BY: (Mohammed Alzarii)

Chapter One in Machine Learning:

(Simple and Multiple) Linear regression

In the following papers we will discuss the core concepts and the basic derivation needed to cover the a comprehensive supervised by regression machine learning models

- Mathematics of linear regression
- Appling simple linear regression models with python, numpy and sklearn
- Appling non linear regression models with python, numpy and sklearn
- Model validation tests

Regression is a way of fitting a function to predict how certain data behave, the simplest form of regression is the linear regression and yes there are non-linear regressions as shown in figure 1 In this explanation paper we will only focus our effort in linear regression as the simplest machine learning.

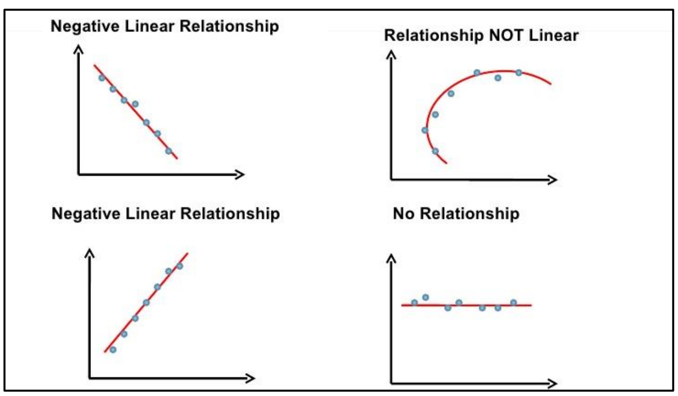

Figure 1: different types of regressions

The Data:

In linear regression the main goal is to find a best fit line, the observed data points have no determined function to follow in fact they are distributed without a linear rule and the task of linear regression is to find a way to estimate their distribution.

There are several examples where linear regression can work for example, the relationship between one independent variable such as temperature to the electricity consumption in the summer or the amount of emitted of CO_2 with respect to the car's engine size we can also use it to estimate the prices of houses in a given city.

In the house's prices however, there can be more than one independent variable that the Type equation here.price will depend on such as the number of rooms, the area of the house or even the age of the house in such a case will switch from single dimension to multidimension linear regression.

Single variable linear regression:

As shown in figure 2 the equation of the predictive line is given by equation (1) as a simple linear relation with an intercept and a slop constant named θ_0 and $\theta_1$respectively

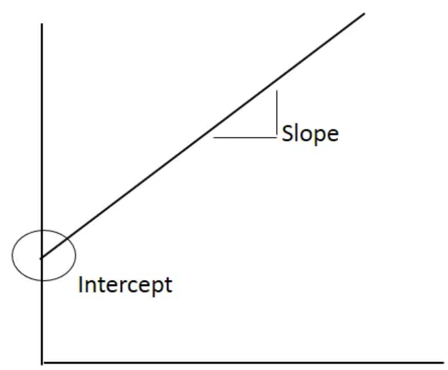

Figure 2: simple linear function.

$$y = \theta_0 + \theta_1 x \text{ ..(1)}$$

The blue dots in figure 1 represents the observed values brought from the datasets each point can have a location in the rectangular coordinate:

$$(\widehat{y_1}, \widehat{x_1}) \ldots\ldots (\widehat{y_n}, \widehat{x_n}) \quad (2)$$

The difference between a point in equation (1) and (2) is a prediction error which AKA residual:

$$E_0 = \widehat{y_0} - y_0 \ (3)$$

If we sum over all residuals (distances between $\hat{y}$ *and* y or the blue points and the point on the line respectively as shown in figure 3:

$$\sum_{i=1}^{n} E_i = \sum_{i=1}^{n} \widehat{y_i} - y_i \ (4)$$

However, due to the negative and positive distances and to avoid the cancelations in the summation we will use the square version of the error also known as residual sum of squares (RSS):

$$\sum_{i=1}^{n} E_i^2 = \sum_{i=1}^{n} (\widehat{y_i} - y_i)^2 \quad (5)$$

from Equations (1) and (5):

$$\sum_{i}^{n} E_{i=1}^2 = \sum_{i=1}^{n} (\widehat{y_i} - (\theta_0 + \theta_1 x_i))^2 \quad (6)$$

$$= \sum_{i=1}^{n} (\widehat{y_i} - \theta_0 - \theta_1 x_i)^2 \ (7)$$

BY: (Mohammed Alzarii)

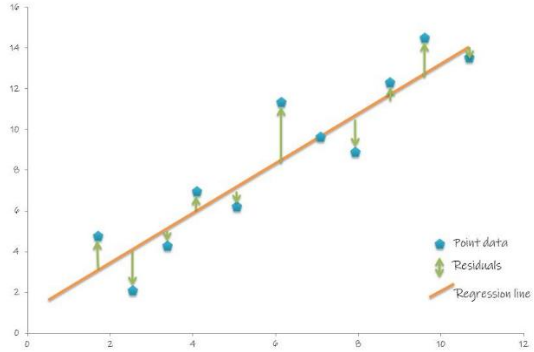

Figure 3: the residual distances.

Our aim is find a way that can minimize E , this can be done by finding the best position of the y-intercept and a slop to fit as a best fit line to the data , what we are going to do next is to find a formulas that enable us to compute the values of these quantities.

Equation (7) is also known as the cost function, and to find its minimum we differentiate with respect to both $\theta_0 and\ \theta_1$ and equate the result to zero

$$\sum_{i=1}^{n} -2\left(\hat{y_i} - \theta_0 - \theta_1\, x_i\right) = 0 \quad (8)$$

$$\sum_{i=1}^{n} -2\left(\hat{y_i} - \theta_0 - \theta_1\, x_i\right)(x_i) = 0 \quad (9)$$

The summation is linear operator hence (8) and (9) becomes:

$$\sum_{i=1}^{n} \hat{y_i} + \sum_{i=1}^{n} \theta_0 + \sum_{i=1}^{n} \theta_1\, x_i = 0 \quad (8)$$

$$\sum_{i=1}^{n} \hat{y_i} x_i + \sum_{i=1}^{n} \theta_0 x_i + \sum_{i=1}^{n} \theta_1\, x_i^2 = 0 \ (9)$$

By evaluating some summations:

$$\sum_{i=1}^{n} \hat{y_i} + n\theta_0 + \theta_1 \sum_{i=1}^{n} x_i = 0 \quad (8)$$

$$\sum_{i=1}^{n} \hat{y_i} x_i + \theta_0 \sum_{i=1}^{n} x_i + \theta_1 \sum_{i=1}^{n} x_i^2 = 0 \ (9)$$