Page No.
Date

Simple linear Model :-

In simple linear Regression
ind variable x (input)
depend variable y (output)

assumed that relationship betn x and y is linear, i.e. can be represented by line

$$y = ax + b$$

$y \leftarrow$ predicted value
$a \leftarrow$ slope of line
$b \leftarrow$ intercept of
$x \leftarrow$ input value.

Goal : we want to determine the value of slope a and b intercept b such that the line best fits the given data points.

① least sq. estimate :-

To find the values a and b, we use least square method. The basic idea of this method is to find the line that minimizes sum of sqred differences betn the actual y-values & the predicted y-values from the line.

for each data point (x_i, y_i) the predicted value from the line is,

$$\hat{y}_i = ax_i + b$$

The residual error of each data points are :-

$$e_i = y_i - \hat{y}_i = y_i - (ax_i + b)$$

The sum of squared residuals is

$$S(a,b) = \sum_{i=1}^{n} (y_i - (ax_i + b))^2$$

② minimizing sum of squared residuals :→

for a) :-

$$\frac{\partial}{\partial a} s(a,b) = \frac{\partial}{\partial a} \sum_{i=1}^{n} (y_i - (ax_i + b))^2$$

using the chain rule, we get,

$$\frac{\partial}{\partial a} s(a,b) = -2 \sum_{i=1}^{n} x_i (y_i - (ax_i + b))$$

we set this derivative equal to zero to minimize the sum of sq. residuals.

$$-2 \sum_{i=1}^{n} x_i (y_i - (ax_i + b)) = 0$$

$$\sum_{i=1}^{n} x_i (y_i - ax_i - b) = 0 \text{ ——— } ①$$

$$\sum_{i=1}^{n} x_i y_i - a \sum_{i=1}^{n} x_i^2 - b \sum_{i=1}^{n} x_i = 0$$

for b) :-

$$\frac{\partial}{\partial a} s(a,b) = \frac{\partial}{\partial b} \sum_{i=1}^{n} (y_i - (ax_i + b))^2$$

using the chain rule, we get.

$$\frac{\partial}{\partial a} (s(a,b)) = -2 \sum_{i=1}^{n} (y_i - (ax_i + b))$$

set this equal to zero :-

$$\sum_{i=1}^{n} (y_i - ax_i - b) = 0$$

$$\sum_{i=1}^{n} y_i - a \sum_{i=1}^{n} x_i - b \sum_{i=1}^{n} 1 = 0$$ ∴ since sum of 1 is just n (the no of data pᵢ)

Rearranging ⇒ $$b = \frac{1}{n} \left(\sum y_i - a \sum x_i \right)$$ ⇒ $\sum y_i - a \sum x_i - b\, n = 0$

now we have to equations :-

solving for slope a

$$a = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

solve for the 2nd equation :-

$$\sum_{i=1}^{n} y_i - a \sum_{i=1}^{n} x_i - nb = 0$$

Rearranging the eqn isolate b :

$$nb = \sum_{i=1}^{n} y_i - a \sum_{i=1}^{n} x_i$$

$$b = \frac{1}{n} \sum_{i=1}^{n} y_i - \frac{a}{n} \sum_{i=1}^{n} x_i$$

. . . . divide by n .

Thus the intercept b is :

$$b = \frac{\sum_{i=1}^{n} y_i}{n} - \frac{a \sum^{n} x_i}{n}$$

Substitute b into first eqn.

$$\sum x_i (y_i - a x_i - b) = 0$$

$$\sum x_i \left(y_i - a x_i - \left(\frac{\sum y_i}{n} - \frac{a \sum x_i}{n} \right) \right) = 0$$

Distribute the summation inside parenthesis.

$$\sum x_i y_i - a \sum x_i^2 - \sum x_i \left(\frac{\sum y_i}{n} \right) + a \sum x_i \left(\frac{\sum x_i}{n} \right) = 0$$

Mean v/s Mean squared distance

perpendicular distance formula =

$D(m, c)$

general eqn of straight line $\qquad y = ax + b + c$

$$ax - y + b = 0$$

The low distance d from point (x_0, y_0) to the line by the eqn $Ax + By + C = 0$ is

$$d = \frac{|Ax_0 + By_0 + c|}{\sqrt{A^2 + B^2}}$$

For :-

$$y = ax + b$$

$$ax - y + b = 0$$

Here,

$$A = a \qquad B = -1 \qquad c = b$$

$$d = \frac{|ax_0 - y_0 + b|}{\sqrt{a^2 + (-1)^2}} \Rightarrow d = \frac{|ax_0 - y_0 + b|}{\sqrt{a^2 + 1}}$$

$D(a, b) =$

$$d_i^2 = \frac{(ax_i - y_i + b)}{a^2 + 1}$$

$$S = \sum_{i=1}^{n} \frac{(ax_i - y_i + b)^2}{a^2 + 1}$$

Date

$$\frac{\partial}{\partial a} S = \frac{\partial}{\partial a}\left(\Sigma \frac{(ax_i - y_i + b)^2}{a^2+1}\right)$$

$=$

$f(a) = (ax_i - y_i + b)^2$

$f(b) = a^2+1$

Now by quotient rule =

$$\frac{d}{da}\left(\frac{f(a)}{g(a)}\right) = \frac{f'(a)\, g(a) - f(a)\, g'(a)}{g(a)^2}$$

$$= \frac{2(ax - y + b)\,x\,(a^2+1) - (ax_i - y_i + b)^2 (2a)}{(a^2+1)^2}$$

$$\frac{\partial s}{\partial a} = \Sigma \frac{2(ax_i - y + b)\,x\,(a^2+1) - 2a\,(ax_i - y_i + b)^2}{(a^2+1)^2}$$

= equating to zero.

$$\Sigma (ax_i - y + b)\left(x(a^2+1) - a(ax - y + b)\right) = 0$$

diff. $\frac{\partial s}{\partial b} = 0$

$$= \Sigma \frac{2(ax_i - y_i + b)}{a^2+1}$$

$$0 = \Sigma(ax_i - y_i + b)$$

$$nb + a\,\Sigma x_i = \Sigma y_i$$

$$b = \frac{\Sigma y_i - a\,\Sigma x_i}{n}$$