Information Engineering and Technology Faculty

German University in Cairo

EX1 Take2 Machine Learning

**Name: Yasmine Walid Ibrahim Helal**

**ID: 40-9594**

## Introduction

The demand of machine learning in our everyday life is highly increasing. Machine learning is the preferred approach in Speech recognition, Natural language processing, Computer vision, Medical outcomes analysis, Robot control, Computational biology, Sensor networks, etc. There are two types of machine learning; supervised learning which is known as predictive learning and unsupervised learning which is known as descriptive learning.

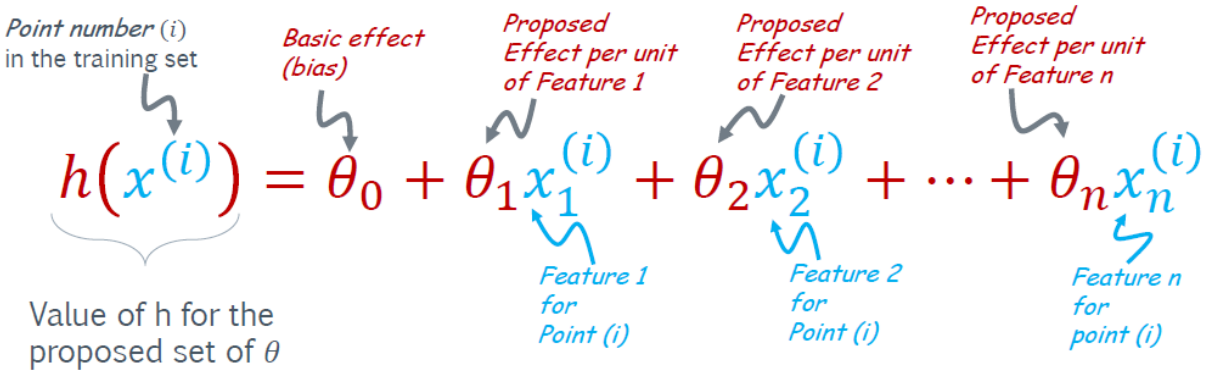
Supervised Learning consists of two types which are regression for continuous value output and classification for discrete value output.

Linear Regression is a regression supervised learning algorithm. Linear Regression is a linear approach to model the relation between a scalar response and one or more explanatory variables (dependent and independent variables). The relationships between the variables are modeled using linear predictor functions whose unknown model parameters are estimated from the data.

In this report we will focus on Linear Regression with multiple variables and its different techniques.

## Linear regression with multiple variables techniques and results:

The hypothesis function of linear regression with multiple variable which means we have n features. So we have (x1,x2,...xn) and (θ1,θ2,...θn).




Here are the steps done in the code:

1- At first, the dataset is loaded. The dataset contains of 18 features and the price.
2- The nan values are dropped
3- The data correlation is calculated and only the features with correlation with the price which is higher than 0.5 are take.

4- The data correlation is calculated and only the features with correlation with the price which is higher than 0.5 are take.
5- Split the data into a Training Set (60%), a Cross Validation (CV) Set (20%) and a Test Set (20%).

$$x_i = \frac{x_i - \mu_i}{S_i}$$

6- Feature Normalization of the data
7- After the featureNormalize function is tested, we now add the intercept term to normalized
8- functions computeCostMulti and gradientDescentMulti to implement the cost function and gradient descent for linear regression with multiple variables.
9- The gradientDecentMulti is calculated for first, second and third degree with different alphas

Graphs of the training data

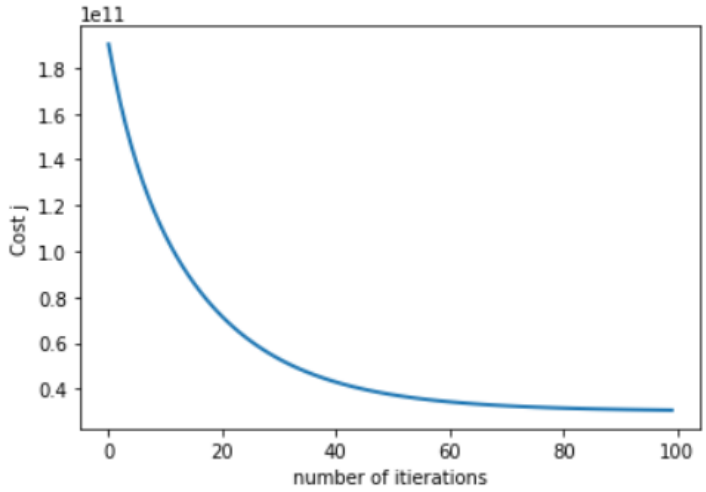



*Figure 1 alpha 0.03 degree 1*

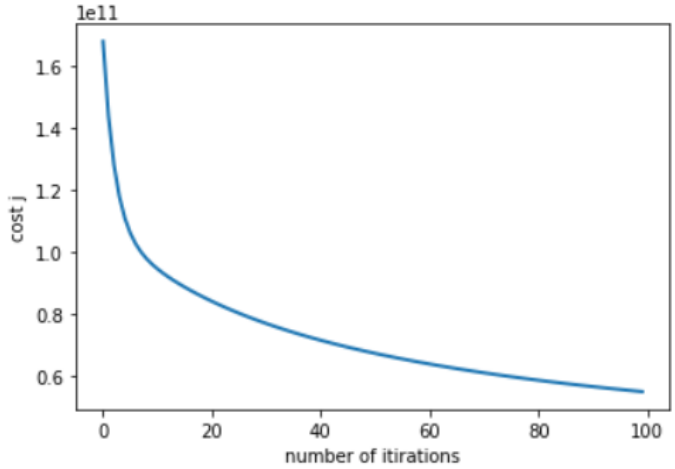



*Figure 2 alpha 0.01 degree 2*

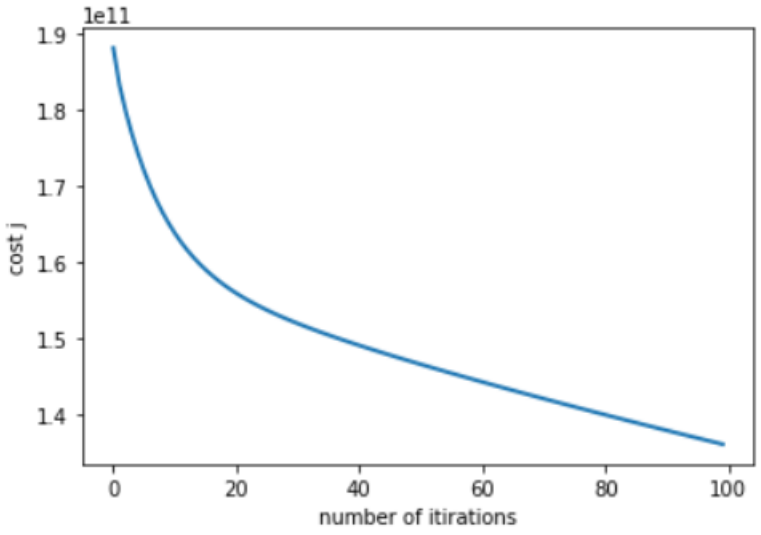



*Figure 3 alpha 0.001 degree 3*

10- Calculating the cost function of the test data and evaluating the error of the model.