Machine Learning: Multiple Linear Regression

Notes by Mannan Ul Haq (BDS-3C)

Multiple Linear Regression is a statistical method used to model the relationship between a dependent variable (Y) and two or more independent variables (X_1, X_2, X_3, ... X_p), where "p" represents the number of independent variables. It extends the concept of simple linear regression, where there is only one independent variable, to scenarios where multiple factors influence the dependent variable. In multiple linear regression, the goal is to find a linear equation that best fits the data by considering all the independent variables simultaneously. The technique enables analysts to determine the variation of the model and the relative contribution of each independent variable in the total variance.

For a multiple regression with **"p"** independent variables, the regression equation takes the form:

$$\mathbf{Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_p X_p}$$

- Y represents the dependent variable.
- X_1, X_2, X_3, ..., X_p are the independent variables.
- β_0 is the intercept, representing the value of Y when all independent variables are zero.
- β_1, β_2, β_3, ..., β_p are the regression coefficients, indicating the change in Y for a one-unit change in the corresponding independent variable, while holding all other variables constant.

Finding Multiple Linear Regression Equation:

① Suppose we have to find Multiple-Linear Equation of X_1 on X_2 & X_3

Here, X_1 = Dependent variable

X_2 & X_3 = Independent variables

Equation of Regression Line:

$$X_1 = a_{1.23} + b_{12.3} X_2 + b_{13.2} X_3$$

Here, $b_{12.3}$ & $b_{13.2}$ are partial regression coefficients

We can find the value of constants $a_{1.23}$, $b_{12.3}$ and $b_{13.2}$ are obtained by solving the following three equations simultaneously:

$$\sum X_1 = N a_{1.23} + b_{12.3} \sum X_2 + b_{13.2} \sum X_3 \rightarrow \text{(i)}$$

$$\sum X_1 X_2 = a_{1.23} \sum X_2 + b_{12.3} \sum (X_2)^2 + b_{13.2} \sum X_2 X_3 \rightarrow \text{(ii)}$$

$$\sum X_1 X_3 = a_{1.23} \sum X_3 + b_{12.3} \sum X_2 X_3 + b_{13.2} (X_3)^2 \rightarrow \text{(iii)}$$

Consider, we have given the following data for X_1, X_2 and X_3:

X_1	4	6	7	9	13	15
X_2	15	12	8	6	4	3
X_3	30	24	20	14	10	4

Here, N = 6

X_1	X_2	X_3	X_1X_2	X_1X_3	X_2X_3	X_2^2	X_3^2
4	15	30	60	120	450	225	900
6	12	24	72	144	288	144	576
7	8	20	56	140	160	64	400
9	6	14	54	126	84	36	196
13	4	10	52	130	40	16	100
15	3	4	45	60	12	9	16
$\Sigma X_1 = 54$	$\Sigma X_2 = 48$	$\Sigma X_3 = 102$	$\Sigma X_1X_2 = 339$	$\Sigma X_1X_3 = 720$	$\Sigma X_2X_3 = 1034$	$\Sigma X_2^2 = 494$	$\Sigma X_3^2 = 2188$

Putting the respective values in eq (i), (ii) & (iii):

$$54 = 6a_{1.23} + 48\,b_{12.3} + 102\,b_{13.2} \rightarrow (iv)$$

$$339 = 48\,a_{1.23} + 494\,b_{12.3} + 1034\,b_{13.2} \rightarrow (v)$$

$$720 = 102\,a_{1.23} + 1034\,b_{12.3} + 2188\,b_{13.2} \rightarrow (vi)$$

Now using Calculator to Solve these Simultaneous equations:

We get,

$a_{1.23} = 16.4776$

$b_{12.3} = 0.3899$

$b_{13.2} = -0.6233$

So, Multiple Linear Regression Equation is:

$$X_1 = 16.479 + 0.389\,X_2 - 0.623\,X_3$$

Note that the coefficient for X_3 is negative, indicating a negative relationship between X_1 and X_3, while the coefficient for X_2 is positive, indicating a positive relationship.

② Suppose we have to find Multiple-Linear Equation of X_2 on X_1 & X_3

Here, X_2 = Dependent Variable

X_1 & X_3 = Independent Variables

Equation of Regression Line:

$$X_2 = a_{2.13} + b_{21.3} X_1 + b_{23.1} X_3$$

Here, $b_{21.3}$ & $b_{23.1}$ are partial regression coefficients

Simultaneous Equations:

$$\sum X_2 = N a_{2.13} + b_{21.3} \sum X_1 + b_{23.1} \sum X_3 \rightarrow (i)$$

$$\sum X_2 X_1 = a_{2.13} \sum X_1 + b_{21.3} \sum (X_1)^2 + b_{23.1} \sum X_3 X_1 \rightarrow (ii)$$

$$\sum X_2 X_3 = a_{2.13} \sum X_3 + b_{21.3} \sum X_1 X_3 + b_{23.1} \sum (X_3)^2 \rightarrow (iii)$$

③ Suppose we have to find Multiple-Liner Equation of X_3 on X_1 & X_2

Here, X_3 = Dependent Variable

X_1 & X_2 = Independent Variables

Equation of Regression Line:

$$X_3 = a_{3.12} + b_{31.2} X_1 + b_{32.1} X_2$$

Here, $b_{31.2}$ & $b_{32.1}$ are partial regression coefficients

Simultaneous Equations:

$$\sum X_3 = N a_{3.12} + b_{31.2} \sum X_1 + b_{32.1} \sum X_2 \rightarrow (i)$$

$$\sum X_3 X_1 = a_{3.12} \sum X_1 + b_{31.2} \sum (X_1)^2 + b_{32.1} \sum X_2 X_1 \rightarrow (ii)$$

$$\sum X_3 X_2 = a_{3.12} \sum X_2 + b_{31.2} \sum X_1 X_2 + b_{32.1} \sum (X_2)^2 \rightarrow (iii)$$