Linear Regression
(Least Square Error Fit for Simple Linear Regression)

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

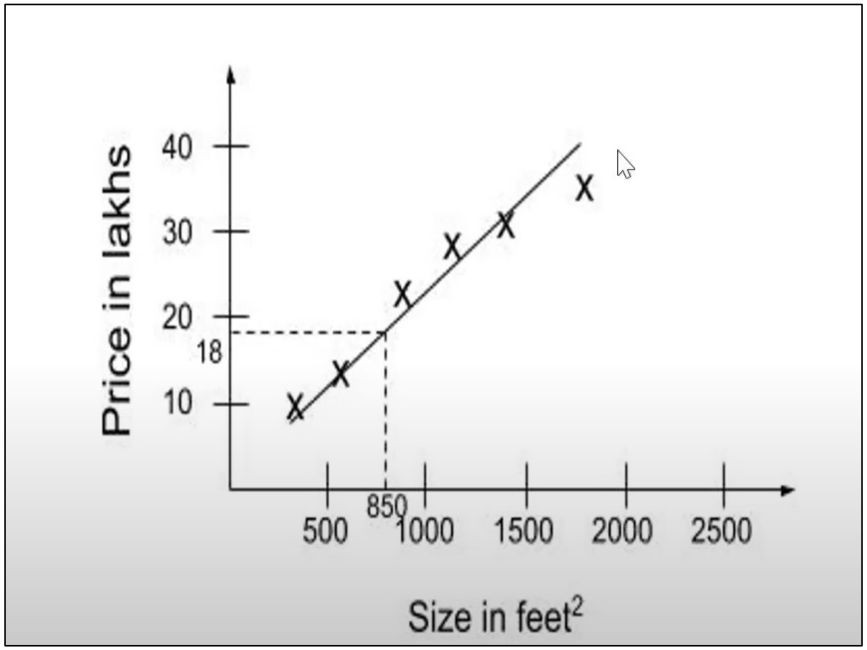
Plot Size	Number of Bedrooms	Covered Area in yards	Granite Flooring	Upgraded Kitchen	Distance from City in Km	Age of flat in years	Price in lakhs
500	3	150	Y	Y	2	2	70
1000	2	250	Y	Y	1	1	140
1800	4	320	N	Y	2	1	200
300	2	130	Y	Y	3	2	60
2000	4	500	Y	N	5	3	200
250	3	160	N	N	1	2	60

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.



Fitting the Straight Line for SLR

- The linear function that binds the input variable x with the corresponding predicted value of (yˆ) can be given by the equation of straight line(slope-intercept form) as:

$$\hat{y} = \beta_0 + \beta_1 x$$

- where β_1 is the slope of line (i.e. it measures change in output variable y with unit change in independent variable x).
- β_0 represents y-intercept i.e. the point at which the line touch x-axis
- $\hat{y}$ is the predicted value of the output for the particular value of input variable x.

Cost/Error function for SLR

- The major goal of SLR model is to fit the straight line that predicts the output variable value quite close to the actual value.
- But, in real world scenario, there is always some error (regression residual) in predicting the values, i.e.

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

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

$$Residual\ Error = \epsilon_i = y_i - y\hat{}_i$$

This error may be positive or negative, as it may predict values greater or lesser than actual values. So we consider **square of each error value.**

Cost/Error function for SLR

- The total error for all the n points in the dataset is given by:

$$Total\ Square\ Error = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - y\hat{}_i)^2$$

$$= \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

- The mean of square error is called the cost or error function for simple linear function denoted by J(β_0, β_1) and given by:

$$J(\beta_0, \beta_1) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

- There exist many methods to optimize (minimize) this cost/error function to **find line of best fit.**

Least Square Method for Line of Best Fit

- The least square method aims to find values $\beta_0\hat{}$and $\beta_1\hat{}$ for β_0 and $\beta_1$for which the square error between the actual and the predicted values is minimum i.e. least (So, the name is least square error fit).
- The values $\beta_0\hat{}$and $\beta_1\hat{}$ for β_0 and $\beta_1$for which the square error function (J (β_0, β_1)) is minimum are computed using second derivative test as below:

  1. Compute partial derivatives of J (β_0, β_1) w.r.t β_0 and β_1 i.e. $\frac{\partial \, J \, (\beta_0, \beta_1)}{\partial \beta_0}$ and $\frac{\partial \, J \, (\beta_0, \beta_1)}{\partial \beta_1}$
  2. Find values $\beta_0\hat{}$and $\beta_1\hat{}$ for which $\frac{\partial \, J \, (\beta_0, \beta_1)}{\partial \beta_0} = 0$ and $\frac{\partial \, J \, (\beta_0, \beta_1)}{\partial \beta_1} = 0$
  3. Find second partial derivative $\frac{\partial^2 J \, (\beta_0, \beta_1)}{\partial {\beta_0}^2}$ and $\frac{\partial^2 \, J \, (\beta_0, \beta_1)}{\partial {\beta_1}^2}$; and prove it be minimum for $\beta_0\hat{}$and $\beta_1\hat{}$.

Least Square Error Fit- Contd.....

$$Total\ Sqaure\ Error = J(\beta_0, \beta_1) = \sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$

Step 1: Compute partial derivatives of J (β_0, β_1) w.r.t β_0 and β_1 i.e. $\frac{\partial\, J\,(\beta_0,\beta_1)}{\partial \beta_0}$ and $\frac{\partial\, J\,(\beta_0,\beta_1)}{\partial \beta_1}$

$$\frac{\partial\, J\,(\beta_0,\beta_1)}{\partial \beta_0} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)$$

$$\frac{\partial\, J\,(\beta_0,\beta_1)}{\partial \beta_1} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)\,(x_i) = -2\sum_{i=1}^{n}(x_i y_i - \beta_0 x_i - \beta_1 x_i^2)$$

Step 2: Find values $\beta_0\hat{}$ and $\beta_1\hat{}$ for which $\frac{\partial\, J\,(\beta_0,\beta_1)}{\partial \beta_0} = 0$ and $\frac{\partial\, J\,(\beta_0,\beta_1)}{\partial \beta_1} = 0$

$$\sum_{i=1}^{n}(y_i - \beta_0\hat{} - \beta_1\hat{}\,x_i) = 0 \qquad \textbf{(1)}$$

and $$\sum_{i=1}^{n}(x_i y_i - \beta_0\hat{}\,x_i - \beta_1\hat{}\,x_i^2) = 0 \qquad \textbf{(2)}$$

Least Square Error Fit- Contd.....

From equation 1:

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \beta_0\hat{} - \sum_{i=1}^{n} \beta_1\hat{}x_i = 0$$

$$\sum_{i=1}^{n} y_i - n\beta_0\hat{} - \beta_1\hat{}\sum_{i=1}^{n} x_i = 0$$

$$n\beta_0\hat{} + \beta_1\hat{}\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \qquad (3)$$

From equation 2:

$$\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} \beta_0\hat{}x_i - \sum_{i=1}^{n} \beta_1\hat{}x_i^2 = 0$$

$$\sum_{i=1}^{n} x_i y_i - \beta_0\hat{}\sum_{i=1}^{n} x_i - \beta_1\hat{}\sum_{i=1}^{n} x_i^2 = 0$$

$$\beta_0\hat{}\sum_{i=1}^{n} x_i + \beta_1\hat{}\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \qquad (4)$$

Least Square Error Fit- Contd…..

Multiply equation 3 with $\sum_{i=1}^{n} x_i$ and equation 4 by n

$$n\beta_0\hat{}\sum_{i=1}^{n} x_i + \beta_1\hat{}(\sum_{i=1}^{n} x_i)^2 = \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i \quad (5)$$

$$\mathrm{n}\beta_0\hat{}\sum_{i=1}^{n} x_i + \beta_1\hat{}\mathrm{n}\sum_{i=1}^{n} x_i^2 = n\sum_{i=1}^{n} x_i y_i \quad (6)$$

Subtracting Equation 5 from 6, we get,

$$\boldsymbol{\beta_1\hat{}} = \frac{\boldsymbol{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}}{\mathbf{n}\boldsymbol{\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}}$$

From Equation (3),

$$n\beta_0\hat{} + \beta_1\hat{}\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$$

$$\beta_0\hat{} = \frac{1}{n}\sum_{i=1}^{n} y_i - \frac{1}{n}\beta_1\hat{}\sum_{i=1}^{n} x_i$$

$$\boldsymbol{\beta_0\hat{} = \overline{y} - \beta_1\hat{}\overline{x}}$$

Least Square Error Fit- Contd.....

Step 3: Find second partial derivative $\frac{\partial^2 J(\beta_0,\beta_1)}{\partial \beta_0{}^2}$ and $\frac{\partial^2 J(\beta_0,\beta_1)}{\partial \beta_1{}^2}$; and prove it be minimum for β_0ˆand β_1ˆ .

$$\frac{\partial^2 J(\beta_0,\beta_1)}{\partial \beta_0{}^2} = \frac{\partial(-2\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)}{\partial \beta_0} = -2 \times (-1) = 2$$

$$\frac{\partial^2 J(\beta_0,\beta_1)}{\partial \beta_1{}^2} = \frac{\partial-2\sum_{i=1}^{n}(x_i y_i-\beta_0 x_i-\beta_1 x_i^2)}{\partial \beta_1} = -2 \times \left(-x_i^2\right) = 2x_i^2$$

Therefore, both are positive i.e. the cost function continuously increases with respect to β_0 and $\beta_1$beyond β_0ˆ and β_1ˆ. But attains it minimum value at β_0ˆ and β_1ˆ

Least Square Error Fit- Summary

- The linear function that binds the input variable x with the corresponding predicted value of (yˆ) can be given by the equation of straight line(slope-intercept form) as:

$$y\hat{}=\beta_0 + \beta_1 x$$

- The square error in prediction is minimized when

$$\beta_1\hat{} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\mathrm{n}\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

$$= r_{xy}\frac{\sigma_y}{\sigma_x} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

and $\beta_0\hat{} = \bar{y} - \beta_1\hat{}\bar{x}$

Least Square Error Fit- Example

The data set (shown in table) gives average masses for women as a function of their height in a sample of American women of age 30–39.

(a) Fit a square line for average mass as function of height using least square error method.

(b) Predict the average mass of women whose height is 1.40 m

Height (m), x_i	Mass (kg), y_i
1.47	52.21
1.50	53.12
1.52	54.48
1.55	55.84
1.57	57.20
1.60	58.57
1.63	59.93
1.65	61.29
1.68	63.11
1.70	64.47
1.73	66.28
1.75	68.10
1.78	69.92
1.80	72.19
1.83	74.46

Least Square Error Fit- Example

i	x_i	y_i	x_i2	x_iy_i
1	1.47	52.21	2.1609	76.7487
2	1.50	53.12	2.25	79.68
3	1.52	54.48	2.3104	82.8096
4	1.55	55.84	2.4025	86.552
5	1.57	57.20	2.4649	89.804
6	1.60	58.57	2.56	93.712
7	1.63	59.93	2.6569	97.6859
8	1.65	61.29	2.7225	101.1285
9	1.68	63.11	2.8224	106.0248
10	1.70	64.47	2.89	109.599
11	1.73	66.28	2.9929	114.6644
12	1.75	68.10	3.0625	119.175
13	1.78	69.92	3.1684	124.4576
14	1.80	72.19	3.24	129.942
15	1.83	74.46	3.3489	136.2618
Total	24.76	931.17	41.0532	1548.2453

Least Square Error Fit- Example

$$\beta_1\hat{} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\mathrm{n}\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

$$\beta_1\hat{} = \frac{15 \times 1548.2453 - 24.76 \times 931.17}{15 \times 41.0532 - 24.76^2} = 61.19$$

$$\beta_0\hat{} = \bar{y} - \beta_1\hat{}\bar{x}$$

$$\beta_0\hat{} = \frac{931.17}{15} - 61.19 \times \frac{24.76}{15} = -38.88$$

Therefore the line of best fit is given by: $y\hat{} = -38.88 + 61.19x$

Predicted value of y when x is 1.4 is

$y\hat{} = -38.88 + 61.19 \times 1.4 = 46.78$