RokunuzJahan Rudro
EECS-658
3011173

Accuracy Score:

Linear regression = 0.9666

Polynomial degree 2 = 0.9666

Polynomial degree 3 = 0.91333

Naive Bayesian = 0.96

kNN = 0.94

LDA = 0.9733

QDA = 0.96

SVM = 0.94

DTC = 0.9266

RFC = 0.94

ETC = 0.9466

NN = 0.9733

4.a) Based on the accuracy scores and also the confusion matrix, NN is the best one.

b) NBClassifier assumes a normal distribution which is not how the real world actually works. It also assumes independence among variables but that is also very rare in a real world scenario.

Linear regression assumes linearity between the independent and dependent variables. It is also prone to noise and overfitting. Polynomial regression works well when we can choose a degree but there are also tradeoffs for that.

LDA requires normal distribution assumption on features and predictors. Moreover, it's not useful when few predictors are involved.

kNN takes a very long time for inference in comparison to training. It is also very much prone to overfitting if there are outliers.

RokunuzJahan Rudro
EECS-658
3011173

DTC does a lot of overfitting and has a very high variation in the prediction.

SVM takes a lot of time for training with large datasets. It doesn't perform well with noises in the data.

RTC has a very slow training and is very biassed when dealing with categorical variables.

ET doesn't return a global optimal decision which produces inaccuracies.

7.a) No

b) 80%

c) 1797 x 0.2 = 359

d) 1797 x 0.8 = 1438

e) 64

f) 64

g) 10

h) 0, 1, 2 ...9