Homework 4

Yuan Li, N19728558, NetID: yl6606

Part II: Programming and Questions

1. (a) Answer:

I set $k = 10$, as 10-fold cross validation. And the score as MSE. This part uses Linear regression model. The scores for both the test set and training set are as follow:

$$score_{train} = 21.806183$$
$$score_{test} = 23.636068$$

1. (b) Answer:

I set $k = 10$, as 10-fold cross validation. And the score as MSE. This part uses Ridge regression model. The scores for both the test set and training set are as follow:

lamda	train score	test score
10.0	21.8929011	23.688583
31.62277	22.2854440	24.0178402
100.0	23.7254883	25.2938525
316.22776	28.1665540	29.4572962
1000.0	38.5321487	39.4894933
31622.77660	78.5307424	78.9216211
100000.0	82.4032501	82.7724042
316227.7660	83.7551488	84.1174862
1000000.0	84.1968033	84.5569941
3162277.6601	84.3379310	84.6974441
10000000.0	84.3827076	84.7420065

2. Answer:

In this part, I made a polymoial transformation of degree 2 on data. Then the results become as follow:

For Linear regression model, the scores are:

$$score_{train} = 5.8088208$$
$$score_{test} = 11.8549682$$

For Ridge regression model, the scores are:

lamda	train score	test score
10.0	10.049055	13.476138
31.62277	12.751706	15.829601
100.0	16.222690	18.9800188
316.22776	19.7002536	22.0686923
1000.0	24.2874579	26.2184755
31622.77660	62.0090019	62.6462888
100000.0	74.2875889	74.7379516
316227.7660	80.692551	81.07998618
1000000.0	83.167235	83.5352423
3162277.6601	84.0057958	84.3677657
10000000.0	84.277006	84.6370805

3. Answer:

I will choose Linear regression model with polymoial transformation.

Comparing the test score for each model, Linear regression model with polymoial transformation got the smallest one. Because I used MSE for the score, when the score is smaller, the model is better. So I will choose this model.

The predict price should be 253.156094.