Jessica Braganza
F17112151
BE Comp 2

PAGE NO :
DATE :

ML Practical 1

Title : Linear Regression

1) What are applications of linear regression ?

Ans 1. Linear Regression is a very powerful statistical technique and can be used to generate insights on consumer behaviour, understanding business and factors influencing probability.

2. Linear regressions can be used in business to evaluate trends and make estimates or forecasts.

3. For example, if a company's sales have increased steadily every month for the past few years, by conducting a linear analysis on the sales data with monthly sales, the company would forecast sales in future months.

4. Linear regression can also be used to analyze the marketing effectively, pricing and promotions on sales of a product.

5. For instance, if company XYZ wants to know if the funds that they have invested in marketing a particular brand has given them substantial return on investment, they can use linear regression.

6. It enables us to capture the isolated impacts of each of the marketing campaigns along with controlling the factors that could influence the sales.

2) What are important function used for linear regression while program implementation and explain their purpose ?

Ans. - It is assumed that the two variables are linearly related. Hence, we try to find a linear function that predicts the response value (y) as accurately as possible as a function of the feature or independent variable (x). Functions are as follows :

1) Dataset for the Linear Model creation :-

Import make_regression ~~model~~ method and create the dataset which contains 1 feature which is x and the target is Y.

Now let's split the dataset into training and validation parts. This can be done with sklearn.model_selection.train_test_split method.

2) Relation between x_train and y_train variables:

Our objective here is to make a model which takes x as input and predicts the corresponding y variable.

Model(x) = y

Consider an input x = -2, then Model(-2) would return the value close to -200.

If the input x = 2, then Model(2) would return the value close to 100.

Model(x) = $w_1 * x + w_0$

$w_1$ and $w_0$ are the parameters, which will be learnt by model during the training. The line which appeared on the graph is the best-fit line for the train set.

Q 3) Solve the problem for given dataset in problem statement and find values of $b_0$ and $b_1$ in equation

Ans.

| Number of hours spent driving (x) | Risk score on a scale of 0-100 (y) |
|---|---|
| 10 | 95 |
| 9 | 80 |
| 2 | 10 |
| 15 | 50 |
| 10 | 45 |
| 16 | 98 |
| 11 | 38 |
| 16 | 93 |

DATE :

| X | Y | $X^2$ | X.Y |
|---|---|---|---|
| 10 | 95 | 100 | 950 |
| 9 | 80 | 81 | 720 |
| 2 | 10 | 4 | 20 |
| 15 | 50 | 225 | 750 |
| 10 | 45 | 100 | 450 |
| 16 | 98 | 256 | 1568 |
| 11 | 38 | 121 | 418 |
| 16 | 93 | 256 | 1488 |

$\Sigma X = 89$ $\Sigma Y = 509$ $\Sigma X^2 = 1143$ $\Sigma X.Y = 6364$

$$\text{Slope } (b_1) = \frac{n \times \Sigma X.Y - \Sigma X . \Sigma Y}{n \times \Sigma X^2 - (\Sigma X)^2}$$

$$= \frac{8 \times 6364 - (89 \times 509)}{8 \times 1143 - (89)^2}$$

$$= \frac{5611}{1223}$$

$$= 4.58$$

$$\text{Intercept } (b_0) = \frac{\Sigma Y - b_0 \times \Sigma X}{n}$$

$$= \frac{509 - 4.58 \times 89}{8}$$

$$= \frac{101.38}{8}$$

$$= 12.67$$

$$y = 4.58x + 12.67$$