ML Assignment - 2

IIT2019044 Shreesh Swaraj

1) ans ~~Y 2006 2008~~ 2009 201

Y	2006	2008	2009	2011	2013	2014	2015	2016	2017	2018	2019
Rev	100.2	98.3	87.1	89.2	88.9	83.5	89.1	84	92.3	96	97

Let us assume least square line fitting data be

$y = w_0 + w_1 x.$

y = expected revenue

x = years

$$J = \sum_{i=1}^{n} \left(h_w(x_i) - y_i\text{actual}\right)^2$$

$$[h_w(x) = w_0 + w_1 x]$$

By minimising J we can find value of w_0 & w_1

$$w_0 = \frac{BC - AD}{CA - A^2} \qquad w_1 = \frac{AB - DA}{A^2 - CA}$$

$$B = \sum_{1}^{n} y_i,\ A = \sum_{1}^{n} x_i,\ C = \sum_{1}^{n} x_i^2,\ D = \sum_{1}^{n} x_i y_i$$

$w_0 = 627.6287$

$w_1 = -0.2663761$

$$\boxed{y = 627.6287 + (-0.2663761)\, x}$$

b) Expected revenue in 2021

$y = 627.6287 + (-0.26637) \times 2021$

$y = 89.3601$ Billion Rupees

c) Expected error : $h_w(x) - y_{actual}$

$h_w(x) = 627.628 + -(0.26633)x$

Mean squared error: $\sum_{1}^{n} (h_w(x_i) - y_{actual})^2$

$= 28.85011$

(2)

ML	85	90	93	75	87	71	98	68	84	87
HUR	82	88	96	72	91	80	95	72	89	84

a) Regression of Y on X

$Y = w_0 + w_1 x \rightarrow$ Best fit line

$y = HUR$; $X = ML$.

$w_0 = \frac{BC - AD}{CA - A^2}$ $\quad w_1 = \frac{AB - AD}{A^2 - CA}$

$A = \sum_{1}^{n} x_i$, $B = \sum_{1}^{n} y_i$, $C = \sum^{n} x_i^2$, $D = \sum^{n} x_i y_i$

So putting value

$w_0 = 25.79$, $w_1 = 0.71$

$\boxed{Y(HUR) = 25.79 + 0.71x}$

b) Regression of x on y

$x = w_0 + w_1 y$

$y = HUR \quad ; \quad x = ML$

$w_0 = \dfrac{BC - AD}{CA - A^2} \qquad w_1 = \dfrac{AB - AD}{A^2 - CA}$

$A = \sum_1^n x_i \, , \quad B = \sum_1^n y_i \, , \quad C = \sum_1^n x_i^2 \, , \quad D = \sum_1^n x_i y_i$

$w_0 = -19.328 \, , \quad w_1 = 1.202$

$\boxed{x(ML) = -19.32 + (1.202) * y(HUR)}$

c) Marks in ML = 96

Expected marks in HUR $= w_0 + w_1 x$

$= 25.79 + (0.71 \times 96)$

$= 94.32$

d) Marks in HUR = 95

Expected marks in ML $= -19.328 + 1.202 \times 95$

$= 94.9496$

e) Mean squared error in case of regression of X on Y = 28.85

Mean squared error in case of regression of Y on X = 9.25

Hence regression of Y on X gives more accurate results

Here both lines regression of Y on X & X on Y are independent, cant be derived from other

(3) a) $PV^n = c$

Taking log

$\ln(PV^n) = \ln(c)$

~~n ln(P~~

$\ln P + n\ln(V) = \ln(c)$.

$\ln(P) = \ln(c) - n\ln(V)$

$\ln(P) = \ln(c) + n\ln(1/V)$.

This equation is form of $Y = w_0 + w_1 x$

where $w_0 = \ln c$

$Y = \ln P$

$w_1 = n$

$x = \ln(1/V)$

$$W_0 = \frac{BC - AD}{AC - A^2} \qquad W_1 = \frac{AB - AD}{A^2 - AC}$$

$$B = \sum_1^n \ln P_i \;,\; A = \sum_1^n \ln(1/v_i)$$

$$C = \sum_1^n \left(\ln \frac{1}{v_1}\right)^2 \;,\; D = \sum_1^n \ln\left(\frac{1}{v_i}\right) \ln(P_i)$$

$W_0 = 9.54$, $c = e^{9.54} = 13904.94$

$W_1 = 1.4$, $n = 1.4$

b) $n = 1.4$ $\quad c = 13904.94$

$PV^{1.4} = 13904.94$

c) $V = 100$

$$P = \frac{13904.94}{(100)^{1.4}} = \frac{13904.94}{630.95} = 22.069$$

(4) $y = w_0 + w_1 x + w_2 x^2$

This linear hypothesis has variable, this there will be 3 equations.

$$\sum^m x_i^2 y_i = w_0 \sum x_i^2 + w_1 \sum x_i^3 + w_2 \sum^m x_i^4$$

$$\sum^m x_i y_i = w_0 m + w_1 \sum^m x_i + w_1 \sum^m x_i^2$$

$$\sum^m y_i = w_0 m + w_1 \sum^m x_i + w_2 \sum x_i^2$$

x	y	x^2	x^3	x^4	xy	x^2y
0	2.4	0	0	0	0	0
1	2.1	1	1	1	2.1	2.1
2	3.2	4	8	16	6.4	12.8
3	5.6	9	27	81	16.8	50.4
4	9.3	16	64	256	37.2	148.8
5	14.6	25	125	625	73	365
6	21.9	36	216	1296	131.4	788.4
21	59.1	91	441	2275	266.9	1367.5

∴ Forming equations:

$7a + 21b + 91c = 59.1$

$91a + 441b + 2275c = 1367.5$

$21a + 91b + 441c = 266.9$

Upon solving:

$y = 2.5095 - 1.2x + 0.733x^2$