Linear Regression
(Least Square Error Fit)

Dr. JASMEET SINGH
ASSISTANT PROFESSOR, CSED
TIET, PATIALA

Linear Regression

- In machine learning and statistics, regression attempts to determine the strength and character of the relationship between one dependent variable (usually denoted by Y) and a series of other variables (known as independent variables).
- Mathematically, regression analysis uses an algorithm to learn the mapping function from the input variables to the output variable (Y) i.e. **Y = f(x) where Y is a continuous or real valued variable.**
- Regression is said to be linear regression if the output dependent variable is a linear function of the input variables.

Regression Example

- **House Value Prediction**- The example below shows that the price variable (output dependent continuous variable) depends upon various input (independent) variables such as plot size, number of bedrooms, covered area, granite flooring, distance from city, age, upgraded kitchen, etc.

Input Attributes

Output or Class

Instances

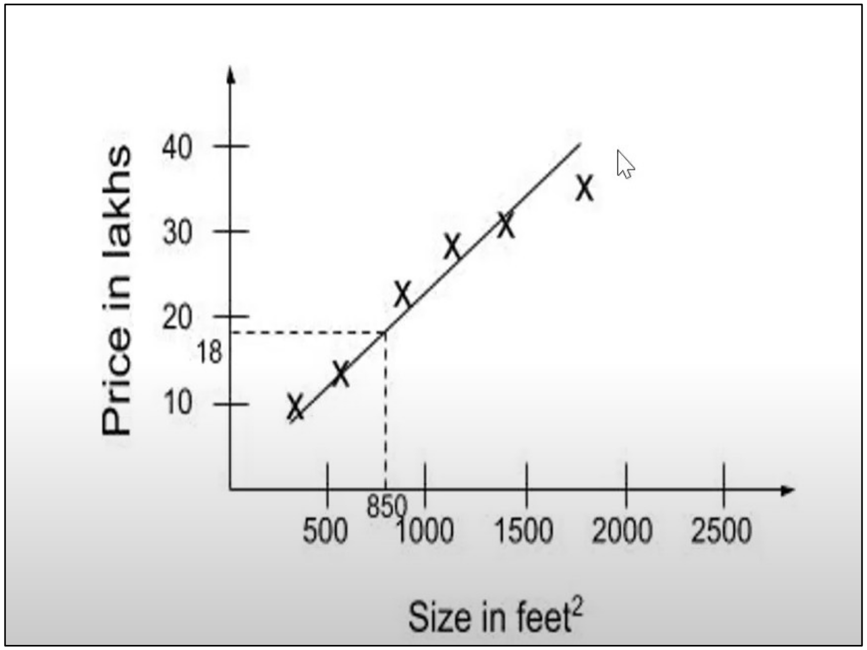
Plot Size	Number of Bedrooms	Covered Area in yards	Granite Flooring	Upgraded Kitchen	Distance from City in Km	Age of flat in years	Price in lakhs
500	3	150	Y	Y	2	2	70
1000	2	250	Y	Y	1	1	140
1800	4	320	N	Y	2	1	200
300	2	130	Y	Y	3	2	60
2000	4	500	Y	N	5	3	200
250	3	160	N	N	1	2	60

Simple Linear Regression (SLR)

- Simple linear regression is a linear regression model with a single explanatory variable.
- It concerns two-dimensional sample points with one independent variable and one dependent variable and finds a linear function (a non-vertical straight line) that, as accurately as possible, predicts the dependent variable values as a function of the independent variable.
- The adjective *simple* refers to the fact that the outcome variable is related to a single predictor.

Simple Linear Regression (SLR) Contd....

- Simple linear regression finds a linear function (a non-vertical straight line) that, as accurately as possible, predicts the dependent variable values as a function of the independent variable.
- For instance, in the house price predicting problem (with only one input variable-plot size), a linear regressor will fit a straight line with x-axis representing plot size and y-axis representing price.



Multiple Linear Regression (MLR)

- Multiple regression models describe how a single response variable Y depends linearly on a number of predictor variables.
- Examples:
  - The selling price of a house can depend on the desirability of the location, the number of bedrooms, the number of bathrooms, the year the house was built, the square footage of the lot and a number of other factors.
  - The height of a child can depend on the height of the mother, the height of the father, nutrition, and environmental factors.

Multiple Linear Regression Model

- A multiple linear regression model with k independent predictor variables $x_1, x_2 ..., x_k$ predicts the output variable as:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots \dots \dots \dots .. + \beta_k x_k$$

- There is always some error (regression residual) in predicting the values, i.e.

$$actual\ value_i = predicted\ value_i + error$$

$$y_i = \hat{y}_i + \epsilon_i$$

$$\hat{y}_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots \dots \dots \dots .. + \beta_k x_{ik} + \epsilon_i$$

The total error can be computed from all the values in dataset i.e. i=1,2,….,n

$$Total\ Error = \sum_{i=1}^{n} \epsilon_i = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ij} \right) \quad (7)$$

Multiple Linear Regression Model

- Equation (7) presented in the previous slide, can be represented in matrix form as:

$$\epsilon = Y - X\beta$$

- Where $\epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \vdots \\ \epsilon_n \end{bmatrix}$; y= $\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{bmatrix}$; β = $\begin{bmatrix} \beta_0 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$

and $X = \begin{bmatrix} 1 & x_{11} & x_{12} \cdots & x_{1k} \\ 1 & x_{21} & x_{22} \cdots & x_{2k} \\ \vdots & \vdots & \vdots \quad \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \cdots & x_{nk} \end{bmatrix}$

Least Square Error Fit for MLR

- According to Least Square Error method, we have to find the values of the matrix β for which total square error is minimum.

$$Total\ Square\ Error = J(\beta) = \sum_{i=1}^{n} \epsilon_i^2 = \epsilon^T \epsilon$$

$$= (Y - X\beta)^T (y - X\beta)$$

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

$$J(\beta) = y^T y - \beta^T X^T y - y^T X\beta + \beta^T X^T\ X\beta$$

$$\boldsymbol{J(\beta) = y^T y - 2y^T X\beta + \beta^T X^T\ X\beta}$$

[Because $y^T X\beta$ and $\beta^T X^T y$ is always equal with only one entry]

- The square error function is minimized using **second derivative test.**

Least Square Error Fit for MLR

- **Step 1: Compute the partial derivate of J($\boldsymbol{\beta}$) w.r.t $\boldsymbol{\beta}$**

$$\frac{\partial J(\beta)}{\partial \beta} = \frac{\partial(y^T y - 2y^T X\beta + \beta^T X^T X\beta)}{\partial \beta}$$

$$= \frac{\partial y^T y}{\partial \beta} - \frac{\partial 2y^T X\beta}{\partial \beta} + \frac{\partial \beta^T X^T X\beta}{\partial \beta}$$

$$= 0 - 2X^T y \frac{\partial \beta}{\partial \beta} + \frac{\partial \beta^T X^T X\beta}{\partial \beta}$$

$$[Because \ \frac{\partial AX}{\partial X} = A^T]$$

$$= -2X^T y + 2X^T X\beta$$

$$[Because \ \frac{\partial X^T AX}{\partial X} = 2AX]$$

Least Square Error Fit for MLR

- **Step 2: Compute $\boldsymbol{\beta}$^ for $\boldsymbol{\beta}$ for which $\frac{\partial \mathbf{J}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{0}$**

$$-2X^T y + 2X^T X\beta\hat{} = 0$$

$$X^T X\beta\hat{}=X^T y$$

$$\boldsymbol{\beta}\hat{}=(\boldsymbol{X}^T \boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

- **Step 3: Compute $\frac{\partial^2 \mathbf{J}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2}$ and prove it to be minimum for $\boldsymbol{\beta}$^**

$$\frac{\partial^2 \mathrm{J}(\beta)}{\partial \beta^2} = \frac{\partial(-2X^T y + 2X^T X\beta)}{\partial \beta} = 0 + 2\ XX^T$$

Being a covariance matrix it is a positive definite matrix. Hence J is minimum when $\boldsymbol{\beta}\hat{}=(\boldsymbol{X}^T \boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$

Least Square Error Fit for MLR- Example

Example: The Delivery Times Data A soft drink bottler is analyzing the vending machine serving routes in his distribution system. He is interested in predicting the time required by the distribution driver to service the vending machines in an outlet. It has been suggested that the two most important variables influencing delivery time (y in min) are the number of cases of product stocked (x_1) and the distance walked by the driver (x_2 in feet). 3 observations on delivery times, cases stocked and walking times have been recorded.

number of cases of product stocked (x_1)	the distance walked by the driver (x_2)	Delivery time (in min) y
7	560	16.68
3	220	11.50
3	340	12.03

(a) Fit a multiple regression line using least square error fit.

(b) Compute the delivery time when 4 cases are stocked and the distance traveled by driver is 80 feet.

Least Square Error Fit for MLR- Example Soln

- The multiple linear regression equation is: y = $\beta_1\hat{}+\beta_2\hat{}x_1+\beta_3\hat{}x_2$

Where $\beta_1\hat{}, \beta_2\hat{}, \beta_3\hat{}$ or $\beta\hat{}=\begin{bmatrix}\beta_1\hat{}\\ \beta_2\hat{}\\ \beta_3\hat{}\end{bmatrix}$ are regression coefficients for line of best fit.

We know, $\boldsymbol{\beta\hat{}}=(\boldsymbol{X^T\ X})^{-1}\boldsymbol{X^T y}$

$$X=\begin{bmatrix}1 & 7 & 560\\ 1 & 3 & 220\\ 1 & 3 & 340\end{bmatrix} and\ X^T=\begin{bmatrix}1 & 1 & 1\\ 7 & 3 & 3\\ 560 & 220 & 340\end{bmatrix}$$

$$X^TX=\begin{bmatrix}3 & 13 & 1120\\ 13 & 67 & 5600\\ 1120 & 5600 & 477600\end{bmatrix}$$

Least Square Error Fit for MLR- Example Soln

$$(X^T X)^{-1} = \begin{bmatrix} 799/288 & 79/288 & -7/720 \\ 79/288 & 223/288 & -7/720 \\ -7/720 & -7/720 & 1/7200 \end{bmatrix}$$

$$\hat{\beta} = (X^T X)^{-1} X^T y = \begin{bmatrix} 7.7696 \\ 0.9196 \\ 0.0044 \end{bmatrix}$$

The line of best fit is, $y = 7.7696 + 0.9196x_1 + 0.0044x_2$

When $x_1 = 4$, $x_2 = 80$

$y = 7.7696 + 0.9196\ X\ 4 + 0.0044\ X\ 80 = 11.80$ min