Name - Shubham Kumar
Roll No - IIT2018146

ML assignment - 2

Q1)

Year	2004	2008	2009	2010	2011	2012	2013	2014	2015	2016
Revenue	61.2	58.3	67.1	69.2	68.9	83.5	89.1	80	92.3	93

Year	2017
Reven	97

(a) Draw a least square line fitting the data.

Ans:-

First of all, assuming linear regression line will be fit for our about dataset.

so, let line eq^n be $y = h_w(x) = w_0 + w_1 x$

where w_0, w_1 are some constant.

Now our aim is to find w_0, and w_1

This can be done by ~~not~~ differient

$$error = J = \Sigma \left(h_w(x) - Y_{actual}\right)^2$$

$$\frac{\partial J(w_0, w_1)}{\partial w_0} = m w_0 + w_1 \Sigma x_i - \Sigma y_i = 0$$

$$\frac{\partial J(w_0, w_1)}{\partial w_1} = w_0 \Sigma x_i + w_1 \Sigma x_i^2 - \Sigma x_i y_i = 0$$

$A = \Sigma x_i$ $\quad$ $B = \Sigma y_i$ $\quad$ $C = \Sigma x_i^2$ $\quad$ $D = \Sigma x_i y_i$

$$w_1 = \frac{AB - Dm}{A^2 - Cm}$$

$$w_0 = \frac{BC - AD}{Cm - A^2}$$

Now Year is my Independent variable and income or revenue is my dependent ~~vla~~ variable.

Revenue = $w_0 + w_1$ Year

$$Y = h_w(X) = w_0 + w_1 X$$

Y (Revenue)	X (Year)	X^2	XY
61.2	2004	4016016	122644.8
58.3	2008	4032064	117066.4
67.1	2009	4036081	134803.9
69.2	2010	4040100	139092.0
68.9	2011	4044121	138557.9
83.5	2012	4048144	168002.0
89.1	2013	4052169	179358.3
80	2014	4056196	161120.0
92.3	2015	4060225	185984.5
93	2016	4064256	187488.0
97	2017	4068289	195649.0
Σ 859.6	22,129	44517661	1729766.8

$$w_1 = \frac{AB - Dm}{A^2 - Cm}$$

$m = 11$

$A = \cancel{2212}\ 22129$

$B = 859.6$

$C = 44517661$

$D = 1729766.8$

w1

$$w_1 = \frac{22129 * 859.6 - 1729766.8 * 11}{(22129)^2 - 44517661 * 11}$$

$$w_1 = \frac{19022088.4 - 19027434.8}{489692641 - 489694271}$$

$$w_1 = \frac{-5346.4}{-1630}$$

$$w_1 = 3.28$$

w0

$$w_0 = \frac{BC - AD}{Cm - A^2} = \frac{38267381395.6 - 38278009517.2}{489694271 - 489692641}$$

$$= \frac{-10628121.6}{1630}$$

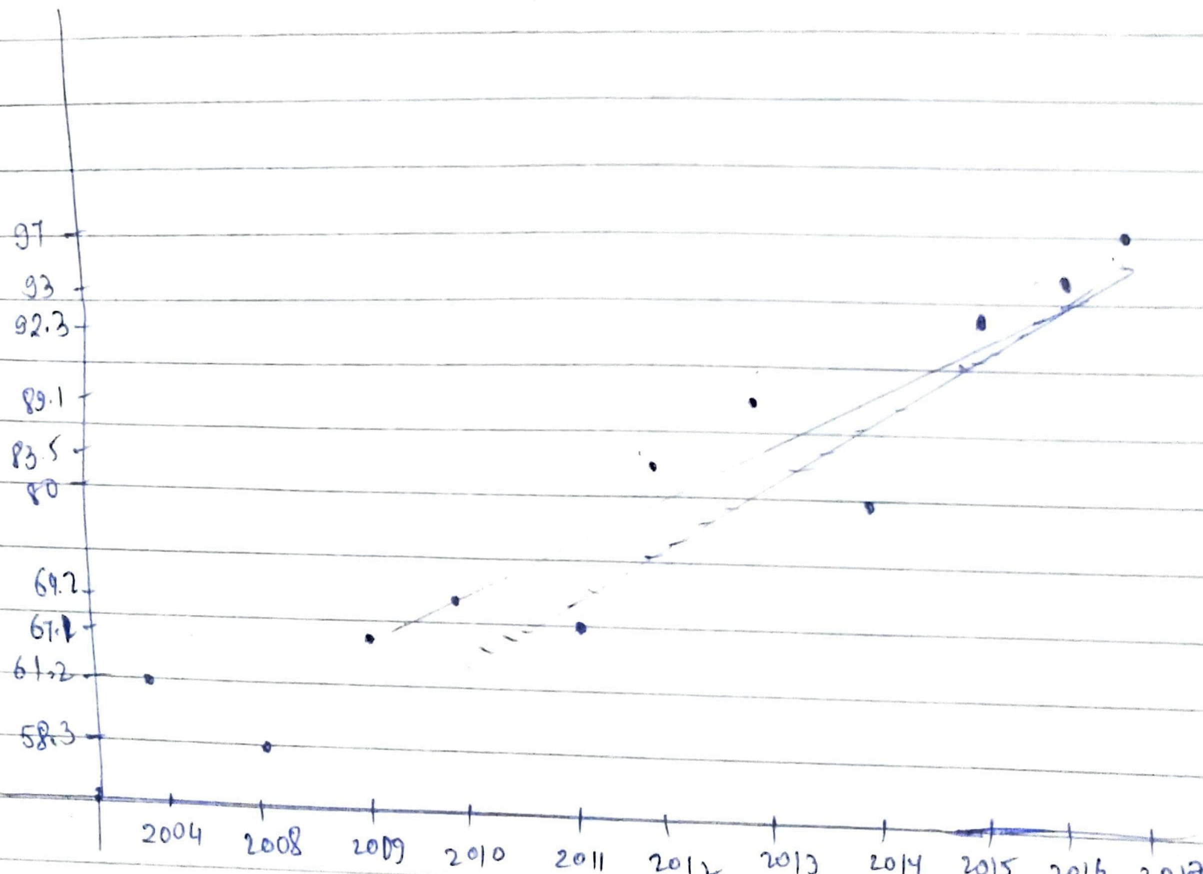
$$= -6520.32$$

so our least square line

Revenue = $-6520.32 + 3.28 * \text{year}$

$$\boxed{y = h_w(x) = -6520.32 + 3.28x}$$

(b)



(b)

Expected Revenue in 2019

Revenue = $-6520.32 + 3.28 * 2019$

$$\boxed{\text{Revenue} = 102}$$

(c) Analyze expected error in prediction.

$$J = \frac{1}{2m} \sum \left(h_w(x) - Y_{actual} \right)^2$$

X (Year)	$h_w(x) = -6520.32 + 3.28 x$	Y_{actual}	$h_w(x) - Y_{actual}$	square
2004	52.80	61.2	-8.40	70.5600
2008	65.92	58.3	7.62	58.0644
2009	69.20	67.1	2.10	4.4100
2010	72.48	69.2	3.28	10.7584
2011	75.76	68.9	6.86	47.0596
2012	79.04	83.5	-4.46	19.8916
2013	82.32	89.1	-6.78	45.9684
2014	85.60	80	5.60	31.3600
2015	88.88	92.3	-3.42	11.6964
2016	92.16	93	-0.84	0.7056
2017	95.44	97	-1.56	2.4336

Summation

$$\sum \left(h_w(x) - Y_{actual} \right)^2 = 302.9080$$

$$\text{Error} = J = \frac{1}{11}(302.9080) = \frac{302.9080}{11}$$

$$\text{error}(J) = 27.5370$$

Q2) ML	75	80	93	65	87	71	98	68	84	77
HUR	82	78	86	72	91	80	95	72	89	74

Find least square line fitting

a) X as independent variable

let first consider ML as independent variable
i.e ML as X and HUR as Y

let the eq^n of best fitting line is
$h_w(X) = Y = w_0 + w_1 x$

	Y (HUR)	X (ML)	X^2	XY
	82	75	5625	6150
	78	80	6400	6240
	86	93	8649	7998
	72	65	4225	4680
	91	87	7569	7917
	80	71	5041	5680
	95	98	9604	9310
	72	68	4624	4896
	89	84	7056	7476
	74	77	5929	5698
Σ	819	798	64722	66045

$A = \Sigma X$ $\quad B = \Sigma Y$ $\quad C = \Sigma X^2$ $\quad D = \Sigma XY$

$A = 798$ $\quad C = 64722$

$B = 819$ $\quad D = 66045$

$$w_1 = \frac{AB - Dm}{A^2 - Cm} = \frac{653562 - 660450}{636804 - 647220}$$

$$= \frac{-6888}{-10416} = 0.6612$$

$$w_0 = \frac{BC - AD}{Cm - A^2} = \frac{53007318 - 52703910}{10416}$$

$$= \frac{303408}{10416}$$

$$= 29.1290$$

$$\boxed{Y = 29.1290 + 0.6612x}$$

$$\boxed{HUR = 29.1290 + 0.6612 * (ML)}$$

(b) Y as independent variable.

So, Now in this case HUR is considered as independent variable and ML as dependent variable

$$Y = w_0 + w_1 x$$
$$ML = w_0 + w_1 (HUR)$$

	Y (ML)	X (HUR)	x^2	xy
Σ	798	819	67675	66045

$A = \Sigma x$, $B = \Sigma y$, $C = \Sigma x^2$, $D = \Sigma xy$

$$w_1 = \frac{AB - Dm}{A^2 - Cm}$$

$A = 819$ $C = 67675$

$B = 798$ $D = 66045$

$$= \frac{653562 - 660450}{670761 \cancel{636804} - 676750} = \frac{-6888}{-\cancel{39946}\ 5989} = 1.150$$

$$w_0 = \frac{BC - AD}{Cm - A^2} = \frac{54004650 - 54090855}{39946}$$

$$= \frac{-86205}{\cancel{39946}\ 5989} = -14.393$$

$\boxed{Y = -14.393 + 1.150X}$

$\boxed{ML = -14.393 + 1.150(HUR)}$

(c) If student receive a mark of 96 in ML Then expected marks in HUR

$HUR = 29.129 + 0.6612 * (96)$

$= 92.6042$

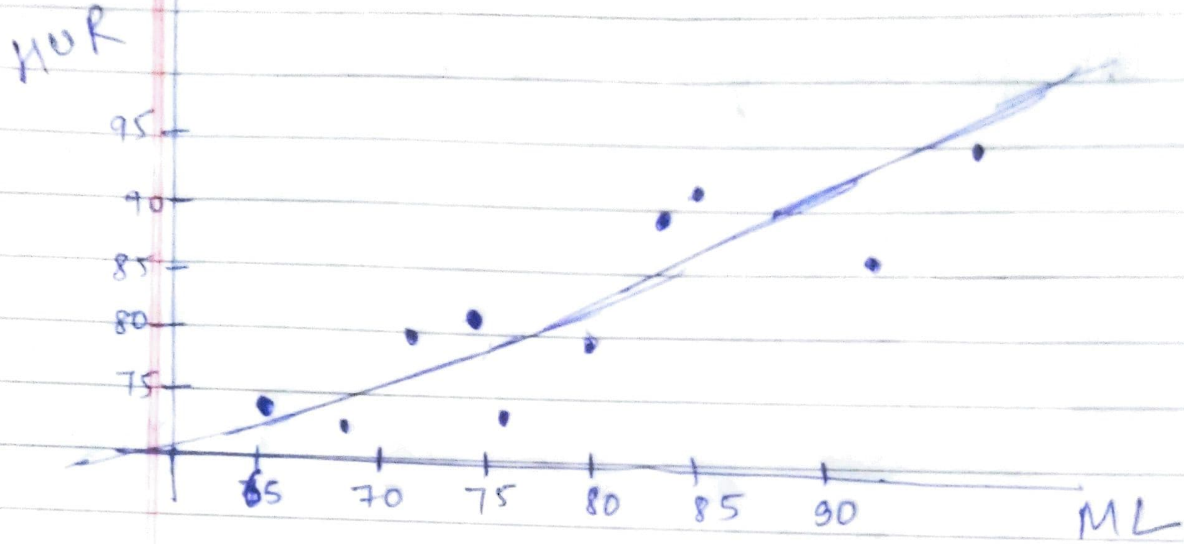
(d) If student recieve 95 in HOR
Then, expected marks in ML

$$\boxed{\begin{aligned} ML &= -14.393 + 1.50 \times 95 \\ &= 128.107 \end{aligned}}$$

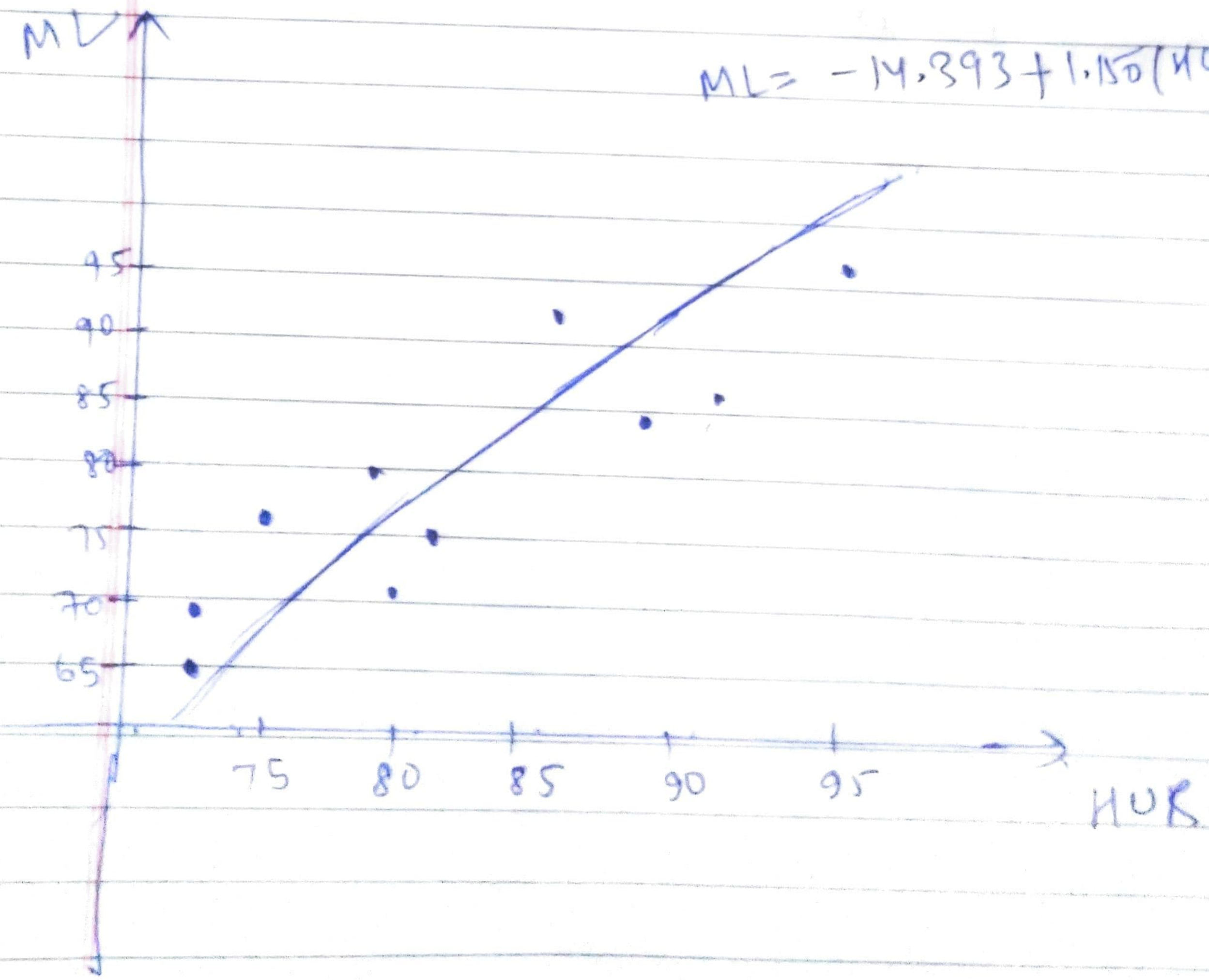
(e)

HUR = 29.1290 + 0.6612(ML)

y = 29.1290 + 0.6612x



ML = −14.393 + 1.150(HUR)



Q3)

V	54.3	61.8	72.4	88.7	118.6	194
P	61.2	49.5	37.5	28.4	19.2	10.1

$$PV^n = C$$

Ans:-

$$\log(PV^n) \quad \log C$$

$$\log P + n\log V = \log C$$

Let $Y = \log P$, $X = \log V$

$\cancel{X = \log C - n\log}$ $Y = \log C - nX$

let $w_0 = \log C$

$w_1 = -n$

$$Y = w_0 + w_1 X$$

This eqn become like simple linear regression like

To find w_0 and w_1

log(V) X	1.7348	1.7910	1.8597	1.9479	2.0741	2.2878
log(P) Y	1.78	1.69	1.57	1.45	1.28	1.00

	Y	X	X^2	XY
	1.78	1.73	3.0095	3.0997
	1.69	1.79	3.2076	3.0350
	1.57	1.85	3.4586	2.9273
	1.45	1.94	3.7944	2.8310
	1.28	2.07	4.3018	2.6617
	1.00	2.28	5.2340	2.2977
Σ	8.79	11.69	23.0061	16.8523

$A = \Sigma x$ $\quad$ $C = \Sigma x^2$

$B = \Sigma y$ $\quad$ $D = \Sigma xy$

A 11.69 $\quad$ $C = 23.0061 \simeq 23.00$

$B = 8.79$ $\quad$ $D = 16.8523 \simeq 16.85$

$$w_0 = \frac{BC - AD}{Cm - A^2} \qquad m = 6$$

$$= \frac{202.17 - 196.97}{138.00 - 136.65}$$

$$= \frac{5.20}{1.35} = 3.85$$

$$w_1 = \frac{AB - Dm}{A^2 - Cm} = \frac{102.75 - 101.1}{-1.35}$$

$$= \frac{1.65}{-1.35} = -1.222$$

$w_1 = -n$

$n = -w_1$

$n = -(-1.222) \Rightarrow n \simeq 1.2$

$w_0 = \log c$

$\log c = w_0$

$c = 10^{w_0} = 10^{3.85}$

(b) $PV^n = C$

$PV^{1.2} = 7079.457$

(c) Estimate value of P when $V = 100$

$$P = \frac{C}{V^n}$$

$$P = \frac{7079.457}{(100)^{1.2}} = \frac{7079.457}{251.188} \simeq 28.1838$$

$P \simeq 28.184$

(4) Find the least square parabola

$$Y = W_0 + W_1 X + W_2 X^2$$

X	0	1	2	3	4	5	6
Y	2.4	2.1	3.2	5.6	9.3	14.6	21.9

Ans:

$$J(w_0, w_1, w_2) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_w(x) - Y_{actual} \right)^2$$

$$\frac{\partial J}{\partial w_0} = 0, \quad \frac{\partial J}{\partial w_1} = 0, \quad \frac{\partial J}{\partial w_2} = 0$$

will result

$$\Sigma Y_i = w_0 m + w_1 \Sigma X + w_2 \Sigma X^2$$

$$\Sigma XY = w_0 \Sigma X + w_1 \Sigma X^2 + w_2 \Sigma X^3$$

$$\Sigma X^2 Y = w_0 \Sigma X^2 + w_1 \Sigma X^3 + w_3 \Sigma X^4$$

Let $A = \Sigma X$ $\qquad A = 21$

$B = \Sigma Y$ $\qquad B = 59.1$

$C = \Sigma X^2$ $\qquad C = 91$

$D = \Sigma X^3$ $\qquad D = 441$

$E = \Sigma X^4$ $\qquad E = 2275$

$F = \Sigma XY$ $\qquad F = 266.9$

$G = \Sigma X^2 Y$ $\qquad G = 1367.5$

Date
Page

Now, eqⁿ will become.

$$59.1 = 7w_0 + 21w_1 + 91w_2$$

$$266.9 = 21w_0 + 91w_1 + 441w_2$$

$$1367.5 = 91w_0 + 441w_1 + 2275w_2$$

$$w_0 = 2.5095$$
$$w_1 = -1.200$$
$$w_2 = 0.7333$$

So eqⁿ of line

$$Y = 2.5095 - 1.2X + 0.733X^2$$