

Multiple Linear Regression

①

It is a statistical technique that uses several explanatory variables to predict the outcome of a response variable by fitting a linear equation. It is an extension of linear regression that uses one explanatory variable.

→ Every value of independent variable x is associated with a value of dependent variable y .

→ The purpose of a multiple regression is to find an equation that best predicts the y variable (dependent) as a linear function of x (many independent) variables.

The formulated equation of multiple linear regression

$$\hat{y}_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + e.$$

where y_i = dependent variable

x_i = explanatory variables

β_0 = y -intercept (constant-term)

β_p = slope coefficients for each explanatory variable.

e = the model's error term (Residuals).

for $i = 1 \dots n$ observations.

General linear model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \epsilon$$

$\epsilon \rightarrow$ error

$\beta_0 \rightarrow$ y-intercept

$\beta_1, \beta_2, \dots, \beta_p \rightarrow$ slope parameters

$x_1, x_2, \dots, x_p \rightarrow$ independent variables

$y \rightarrow$ dependent variable.

for i th observation,

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon_i.$$

The observations are represented as in Table.

Observation number	Response y	Predictors					
		x_1	x_2	x_3	$\dots$	x_p	
1	y_1	x_{11}	x_{12}	x_{13}	$\dots$	x_{1p}	
2	y_2	x_{21}	x_{22}	x_{23}	$\dots$	x_{2p}	
3	y_3	x_{31}	x_{32}	x_{33}	$\dots$	x_{3p}	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
n	y_n	x_{n1}	x_{n2}	x_{n3}	$\dots$	x_{np}	

To fit best model, need to minimize sum of squares of the errors.

using matrix notations:

Slopes $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$ $\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_p \end{bmatrix}$ Residual (errors)

matrix form of multiple linear regression is given by. (2)

$$\hat{y} = X\beta + E.$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

where $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ $X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$ $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$ & $E = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$

which is equivalent to.

$$y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_p x_{1p} + e_1$$

$$\beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_p x_{2p} + e_2$$

$$\vdots$$

$$\beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_p x_{np} + e_n$$

ie $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + e_i$

Estimation of Model Parameters - $e_i = y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij}$

$$E = \hat{y} - X\beta \quad \text{ie} \quad y - \hat{y} \quad \text{ie} \quad \begin{bmatrix} y_1 - \hat{y}_1 \\ y_2 - \hat{y}_2 \\ \vdots \\ y_n - \hat{y}_n \end{bmatrix} = y_i - \sum_{j=0}^p \beta_j x_{ij}$$

To find residual sum of squares, so that it fits the model.

$$RSS = \sum_{i=1}^n e_i^2 \quad \sum_{i=1}^n e_i^2 = \sum_{i=1}^n \left[y_i - \sum_{j=0}^p \beta_j x_{ij} \right]^2$$

which is also written as $Res = e^T e$.
Choose β in such a way that SSE is minimize.

~~Here~~ Now to estimate β :-

Thus, ^{now} we can substitute e with $y - \hat{y}$

$$\Rightarrow \text{Res} = (y - \hat{y})^T (y - \hat{y})$$

$$\Rightarrow \text{Res} = (y - x\beta)^T (y - x\beta)$$

where $\hat{y} = y \times \beta$

$$= (y^T - \beta^T x^T) (y - x\beta)$$

multiplying the terms we get,

where $(AB)^T = B^T A^T$.

$$= y^T y - y^T x \beta - \beta^T x^T y + \beta^T x^T x \beta$$

— (A)

To ~~reduce~~ minimize error, take derivative of Res w.r.t β . and slope is zero at that point.

$$\therefore \frac{\partial (\text{RSS})}{\partial \beta} = \frac{\partial (y^T y - y^T x \beta - \beta^T x^T y + \beta^T x^T x \beta)}{\partial \beta} = 0.$$

$$\frac{\partial (y^T y)}{\partial \beta} - \frac{\partial (y^T x \beta)}{\partial \beta} - \frac{\partial (\beta^T x^T y)}{\partial \beta} + \frac{\partial (\beta^T x^T x \beta)}{\partial \beta} = 0.$$

$$0 - y^T x - (x^T y)^T + 2 \beta^T x^T x = 0.$$

— (B)

where, $\frac{\partial y}{\partial a}$ for $y=A=0$.

$$\frac{\partial y}{\partial a} \text{ at } y=Aa=A.$$

$$\frac{\partial y}{\partial a} \text{ at } y=a-A=A^T.$$

$$\frac{\partial y}{\partial a} \text{ at } y=a^T A a = 2a^T A.$$

are used for matrix differentiation where $a = n \times 1$ matrix and $A = n \times n$ matrix and A is independent of a .

$\therefore$ now equation (B) can be rewritten as

$$0 - y^T x - y^T x + 2\beta^T x^T x = 0$$

$$\Rightarrow 2\beta^T x^T x = 2y^T x.$$

$$\Rightarrow \beta^T x^T x = y^T x.$$

$$\beta^T = y^T x (x^T x)^{-1}$$

multiple $(x^T x)^{-1}$ on both sides

~~$$\beta^T = (x^T x)^{-1} x^T y$$~~

$$\boxed{\beta = (x^T x)^{-1} x^T y}$$

[taking Transpose of β , we get]

Thus for given set of features, $x_1, x_2, \dots, x_n$ & dependent variable y , using least squares,

$$\beta = (x^T x)^{-1} x^T y,$$

Finally, $\boxed{\hat{y} = x\beta}$

is evaluated.

for set of independent variables. where x is a matrix & β is slope.