PREDICTION

To predict a continuous value rather than a categorical label.

The prediction of continuous values can be modeled by statistical techniques of Regression.

(I) Linear Regression.

Data are modeled using a straight line.

(II) ~~Bivariant~~ Bivariate Linear Regression.

It models a random ~~valued~~ variable (Y) (called a response var) as a linear function of another random variable (X) (called a predictor var)

i.e $\boxed{Y = \alpha + \beta X}$

variance of Y is assumed to be constant

α is the regression coefficient specifing the Y intercept

β is the regression coefficient specifing slope of the line.

Coefficients can be solved by the method of Least Square, which minimised the error b/w the actual data and estimate of the line.

Samples or data points are $(x_1, y_1), (x_2, y_2), \ldots (x_s, y_s)$

$$\boxed{\beta = \frac{\sum_{i=1}^{s} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{s} (x_i - \bar{x})^2}}$$

$$\boxed{\alpha = \bar{y} - \beta \bar{x}}$$

→ There are marks on steps as well

Develop a model to predict the salary of college graduates with 10 years of work experience, predict the salary of a college graduate.

	(years-experience) X	Salary Y (salary) $1000s	$(X-\bar{X})$	$(X-\bar{X})$	$(Y-\bar{Y})$
1	3	30			
2	8	57			
3	9	64			
4	13	72			
5	3	36			
6	6	43			
7	11	59			
8	21	90			
9	1	20			
10	16	83			

$\Sigma X = 91$ $\Sigma Y = 554$ $n = 10$

predict where $X = 10$

$\bar{X} = 9.1$ $\bar{Y} = 55.4$

$$\beta = \frac{(3-9.1)(30-55.4)}{(3-9.1)^2} + \dots$$

$$\beta = 3.5374$$

$$\alpha = \bar{Y} - \beta\bar{X}$$

$$\alpha = 55.4 - 3.5374(9.1)$$

$$= 23.2096$$

$Y = \alpha + \beta X$ For $X = 10$

$$Y = 23.2096 + 3.5374(10)$$

$$\boxed{Y = 58.58}$$

Ans) So the salary of a college graduate with 10 years of experience is \$ 58.58 K