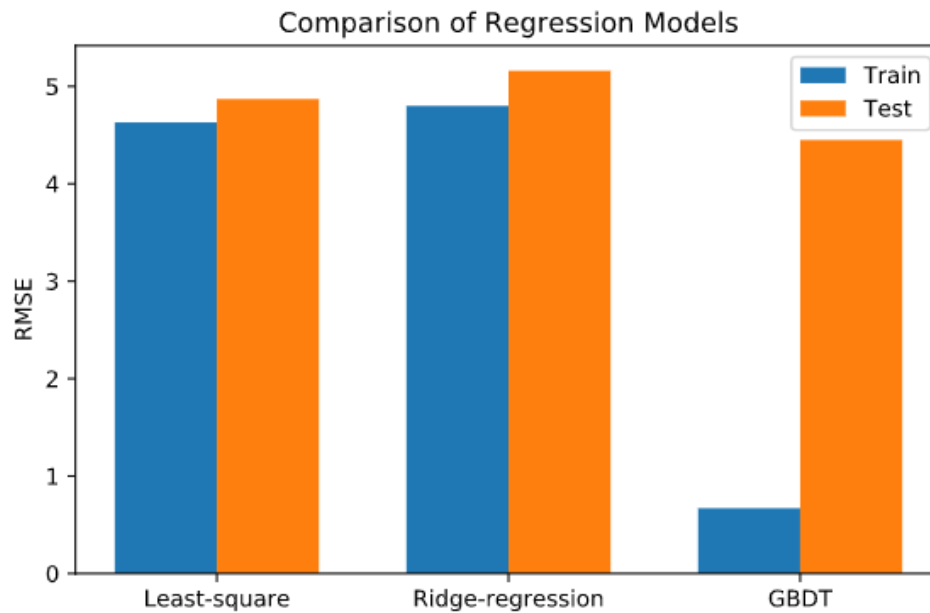Problem 4(e)

Boston Housing Price GBDT

Train RMSE: 0.6733030286580176

Test RMSE: 4.448467175942173

From the above plot it is evident that both in train and test cases for Boston housing data GBDT is performing best than least-square and ridge regression.

Problem 4(f)

Credit-g-dataset:

Train Accuracy: 0.9742857142857143

Test Accuracy: 0.7466666666666667

Breast-cancer dataset:

Train Accuracy: 1.0

Test Accuracy: 0.9707602339181286

Problem 4(g)

For all cases, GBDT's accuracy and RMSE score is better than random forest. For RF each iteration the classifier is trained independently from the rest. Where GBDT is a boosting method, which builds on weak classifiers. It adds a classifier at a time, so that the next classifier is trained to improve the already trained ensemble, that makes the GBDT stronger than another classifier.