Rebecca Dias
182027/19
TE CMPNA2

ML

classmate

Date
Page

Experiment NO:-02

Aim: i) To implement linear regression in statistical method

ii) To implement linear regression in scikit learn.

Theory:

- What is linear regression?

Ans: ① It is a statistical method that allows us to summarize and study relationships between two continuous (quantitative) variables

② One variable (x) is predictor, explanatory or independent variable

③ One variable (y) is response, outcome or dependant variable.

④ Multiple linear regression involves two or more predictors variables

⑤ Deterministic relationship is the equation which describes the relationship between these two variables.

- How do you compute the coffiients of regression line?

Ans: The normal equations can be solved simultaneously for b_0 and b_1

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

18 2027 /19
TE CMPNA2

ML

$$b_0 = \frac{1}{n}(\Sigma y_i - b_i \Sigma x_i) = \bar{y} - b_1 \bar{x}$$

where x and y are the means of the x_i and y_i abrevations respectively

- How do you check for the correctness of the model?

Ans: The correctness of the model can be checked using the Karl pearsons coffficient and R-square value

i) Karl pearsons correlation coefficient

$$r = \frac{N\Sigma xy - \Sigma x \Sigma y}{\sqrt{[N\Sigma x^2 - (\Sigma x)^2][N\Sigma y^2 - (\Sigma y)^2]}}$$

It has a value between -1 and $+1$, 1 is total positive linear ~~regression~~ correlation; 0 is no linear correlation and -1 is total negative linear correlation.

ii) R-square (R^2)

$$R^2 = 1 - (SSE / SSTO).$$

Example.

Height	Weight
151	63
174	81
138	56
18·6	91
128	47
136	57
179	76
163	72
152	62
131	48

x	y	$x_i-\bar{x}$	$y_i-\bar{y}$	$(x_i-\bar{x})(y_i-\bar{y})$	$(x_i-\bar{x})^2$	xy	x^2	y^2
151	63	−2·8	−2·3	6·44	7·84	9513	22801	3969
174	81	26·2	15·7	311·14	408·04	14094	30276	6561
138	56	−15·8	−9·3	146·44	249·64	7728	19044	3136
186	91	32·2	25·7	827·54	1036·8	16926	34596	8281
128	47	−25·8	−18·3	472·14	665·64	8896	16384	2209
136	57	−17·8	8·3	147·74	316·84	7752	18496	3249
179	76	25·2	10·7	269·64	635·04	13604	32041	5776
163	72	9·2	6·7	61·64	84·64	11736	26569	5184
152	62	−1·8	3·3	5·94	3·24	9424	23104	3844
131	48	−22·8	−17·3	394·44	519·84	6288	17161	2304
$\bar{x}=$	$\bar{y}=$			$\Sigma=$	$\Sigma=$	$\Sigma=$	$\Sigma=$	$\Sigma=$
153·8	65·3			2649·6	3921·6	163081	294452	44513

$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2} = \frac{2649.6}{3927.6} = 0.6746$$

$$b_0 = \frac{1}{n}\left(\Sigma y_i - b_1 \Sigma x_i\right) = \bar{y} - b_1\bar{x}$$

$$= 65.3 - (0.6746 \times 153.8)$$

$$= -38.4581$$

Karl pearsons coffiicient,

$$r = \frac{(10 \times 103681) - (1538 \times 653)}{\sqrt{[(10 \times 240272) - (1538)^2][(10 \times 44513) - 653^2]}}$$

$$r = 0.9771$$

$$\hat{y} = b(x_i) + b_0$$

x	y	$\hat{y}$	SSR	SST	SSE
181	63	63.411	3.57	5.29	0.1689
179	81	78.927	185.67	246.49	4.2971
138	56	54.641	113.61	86.49	1.8466
186	91	87.622	471.86	660.49	15.824
128	47	47.895	302.93	334.89	0.8009
136	57	53292	144.195	68.89	13.75
179	71	82.3	289.803	114.49	39.69
163	72	71.806	38.519	44.89	0.2438
152	62	64.088	1.475	10.89	4.3498
131	48	49.919	236.581	299.29	3.6818
Σ	$\bar{y} = 65.3$		1781.44	1812.1	84.6524

$$R^2 = SSR/SST = 1781.44/1812.1 = 0.9547$$

$$R = \sqrt{R^2} = 0.9771$$