Beryl Sin

PeopleSoft ID: 4522433

Problem Set 8

1. a.

A group of black squares with white letters

Description automatically generated

b. See ps8.py for the data splitting

c. See ps8.py for bagging implementation

d.

A black and white screen with white text

Description automatically generated

e.

A black and white screen with numbers

Description automatically generated

f.

A black and white screen with numbers and text

Description automatically generated

g.

A black and white screen with white text

Description automatically generated

h.

A black and white screen with numbers and text

Description automatically generated

i.

A black and white screen with white text

Description automatically generated

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.