Linear Regression

$\underset{\text{Independent}}{\textcircled{X}} \longrightarrow \underset{\text{dependent}}{\textcircled{Y}}$ $\Longrightarrow$ try to predict Y interms of X

step 1

Correlation
- → Positive
- → Negative
- → Uncorrelated

step 2

→ Perform Linear regression

→ Find the best fit line with least error

error = difference between actual value & predicted value

total error $= \Sigma(y - \hat{y})^2$

$\hat{y}$ (predicted) $= ax + b$

average model $= \Sigma(y_i - \bar{y})^2$

Linear reg model $= \Sigma(y_i - \hat{y}_i)^2$

$$f(a,b) = \Sigma(y_i - ax_i - b)^2$$

$$\frac{\delta f}{\delta a} = 0 \qquad \frac{\delta f}{\delta b} = 0$$

more math

$$a = \frac{\Sigma(x_i - \bar{x})(y_i - y)}{\sqrt{\Sigma(x_i - \bar{x})^2 \, \Sigma(y_i - \bar{y})^2}}$$

Step 3

Score = Worthiness ie

Coefficient of determination

$0 \le r^2 \le 1$

$\leftarrow$ bad model $\quad\quad \rightarrow$ good model

$$\frac{\text{error from LR mode}}{\text{error from avg model}}$$ tending to 0 its a good model

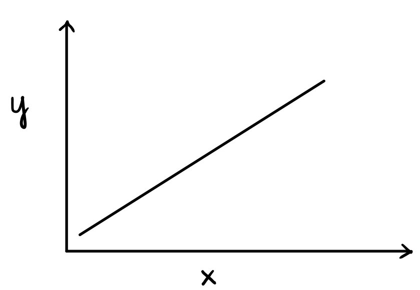
i.e $$\frac{\Sigma (y_i - \hat{y}_i)^2}{\Sigma (y_i - \bar{y}_i)^2}$$

i.e error from LR is much much less Than avg meaning LR is a good model.

$$0 \le \underbrace{1 - \frac{\Sigma (y_i - \hat{y}_i)^2}{\Sigma (y_i - \bar{y}_i)^2}}_{r^2} \le 1$$

Coefficient of determination

Correlation



$\longrightarrow$ $y = ax + b$

$$R^2 = 1 - \frac{\Sigma (y_i - \hat{y})^2}{\Sigma (y_i - \bar{y})^2}$$

$$\underset{\text{-ve correlation}}{-1} \le \underbrace{\frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma (x_i - \bar{x})^2 (y_i - \bar{y})^2}}}_{\substack{\Downarrow \\ 0 \\ \text{un-correlated}}} \le \underset{\text{+ve correlation}}{1}$$

Linear regression example : Waist circumference Adipose Tissue

$\uparrow$ AT in abdominal region $\Rightarrow$ $\uparrow\uparrow$ cardiovascular disease.

OLS = Ordinary list square

adjusting the R^2 values

MultiLinear Regression

$$Y = a_1x_1 + a_2x_2 + a_3x_3 + \cdots + b + \varepsilon$$

$\Rightarrow$ the features are independent of each other