ASSIGNMENT-3

1)

Basis	Simple Linear Regression	Multiple Linear Regression
Independent variables	One	Two or more
Equation	$y = \beta_0 + \beta x$	$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots$
Complexity	Simple	More Complex
Interpretation	Easy	Comparatively Difficult
Usage	One factor affects output	Multiple factors affect output

2) Given :

X (sunshine hours)	Y (ice-creams sold)
2	4
3	5
5	7
7	10
9	15

Step-1 : Required Value

$n = 5$

$\Sigma x = 26$ $\quad \Sigma x^2 = 168$

$\Sigma y = 41$ $\quad \Sigma y^2 = 269$

Step-2 : Calculated Slope (b)

$$b = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2} = \frac{5(269) - 26(41)}{5(168) - 26^2} = \frac{279}{196} = 1.42$$

Step-3 : Calculate Intercept (a)

$a = \bar{y} - b\bar{x} \Rightarrow a = 8.2 - (1.42)(5.2) = 0.81$

$\therefore$ Regression Equation $\Rightarrow y = 0.81 + 1.42x$.

3) Given :

X	Y
1	68
2	77
2	81
3	82
4	88
5	90

a) Regression Line-

$\bar{x} = 2.83$, $\bar{y} = 81$

Slope $b = 4.69$

Intercept $a = 67.72$

$y = 67.72 + 4.69x$

$$\text{Slope} = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\Sigma(x - \bar{x})^2} = \frac{54.99}{11.84} = 4.69$$

$$\text{Intercept}(a) = \bar{y} - b\bar{x} = 81 - (4.69)(2.83) = 67.72$$

Name:_____ Regd. Number:_____

b) Sum of Squares Error (SSE) = $\Sigma(y-\hat{y})^2$

Actual y	Predicted y	$y-\hat{y}$	$(y-\hat{y})^2$
68	72.41	-4.41	19.45
77	77.10	-0.10	0.01
81	77.10	3.90	15.21
82	81.79	0.21	0.04
88	86.48	1.52	2.31
90	91.17	-1.17	1.37

$\therefore SSE = 19.45 + 0.01 + 15.21 + 0.04 + 2.31 + 1.37$
$= \underline{27.04}$

c) Sum of Squares Regression (SSR) = $\Sigma(\hat{y}-\bar{y})^2 \Rightarrow \bar{y} = 81$

Predicted $\hat{y}$	$\hat{y}-\bar{y}$	$(\hat{y}-\bar{y})^2$
72.41	-8.59	73.80
77.10	-3.90	15.21
77.10	-3.90	15.21
81.79	0.79	0.62
86.48	5.48	30.03
91.17	10.17	103.00

$\therefore SSR = 73.80 + 15.21 + 15.21 + 0.62 + 30.03 + 103.00$
$= \underline{310.86}$

d) Sum of Squares Total (SST) = $SSE + SSR = 27.04 + 310.86 = \underline{337.90}$.

e) Coefficient of Determination (r^2) = $\frac{SSR}{SST} = \frac{310.86}{337.90} = \underline{0.92}$.

f) $MSE = \frac{SSE}{n-2} = \frac{27.04}{4} = \underline{6.76}$.

$S = \sqrt{MSE} = \underline{2.60}$.

4) Given:

x	x^2	y	y^2	xy
1	1	2	4	2
2	4	3	9	6
3	9	5	25	15
4	16	7	49	28
5	25	8	64	40

$\Sigma x = 15$ $\quad n = 5$
$\Sigma y = 25$
$\Sigma x^2 = 55$
$\Sigma y^2 = 151$
$\Sigma xy = 91$

Name:_____ Regd. Number:_____

$$r = \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[n\Sigma x^2 - (\Sigma X)^2][n\Sigma Y^2 - (\Sigma Y)^2]}}$$

$$= \frac{5(91) - 15(25)}{\sqrt{[5(55) - 15^2][5(151) - (25)^2]}}$$

$$= \frac{455 - 375}{\sqrt{(275 - 225)(755 - 625)}}$$

$$= \frac{80}{\sqrt{50 \times 130}} = \frac{80}{\sqrt{6500}} = \frac{80}{80.62} = 0.997.$$

5) a) $X' = \frac{X - X_{min}}{X_{max} - X_{min}}$

Record	Age'	Income'
1	0.00	0.79
2	0.25	0.03
3	0.14	0.17
4	0.66	0.00
5	0.07	0.83
6	0.39	0.33
7	0.71	0.17
8	0.73	0.89
9	0.64	0.79
10	1.00	0.44

b) Euclidean Distance from Record

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Record	Distance
9	0.44
8	0.47
7	0.60
6	0.61
4	0.71

c) Neighbours (Nearest) for $k = 3 \rightarrow$ Records = 9, 8, 7.

Name:_____ Regd. Number:_____

b) Sum

Ad

c) Sum e

Pr

d) Sum e

e) Coeffi

f) MSE

S

4) Given:

$\Sigma X = 15$
$\Sigma Y = 25$
$\Sigma X^2 = 55$
$\Sigma Y^2 = 151$
$\Sigma XY = 91$

6) Small k:

Advantages: • Captures local patterns
• Low bias

Drawbacks: • Sensitive to noise
• Overfitting

Large k:

Advantages: • Smooth decision boundary
• less-noise sensitivity

Drawbacks: • High bias
• Underfitting