LECTURE NOTES

Campus: PCE Course: BTECH Class/Section: III yr IV/C Date: 22.02.22
Name of Faculty: Rawesh kv. YADAV Name of Subject: Machine Learning Code: 6CS4-02
Date (Prep.): 22.02.22 Date (Del.): 2.03.22 Unit No./Topic: 1 Lect. No: 3

OBJECTIVE: To be written before taking the lecture (Pl. write in bullet points the main topics/concepts etc., which will be taught in this lecture)

Linear Regression

IMPORTANT & RELEVANT QUESTIONS:

① what is Linear Regression? Explain with the help of any eg.

FEEDBACK QUESTIONS (AFTER 20 MINUTES):

① what are the differences between dependent and Independent variable.

OUTCOME OF THE DELIVERED LECTURE: To be written after taking the lecture (Pl. write in bullet points about students' feedback on this lecture, level of understanding of this lecture by students etc.)

REFERENCES: Text/Ref. Book with Page No. and relevant Internet Websites:

Hands on Machine Learning, Aurlien Geron.

Linear Regression Model:-

- R.p.t estimation of a company.
- Independent variable - A variable whose value doesn't change by the effect of other variable and used to manipulate the dependent variable. It is often denoted as x.

eg- Rainfall - Independent variable (bcoz rainfall can't be controlled by human being but rainfall can effect the crop.

- Dependent variable - Are those variable whose value change when there is a change in the independent variable. It is often denoted as y.

eg- crop yield depends on the amount of rainfall received.

Applications of Linear Regression-

1. Economic Growth - Used to determine the economic growth of a country or state in coming quarter.
   - can be used to predict the GDP of a country.
   - " " " " the price of a product in the future.
   - To estimate the number of houses a builder would sell and at what prices in the coming month.
2. Score prediction - To predict the number of runs a player would score in the coming match based on previous performance.

Understanding Linear Regression:- Linear Regression is a statistical model used to predict the relationship between independent and dependent variable.

The simplest form of a simple linear regression equation with one dependent and one independent variable is represented by-

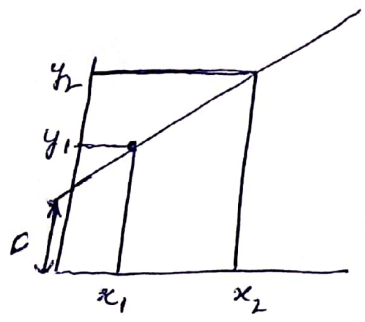
$$y = mx + c$$

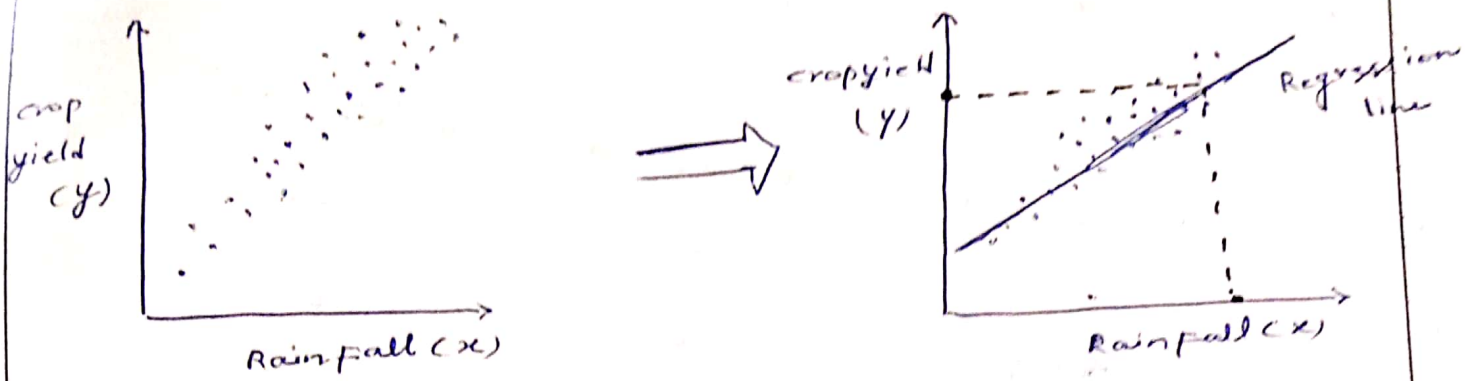
y - dependent variable

x - independent variable

m - slope of line $m = \frac{y_2 - y_1}{x_2 - x_1}$

c - intercept or coefficient of the line.





$$m = \frac{(n * \Sigma(x*y)) - (\Sigma(x) * \Sigma(y))}{(n * \Sigma(x^2)) - (\Sigma(x))^2}$$

x	y (actual y values)	$\hat{y}$ (predicted)
1	2	2.8
2	4	3.4
3	5	4
4	4	4.6
5	5	5.2

$$= \frac{(5*66) - (15*20)}{(5*55) - 225} = 0.6$$

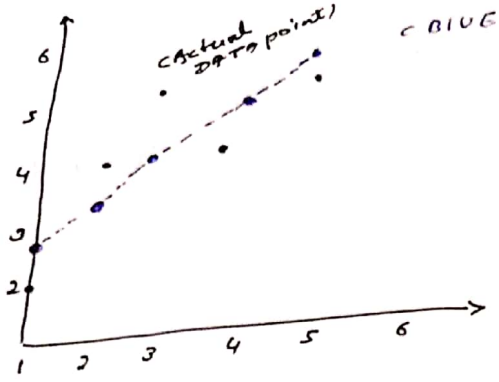
$$c = \frac{(\Sigma(y) * \Sigma(x^2)) - (\Sigma(x) * \Sigma(x*y))}{(n * \Sigma(x^2)) - (\Sigma(x))^2)}$$

$$= 2.2$$

$$y = m * x + c$$

$$= 0.6 * 3 + 2.2$$

$$= 4$$



BLUE points are - predicted data points.

x	y	ypred	y - ypred	(y - ypred)^2
1	2	2.8	- 0.8	0.64
2	4	4.4	0.6	0.36
3	5	4	1	1
4	4	4.6	- 0.6	0.36
5	5	5.2	- 0.2	0.04
				$\Sigma = 2.4$

Residual of errors - The distance between actual and predicted values are known as residual or errors.

The best fit line should have the least sum of squares of these errors also known as e-square.

The sum of squared errors of this regression line is 2.4.

we check this error for each line and conclude the best fit line having the least e-square value.

MULTIPLE LINEAR REGRESSION:-

PAGE NO.

Multiple Linear Regression $\longrightarrow$ $y = m_1x_1 + m_2x_2 + \ldots + m_nx_n + c$

- y: Dependent variable
- $m_1, m_2, \ldots, m_n$: slopes
- $x_1, x_2, \ldots, x_n$: Independent variables (IdV's)
- c: coefficient