Workshop 2: Linear Regression using Gradient Descent.

In this practical, we address the problem of Linear Regression using Gradient Descent. Model answers are provided in the weekly lectures.

Task 1

We will minimize a toy example function:

$$f(x) = X^2$$

which has the following gradient:

$$\nabla = \frac{\partial}{\partial x} f(x) = 2x$$

Run a gradient descent algorithm for four iterations, starting with an arbitrary point: x= -2 and a step size of 0.1 [5 marks]

Example iteration 1:

X (next step) = X – η* ∇(f(X))

Starting point is: X= -2

X (next step) = (-2) – 0.1*(2X)

X (next step) = (-2) – 0.1 (-4)

X (next step) = -2 + (0.1)4

X (next step) = -2 + 0.4

X (next step) = -1.6

Iteration 2:

X (next step) = X – η* (∇)f(X)

X = X – (η)(2X)

X = -1.6 – (0.1)(2(-1.6))

X = -1.6 + 0.32

X = -1.28

Iteration 3:

X (next step) = X – η* (∇)f(X)

X = X – (η)(2X)

X = -1.28 – (0.1)(2(-1.28)

X = -1.28 + 0.256

X = -1.024

Iteration 4:

X (next step) = X – η* (∇)f(X)

X = X – (η)(2X)

X = -1.024 – (0.1)(2(-1.024))

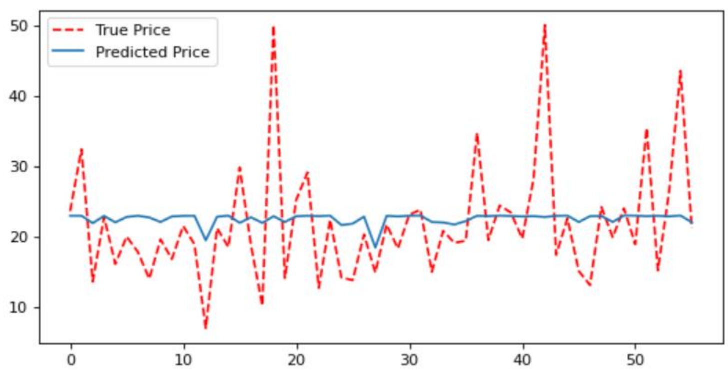
X = -1.024 + 0.2048

X = -0.8192

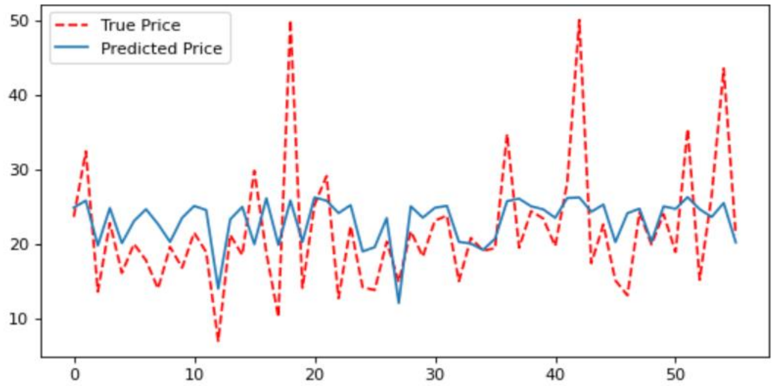
Task 2: Develop, train, and test a Linear Regression Model with Boston housing dataset using Gradient Descent. Goal is to predict the houses price in Boston.

2.1. Please run the jupyter notebook provided on Canvas to plot the predictions after 100, 1000, 10000, and 40000 iterations, and explain the results. [2 marks]

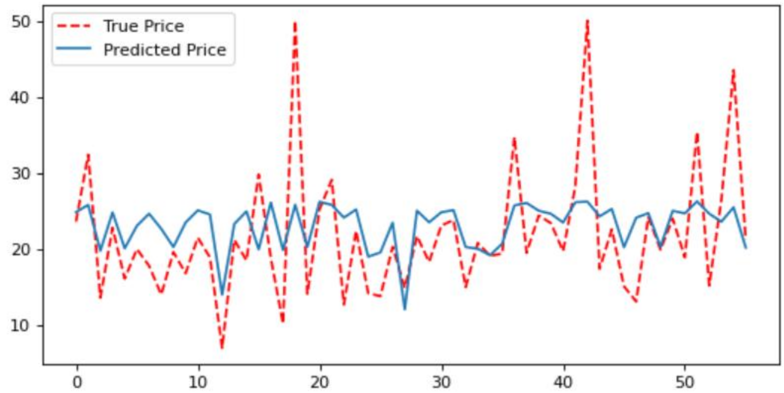
a. After 100 iterations:



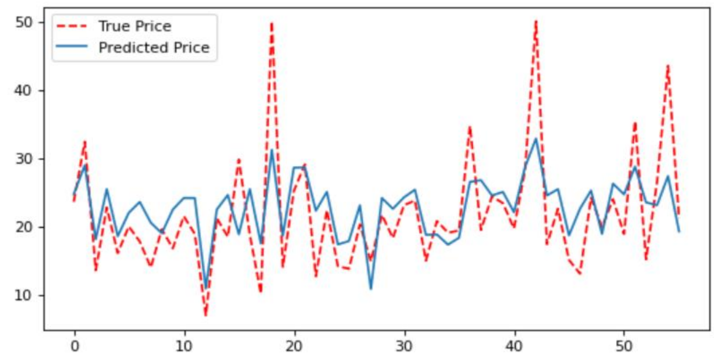
b. After 1000 iterations:



c. After 10000 iterations:



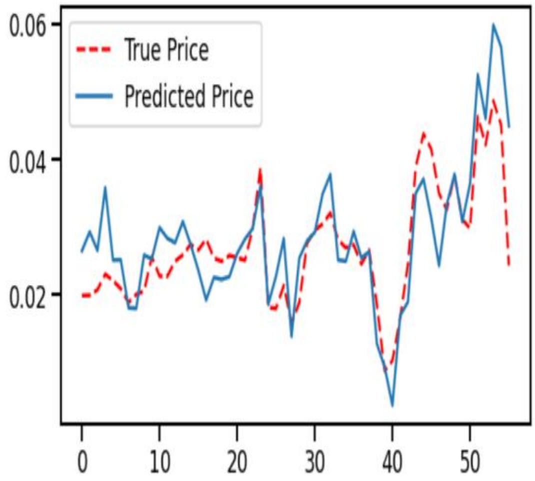
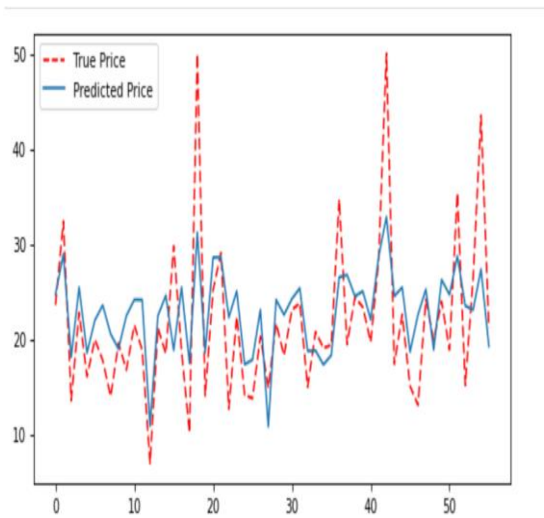
d. After 40000 iterations:



From the plot, it can be seen that the predicted prices improve after each iteration. The first 100 iterations behaved very badly and follows no visible trend as the true price. After 40000 iterations, the model behaves way better and follows the true price trend very closely.

2.2. After 40000 iterations, compare the results with WS1-task 1 and explain the difference in your own words [2 marks]

A**fter 4000 iterations the difference looks something like is:**

Week 1 workshop	Current workshop
	

The current workshop model fares better at predicting the price .

2.3 After 40000 iterations, print network weights and identify the inputs with the highest weight. Compare with the results of WS1-task 2.2 [1 mark]

Use the table below to compare the weights.

Input		Weights Wk 1	Theta_Best (10000 its.) Wk 2
CRIM	*Per capita crime rate by town*	**-0.088559**	**-1.273796**
ZN	*Proportion of residential land zoned for lots over 25000 sq.ft*	**0.035083**	**6.467136**
INDUS	*Proportion of non-retail business acres per town*	**0.013968**	**-1.469107**
CHAS	*Charles River dummy variable*	**1.672166**	**0.060507**
NOX	*Nitric oxides concentration (parts per 10 million)*	**-5.153234**	**0.004018**
RM	*Average number of rooms per dwelling*	**6.849525**	**0.929855**
AGE	*Proportion of owner-occupied units built prior to 1940*	**-0.015810**	**-2.024326**
DIS	*Weighted distances to five Boston employment centres*	**-1.016778**	**0.202150**
RAD	*Index of accessibility to radial highways*	**0.266278**	**-0.150592**

TAX	*Full-value property-tax rate per $10000*	**-0.013787**	**-2.786779**
PTRATIO	*Pupil-teacher ratio by town*	**-0.591467**	**-0.213424**
B	*1000(Bk-0.63)^2 where Bk is the proportion of blacks by town*	**0.014715**	**19.184033**
LSTAT	*% lower status of the population*	**-0.332065**	**-4.213052**

Note:

Input with highest weight week2:   B

Input with highest weight week 1: RM