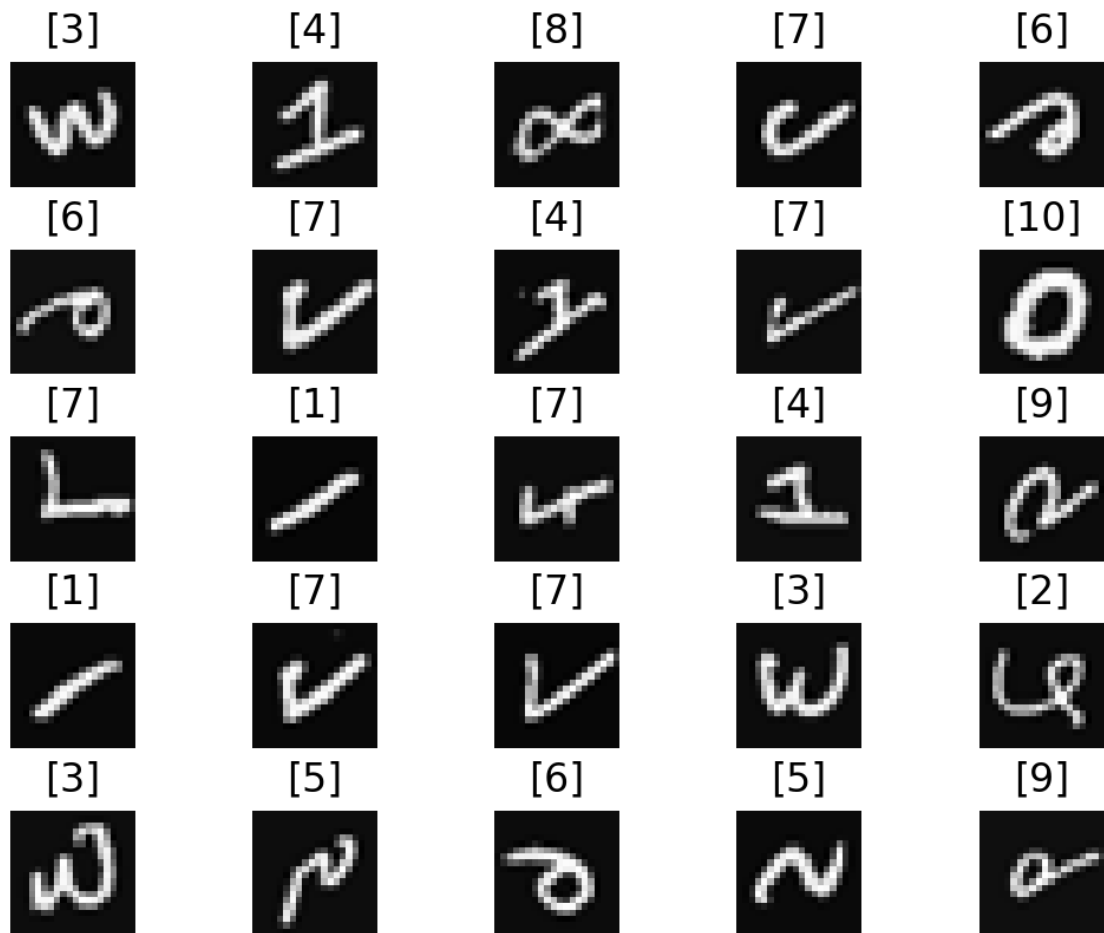Beryl Sin

PeopleSoft ID: 4522433

Problem Set 8

1) a.

b. See ps8.py for the data splitting
c. See ps8.py for bagging implementation

d.

Question 1-D :

Set	Accuracy	Classification Error
X_1	0.983	0.017
X_2	0.949	0.051
X_3	0.949	0.051
X_4	0.94	0.06
X_5	0.942	0.058
X_test	0.904	0.096

e.

Question 1-E :

Set	Accuracy	Classification Error
X_1	0.913	0.087
X_2	0.934	0.066
X_3	0.909	0.091
X_4	0.899	0.101
X_5	0.894	0.106
X_test	0.874	0.126

f.

```
Question 1-F :
```

Set	Accuracy	Classification Error
X_1	0.823	0.177
X_2	0.831	0.169
X_3	0.925	0.075
X_4	0.831	0.169
X_5	0.827	0.173
X_test	0.848	0.152

g.

```
Question 1-G :
```

Set	Accuracy	Classification Error
X_1	0.743	0.257
X_2	0.716	0.284
X_3	0.755	0.245
X_4	1	0
X_5	0.718	0.282
X_test	0.662	0.338

h.

Question 1-H :

Set	Accuracy	Classification Error
X_1	0.913	0.087
X_2	0.931	0.069
X_3	0.934	0.066
X_4	0.921	0.079
X_5	1	0
X_test	0.878	0.122

i.

Question 1-I :

Set	Accuracy	Classification Error
X_1	0.983	0.017
X_2	0.949	0.051
X_3	0.949	0.051
X_4	0.94	0.06
X_5	0.942	0.058
X_test	0.904	0.096

j.
Overall, it seems like the one-vs-all SVM model performed the best, whereas the decision tree classifier performed the worst in terms of accuracy/classification error. The one-vs-all SVM outperformed with the lowest classification error for all sets of data except X_4 and X_5. The decision tree classifier did the best on X_4 since it used X_4 for training, similar to the random forest classifier but for X_5. As can be seen, each model made good predictions on the data set they were trained on.

Bagging did not really improve performance since it yielded the exact same results (same classification errors) as the SVM model.