Homework 1

Chang Wang

1.

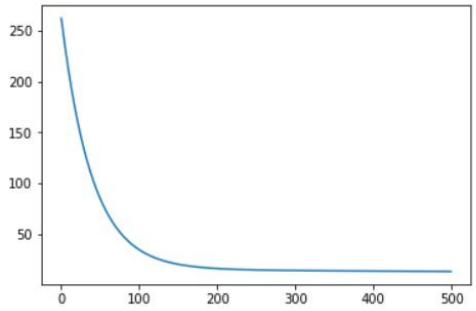
$\theta_0 = 10.1128$

$\theta_1 = 8.2718$

The output of my code is:

```
theta 0: 10.11283406777275
theta 1: 8.271831294479673
```

2.



3.

RMSE = 3.6403

The output of my code is:

```
theta 0: 10.11283406777275
theta 1: 8.271831294479673
TV training RMSE = 3.6403454893687783
```

This is for the training set, when I use TV feature.

4.

RMSE = 3.9086

The output of my code is:

```
TV
theta 0: 10.11283406777275
theta 1: 8.271831294479673
TV training RMSE = 3.6403454893687783
TV test RMSE = 3.9085603448797355
```

This last line is the RMSE for the test set, when I use TV feature.

5.

RMSE = 4.2004

The output of my code is:

```
Radio
theta 0: 10.822122481737317
theta 1: 7.013200578794314
Radio test RMSE = 4.200425795112539
```

This is for the test set, when I use Radio feature.

6.

RMSE = 5.4279

The output of my code is:

```
newspaper
theta 0: 12.96665930015575
theta 1: 3.8374141504698636
newspaper test RMSE = 5.427909854899054
```

This is for the test set, when I use newspaper feature.

7. TV feature has the smallest RMSE and newspaper feature has the largest RMSE, smaller

RMSE means its better.
TV > Radio > newspaper

Here is my code:

```
import matplotlib.pyplot as plt
import csv
import math

learning_rate = 0.01
max_iteration = 500

def get_data(file_name):
    data = []
    with open(file_name) as file:
        csv_file = csv.reader(file)
        for line in csv_file:
            data.append(line)
        return data[1:]

data_line = get_data('Advertising.csv')

def transformation(data):
    for index in range(1,4):
        max_number = 0
        min_number = 9999
        for row in range(len(data)):
            if float(data[row][index]) > max_number:
                max_number = float(data[row][index])
            if float(data[row][index]) < min_number:
                min_number = float(data[row][index])
        for row_a in range(len(data)):
            data[row_a][index] = (float(data[row_a][index]) - min_number)/(max_number - min_number)
    return data

data = transformation(data_line)
training_data = data[:190]
test_data = data[190:]

def training(training_data, row, theta_0, theta_1, learning_rate, max_iteration):
    training_x = []
    training_y = []
    J_theta_list = []
    J_theta = 0
    for index in range(len(training_data)):
```

```
        training_x.append(float(training_data[index][row]))
        training_y.append(float(training_data[index][4]))
    for i in range(max_iteration):
        theta_j_sum_0 = 0
        theta_j_sum_1 = 0
        for j in range(len(training_data)):
            h_theta = theta_0 * 1 + theta_1 * training_x[j]
            J_theta = J_theta + (training_y[j] - h_theta) * (training_y[j] - h_theta)
            theta_j_sum_0 = theta_j_sum_0 + (training_y[j] - h_theta) * 1
            theta_j_sum_1 = theta_j_sum_1 + (training_y[j] - h_theta) * training_x[j]
        J_theta = J_theta * (1/len(training_data))
        J_theta_list.append(J_theta)
        theta_0 = theta_0 + (learning_rate * theta_j_sum_0 * (1/len(training_data)))
        theta_1 = theta_1 + (learning_rate * theta_j_sum_1 * (1/len(training_data)))
    return theta_0, theta_1, J_theta_list


def Evaluation(data, row, theta_0, theta_1):
    training_x = []
    training_y = []
    J_theta = 0
    for index in range(len(data)):
        training_x.append(float(data[index][row]))
        training_y.append(float(data[index][4]))
    for i in range(len(data)):
        h_theta = theta_0 * 1 + theta_1 * training_x[i]
        J_theta = J_theta + (training_y[i] - h_theta) * (training_y[i] - h_theta)
    RMSE = math.sqrt(J_theta * (1 / len(data)))
    return RMSE


row_1 = 1   # 1 is for TV, 2 is for Radio, 3 is for newspaper
theta_0 = -1
theta_1 = -0.5
theta_0, theta_1, cost = training(training_data,row_1,theta_0,theta_1,learning_rate,max_iteration)
print('TV')
print('theta 0:', theta_0)
print('theta 1:', theta_1)
RMSE_1_training = Evaluation(training_data, row_1, theta_0, theta_1)
RMSE_1_test = Evaluation(test_data, row_1, theta_0, theta_1)
print('TV training RMSE =', RMSE_1_training)
print('TV test RMSE =', RMSE_1_test)
plt.plot(cost)
plt.show()


row_2 = 2
```

```
theta_0 = -1
theta_1 = -0.5
theta_0, theta_1, cost_2 = training(training_data,row_2,theta_0,theta_1,learning_rate,max_iteration)
print('Radio')
print('theta 0:', theta_0)
print('theta 1:', theta_1)
RMSE_2_test = Evaluation(test_data, row_2, theta_0, theta_1)
print('Radio test RMSE =', RMSE_2_test)


row_3 = 3
theta_0 = -1
theta_1 = -0.5
theta_0, theta_1, cost_3 = training(training_data,row_3,theta_0,theta_1,learning_rate,max_iteration)
print('newspaper')
print('theta 0:', theta_0)
print('theta 1:', theta_1)
RMSE_3_test = Evaluation(test_data, row_3, theta_0, theta_1)
print('newspaper test RMSE =', RMSE_3_test)
```