SILVER OAK UNIVERSITY

EDUCATION TO INNOVATION

SILVER OAK COLLEGE OF COMPUTER APPLICATION

SUBJECT :MACHINE LEARNING

TOPIC :Unit:-2 Supervised Learning(Linear Regression)

Linear Regression Equation

Linear regression is a way to model the relationship between two variables. You might also recognize the equation as the **slope formula**. The equation has the form Y= a + bX, where Y is the dependent variable (that's the variable that goes on the Y axis), X is the independent variable (i.e. it is plotted on the X axis), b is the slope of the line and a is the y-intercept.

$$a = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$b = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

Step 1: *Make a chart of your data, filling in the columns in the same way as you would fill in the chart if you were finding the Pearson's Correlation Coefficient.*

1439942571849980132579197562562415578749593249756 9

Subject	Age x	Glucose Level y	xy	x^2	y^2
2	21	65	1365	441	4225
4	42	75	3150	1764	5625
6	59	81	4779	3481	6561
Σ	247	486	20485	11409	40022

Step 2: Use the following equations to find a and b.

$$a = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$b = \frac{n(\Sigma xy) - (\Sigma x)(\Sigma y)}{n(\Sigma x^2) - (\Sigma x)^2}$$

a = **65.1416**

b = **.385225**

Find a:

$((486 \times 11{,}409) – ((247 \times 20{,}485)) / 6\ (11{,}409) – 247^2)$

484979 / 7445

=**65.14**

Find b:

$(6(20{,}485) – (247 \times 486)) / (6\ (11409) – 247^2)$

$(122{,}910 – 120{,}042) / 68{,}454 – 247^2$

2,868 / 7,445

= **.385225**

Step 3: *Insert the values into the equation.*

y' = a + bx

y' = 65.14 + .385225x

Solved Examples

1. Find a linear regression equation for the following two sets of data:

x	2	4	6	8
y	3	7	5	10

x	y	x^2	xy
2	3	4	6
4	7	16	28
6	5	36	30
8	10	64	80
$\Sigma x = 20$	$\Sigma y = 25$	$\Sigma x^2 = 120$	$\Sigma xy = 144$

$$a= \frac{\left(\sum_{Y}\right)\left(\sum_{X^2}\right)-\left(\sum_{X}\right)\left(\sum_{XY}\right)}{n\left(\sum_{x^2}\right)-\left(\sum_{x}\right)^2}$$

Now put the values in the equation

$$a = \frac{25 \times 120 - 20 \times 144}{4 \times 120 - 400}$$

$$a= \frac{120}{80}$$

a=1.5

b= $\frac{n\left(\sum_{XY}\right)-\left(\sum_{X}\right)\left(\sum_{Y}\right)}{n\left(\sum_{x^2}\right)-\left(\sum_{x}\right)^2}$

Put the values in the equation

$$b = \frac{4 \times 144 - 20 \times 25}{4 \times 120 - 400}$$

b= $\frac{76}{80}$

b=0.95

Hence we got the value of a = 1.5 and b = 0.95

The linear equation is given by

$Y = a + bx$

Now put the value of a and b in the equation

Hence equation of linear regression is $y = 1.5 + 0.95x$

Polynomial Regression

- $y=a_0+a_1*x+a_2*x^2 +\ldots\ldots\ldots+ +a_n*x^n$

Problem Deninition:

Find a quadratic regression model for the following data:

X	Y
3	2.5
4	3.2
5	3.8
6	6.5
7	11.5

Solution:

Let the quadratic polynomial regression model be

$$y=a_0+a_1*x+a_2*x^2$$

The values of **a_0, a_1, and a_2** are calculated using the following system of equations:

$$\sum y_i = na_0 + a_1(\sum x_i) + a_2(\sum x_i^2)$$
$$\sum y_i x_i = a_0(\sum x_i) + a_1(\sum x_i^2) + a_2(\sum x_i^3)$$
$$\sum y_i x_i^2 = a_0(\sum x_i^2) + a_1(\sum x_i^3) + a_2(\sum x_i^4)$$

$$\sum y_i x_i^2 = a_0(\sum x_i^2) + a_1(\sum x_i^3) + a_2(\sum x_i^4)$$

First, we calculate the required variables and note them in the following table.

	x	**y**	**x^2**	**x^3**	**x^4**	**y*x**	**$y*x^2$**
	3	2.5	9	27	81	7.5	22.5
	4	3.2	16	64	256	12.8	51.2
	5	3.8	25	125	625	19	95
	6	6.5	36	216	1296	39	234
	7	12	49	343	2401	80.5	563.5
Σ	**25**	**27.5**	**135**	**775**	**4659**	**158.8**	**966.2**

Using the given data we,

$$27.5 = 5a_0 + 25a_1 + 135a_2$$

$$158.8 = 25a_0 + 135a_1 + 775a_2$$

$$966.2 = 135a_0 + 775a_1 + 4659a_2$$

Solving this system of equations we get

a_0=12.4285714

a_1=-5.5128571

a_2=0.7642857