21	0.19097
70	0.189214
65	0.16932
40	0.167345
29	0.165045
19	0.140196
57	0.125454
20	0.121208
24	0.099525
30	0.095024
12	0.085818
47	0.084644
61	0.060677
10	0.057916
34	0.052678
27	0.046216
39	0.046114
41	0.042185

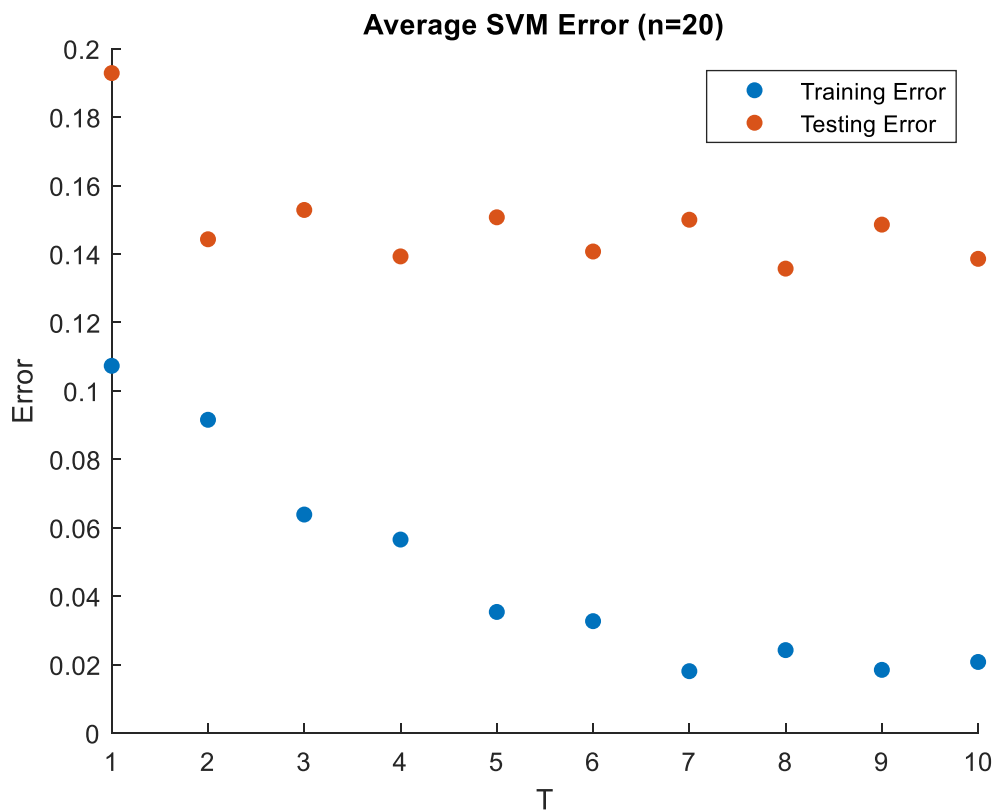
1b)

Attribute	Score
25	0.733986
29	0.683691
11	0.669465
47	0.666084
19	0.631452
34	0.617382
32	0.602107
30	0.602068
9	0.599969
56	0.597054
27	0.595305
60	0.592856
51	0.588075
26	0.587376
53	0.584499
7	0.57968
10	0.570896
61	0.568563
43	0.55667
44	0.542172

The lists are not the same but do share a couple of attributes in the top 20, most notably attribute 25 which is second when using Fisher score and first when using AUROC score. In general, the lists should be similar since both scores are measuring the same thing: how distinguishable the 2 classes are. However, they are calculated differently so we don't expect the lists to be exactly the same.

2a)

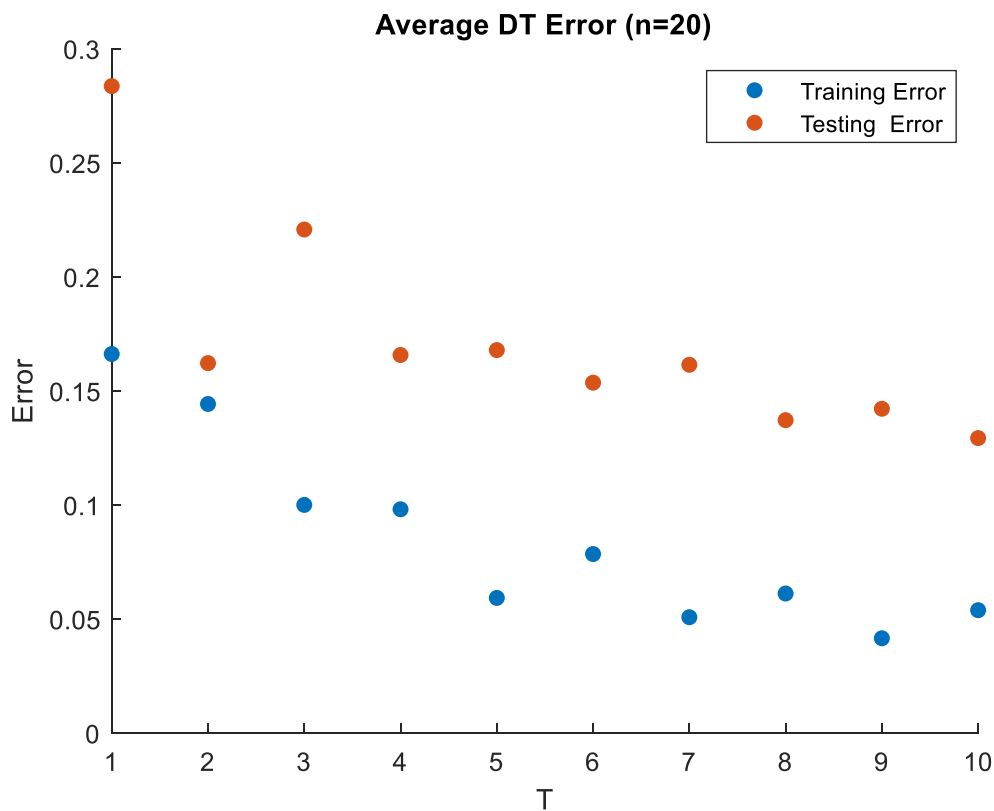
T	Training Error	Testing Error
1(base)	0.107307692	0.192857143
2	0.091538462	0.144285714
3	0.063846154	0.152857143
4	0.056538462	0.139285714
5	0.035384615	0.150714286
6	0.032692308	0.140714286
7	0.018076923	0.15
8	0.024230769	0.135714286
9	0.018461538	0.148571429
10	0.020769231	0.138571429



As the number of models increases, the average classification error decreases in the training set. After the base model ( $T=1$ ) the testing error does not seem to decrease at the same rate as the training error or at all on average.

2b)

T	Training Error	Testing Error
1(base)	0.166153846	0.283571429
2	0.144230769	0.162142857
3	0.1	0.220714286
4	0.098076923	0.165714286
5	0.059230769	0.167857143
6	0.078461538	0.153571429
7	0.050769231	0.161428571
8	0.061153846	0.137142857
9	0.041538462	0.142142857
10	0.053846154	0.129285714



Similar to the Support Vector Machine model, as more models are added bagging decreases the classification error when using the Decision Tree model. The testing error continuous to decrease as T increases unlike the SVM model which plateaued after the base model.