COMP9417 20T1 Assignment 1

Student Name: Minrui Lu

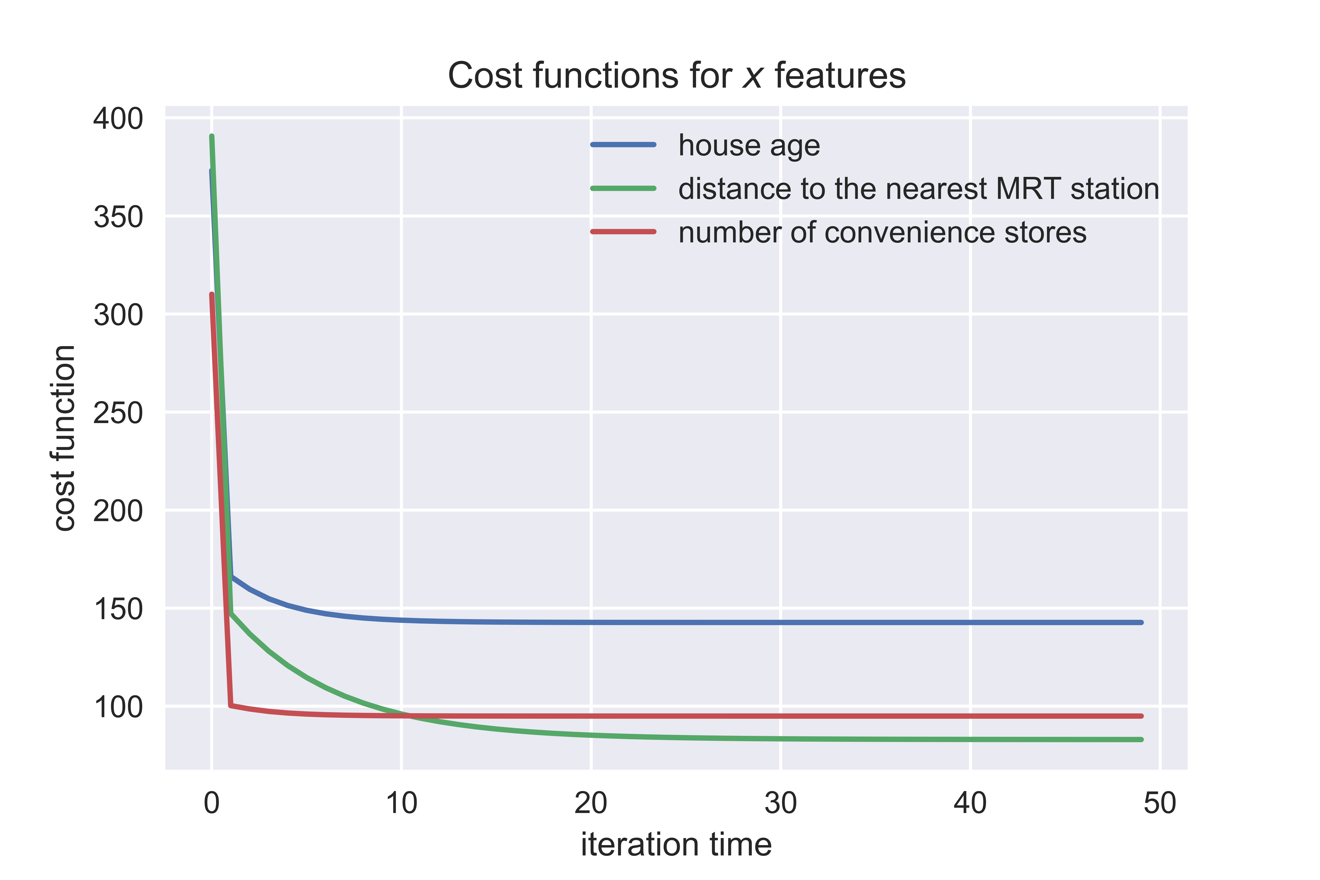
Student Number: 5277884

**Question 1: The 𝜃 parameters () from step 3 when you are using house age feature. (2 marks)**

In step 3, with the iteration times 50 and the training data set row number 300, we have updated the and times respectively. Each time, we update the value by .

From the program output, it can be seen that from step 3, using the house age feature, the value of is 42.54098352098717, and the value of is -10.321581018919572.

**Question 2: A plot, which visualises the change in cost function 𝐽(𝜃) at each iteration. (1 mark)**



**Question 3: RMSE for your training set when you use house age feature. (0.5 mark)**

By calculating the RMSE formula for the house age feature in the training set:

We can get the RMSE for training set when I use house age feature is 12.045471635151399.

**Question 4: RMSE for test set, when you use house age feature. (0.5 mark)**

By calculating the RMSE formula for the house age feature in the test set:

We can get the RMSE for test set when I use house age feature is 16.587314577458564.

**Question 5: RMSE for test set, when you use distance to the station feature. (0.25 mark)**

By calculating the RMSE formula for the distance to the station feature in the test set:

We can get the RMSE for test set when I use distance to the station feature is 12.65187816696171.

**Question 6: RMSE for test set, when you use number of stores feature. (0.25 mark)**

By calculating the RMSE formula for the number of stores feature in the test set:

We can get the RMSE for test set when I use number of stores feature is 14.732079954030375.

**Question 7: Compare the performance of your three models and rank them accordingly. (0.5 mark)**

In order to compare the performance of three models, we need to calculate the difference of RMSE between training model and test model among the three features respectively.

*House Age Feature:*

RMSE for house age training set is 12.045471635151399, RMSE for house age test set is 16.587314577458564, with .

*Distance to Station Feature:*

RMSE for distance to station training set is 9.165812661768193, RMSE for distance to station test set is 12.65187816696171, with .

*Number of Stores Nearby Feature:*

RMSE for number of stores nearby training set is 9.834850879113743, RMSE for number of stores nearby test set is 14.732079954030375, with .

Considering that RMSE is the square root square root of the variance of the residuals, and it indicates how close are the observation data to the predicted values, lower values of RMSE shows better fit.

Therefore, by comparing the RMSE values of test set data, it could be seen that the model fitted by Distance to Station Feature has the best performance, followed by the model fitted by Number of Stores Nearby Feature, with the model fitted by House Age Feature ranking the last.

Namely, the performance ranking is:

***Distance to Station > Number of Stores Nearby > House Age***

The Python code is given below.