Report for homework 4

1, logistic regression.

Function implemented in logistic\_train file. Training and testing executed by main file. The plot for training data versus testing data accuracy. The more training data used for training; the higher accuracy achieved. This is reasonable because more data used in training more information used for building the model and higher performance model built.

A close up of a map

Description automatically generated

2, sparse logistic regression.

A close up of a device

Description automatically generated

The feature number decreased with increasing lambda. This phenomenon is reasonable because the larger lambda the more penalty on L-1 norm of weights which promotes less features kept in modeling.

A close up of a map

Description automatically generated

The AUC increased from 0 to 0.1 and decreased later for lambda. The reason is some penalty on weights could decrease model complexity so reduce model overfitting but too much penalty on weights would make model too simple for prediction.

A close up of a map

Description automatically generated

ROC plot also shows 0.1 lambda is the best in the provided lambda values which produces high true positive but low false positive.

https://github.com/xiaoyanLi629/CSE-847-Homework-4