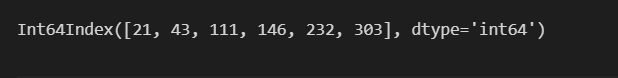
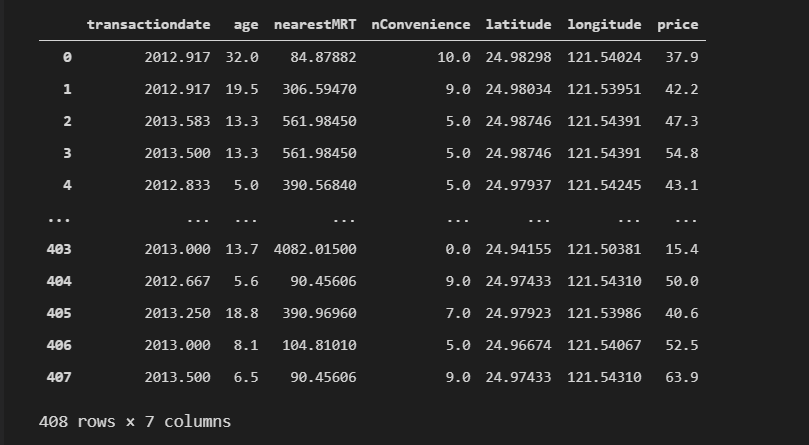
Question 1 (a)

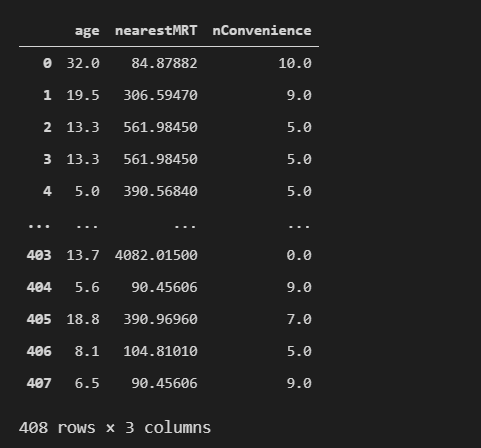
List the indices of the removed data points.



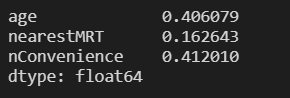
then Remove all rows of the data that contain a missing (‘NA’) value



delete all features from the dataset apart from: age, nearestMRT and nConvenience.



Normalization

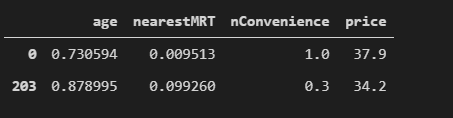


Question 2

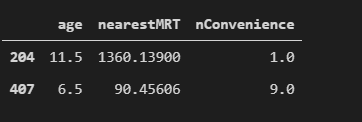
Print out the first and last rows of training set



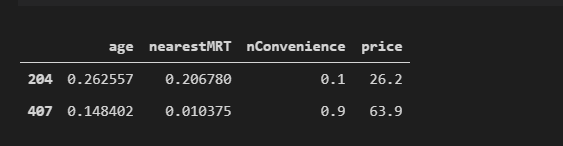
After normalization and add price is



Print out the first and last rows of test set



After normalization and add price is



Question 3

In this question all the = w0+ w1x1+w2x2+w3x3

so

will be

will be and so on(others just change x1 to x2 and x3)

Question 4

GD

i = 0

#iteration number

While not convergence:

#check if convergence

Wx\_change= )for k from 0 to index

Wx = Wx – lr\*Wx\_change/index\*xi

i+=1

#add one iteration time

SGD:

While not convergence:

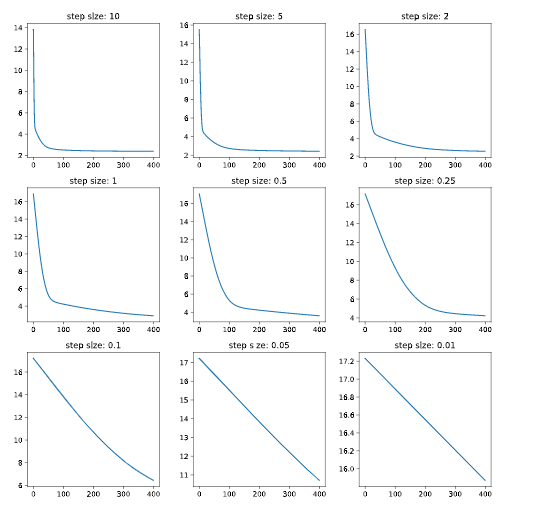
Theta = theta +learning\_rate\*()

i+=1

#add one iteration time

Question 5

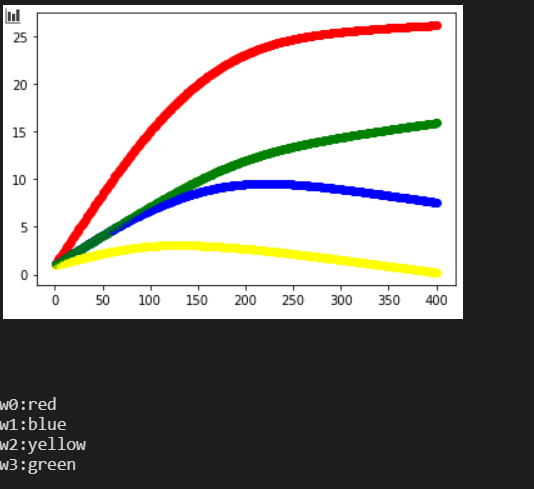
(a)



(b)

for step size I chose 0.5，cause all the lines are smooth so we pick the ones that have the most obvious change but aren't sudden,I think is 0.5s

(c)



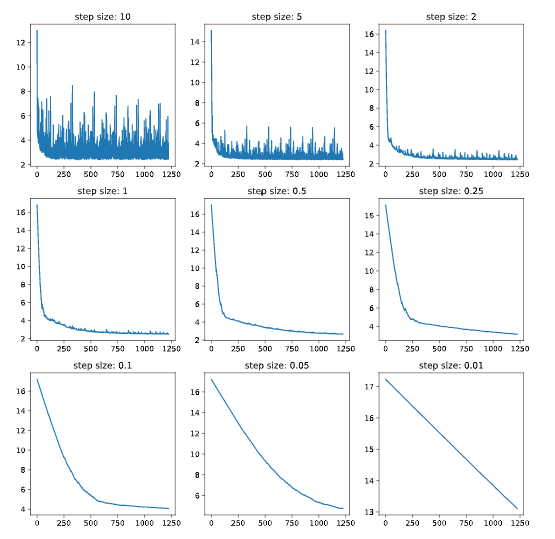
and print out the final weight vector , train loss and test loss





Question 6

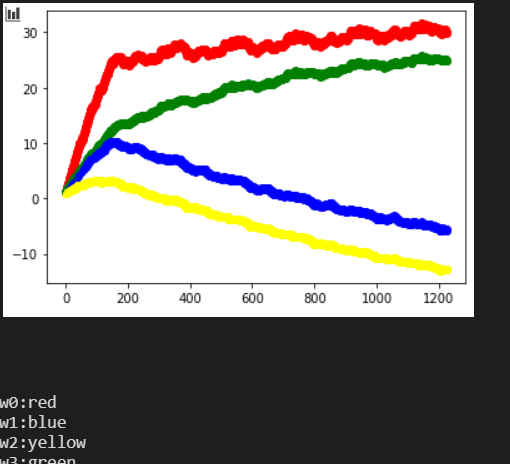
(a)



(b)

for step size I chose 0.1，

firstly since we allways use big step size to check vast amounts of data to Save the number of code operations and small step size for small amounts of data to ensure the accuracy, If step size too small like 0.01 is hard to show the change and also waste too much time but too big step size like 10 is unstable so it has poor accuracy,so i chose '0.1' which is smooth and not cost a lot number of runs

(c)

and print out the final weight vector , train loss and test loss

Question 7

step size for GD and SGD is use to precise the graph,if too big the graph will be unstable,if too small then the change will tends to be constant,so find a suitable step size is important.

After we done question 5 and 6,that we can find the line in question 5 is much smooth than question 6

First we have to think about what kinds of data are GD and SGD oriented,

Gd check all of the data in every iteration to get the mean value of weight that will take long time but the data will be more accurate.(so we always use it to check the dataset that not to big)

SGD choose one random data to calculate the lose that save time but not accurate.(so we always use it to check huge amount dataset)

They are like two extremes, so for the same data set there will always have the better one in GD and SGD，for this question the largest dataset that we have is training data and test data which is 208,it's not a big number of data set,so obviously GD is better,thats the reason why it looks much smooth than SGD