Multiple Linear Regression

Multiple linear regression

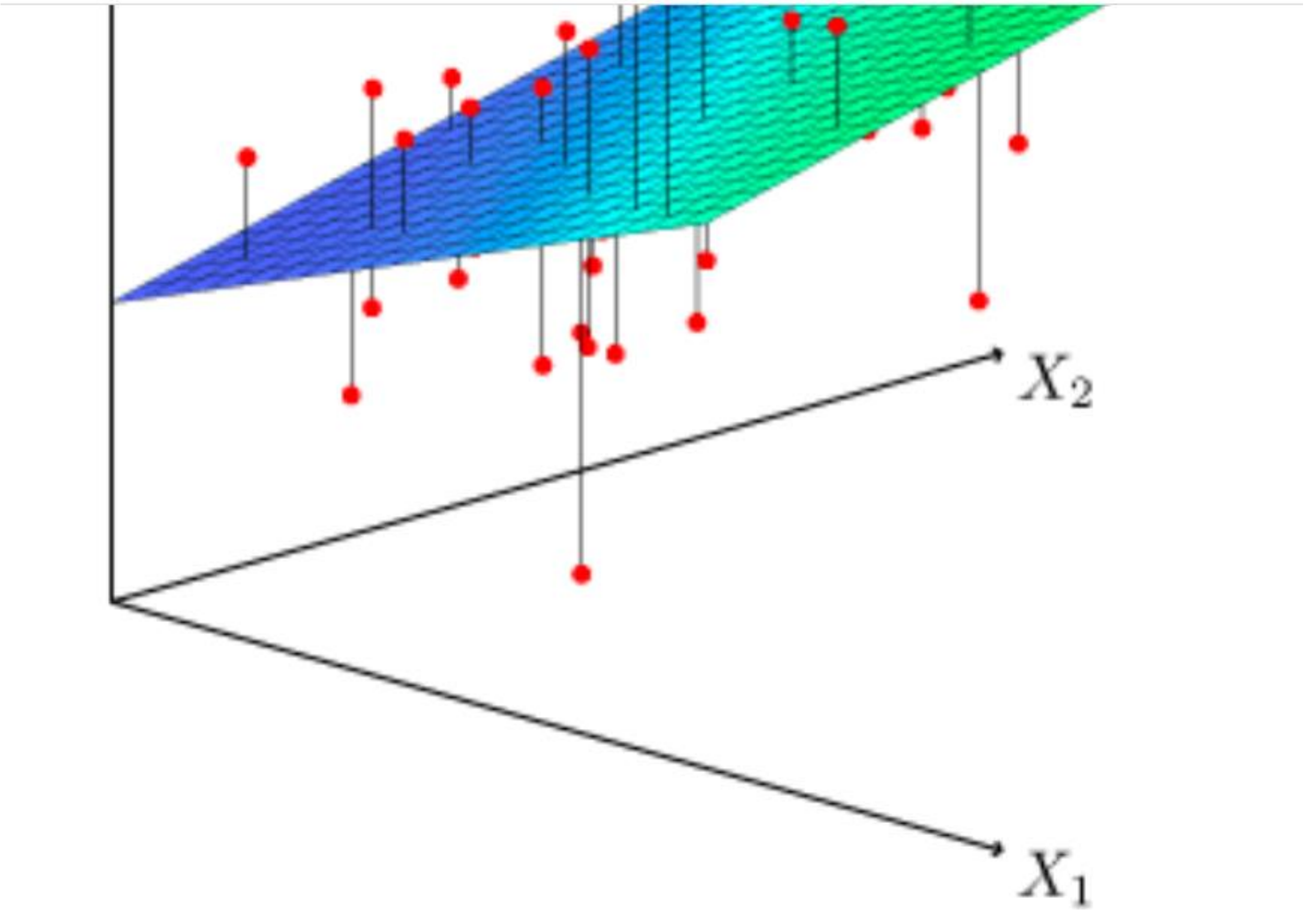
- Multiple linear regression is a method we can use to quantify the relationship between two or more predictor variables and a response variable.
- Multiple linear regression models are defined by the equation

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon$$

Multiple linear regression

- there is more than one independent variables (X1,X2,…,Xp).
- Estimation of the parameters β0,…,βp is based on the same principle as that of simple linear regression, but applied to p dimensions.
- It is thus no longer a question of finding the best line (the one which passes closest to the pairs of points (yi,xi),
- but finding the p-dimensional plane which passes closest to the coordinate points (yi,xi1,…,xip).
- This is done by ***minimizing*** **the sum of the squares of the deviations of the points on the plane**:

Multiple linear regression



Applications

- Businesses can use this method to help evaluate their long-term outlook
- Use multiple linear regression to know:

1. How strong the relationship is between two or more independent variables and one dependent variable

- Ex: how rainfall, temperature, and amount of fertilizer are related to crop growth.

2. The value of the dependent variable at a certain value of the independent variables

Ex: the expected yield of a crop at certain levels of rainfall, temperature, and fertilizer addition

Multiple linear regression

$$b_1 = \frac{(\sum x_2^2)(\sum x_1 y) - (\sum x_1 x_2)(\sum x_2 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

and

$$b_2 = \frac{(\sum x_1^2)(\sum x_2 y) - (\sum x_1 x_2)(\sum x_1 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

Simple linear regression

$$b = \frac{\sum xy}{\sum x^2}$$

$$\overline{Y} = a + b\overline{X}$$

$$\overline{Y} - b\overline{X} = a + b\overline{X} - b\overline{X}$$

$$a = \overline{Y} - b\overline{X}$$

Regression sums

$$\Sigma x_1^2 = \Sigma X_1^2 - (\Sigma X_1)^2 / n$$

$$\Sigma x_2^2 = \Sigma X_2^2 - (\Sigma X_2)^2 / n$$

$$\Sigma x_1 y = \Sigma X_1 y - (\Sigma X_1 \Sigma y) / n$$

$$\Sigma x_2 y = \Sigma X_2 y - (\Sigma X_2 \Sigma y) / n$$

$$\Sigma x_1 x_2 = \Sigma X_1 X_2 - (\Sigma X_1 \Sigma X_2) / n$$

Note: Division is for the second term only

Multiple linear regression

y	X_1	X_2
140	60	22
155	62	25
159	67	24
179	70	20
192	71	15
200	72	14
212	75	14
215	78	11

Multiple linear regression

- Step 1: Calcúlate $X1^2$, $X2^2$, X1y, X2y and X1X2
- Step 2: Calculate Regression Sums
- Step 3: Calculate b0, b1, and b2.
- Step 4: Place b0, b1, and b2 in the estimated linear regression equation.

Multiple linear regression

- How to Interpret a Multiple Linear Regression Equation
- Coefficients
- The coefficients in a multiple regression model represent the change in the dependent variable for every one-unit increase in the independent variable, holding all other variables constant.
- A positive coefficient indicates a positive relationship between the independent variable and the dependent variable,
- a negative coefficient indicates a negative relationship.
- The size of the coefficient reflects the strength of the relationship.

Multiple linear regression

- How to Interpret a Multiple Linear Regression Equation
- $\hat{y} = -6.867 + 3.148x1 – 1.656x2$
- b0 = -6.867. When both predictor variables(x1 & x2) are equal to zero, the mean value for y is -6.867.
- b1 = 3.148. A one unit increase in x1 is associated with a 3.148 unit increase in y, on average, assuming x2 is held constant.
- b2 = -1.656. A one unit increase in x2 is associated with a 1.656 unit decrease in y, on average, assuming x1 is held constant.

Multiple linear regression – Problem Statement

SUBJECT	Y	X_1	X_2
1	-3.7	3	8
2	3.5	4	5
3	2.5	5	7
4	11.5	6	3
5	5.7	2	1