—: LAB-04 :—

Date __/__/__
Page ____

A Linear Regression :—

a relation between one dependent variable {the outcome we want to predict} and one or more independent variable {the predictors}

A assumsation of a linear relationship between the independent variables and dependent variables.

Involving only one independent variable.

To find a line $y = mx + c$ i.e. minimizes the error b/w the predicted values (y) and actual values

Algo

01 Define the Problem :

Select the dependent variable y (target variable) and the independent Variables x, (features)

02 Initialize Parameters :

Set initial values for the co-efficients ($\beta_0 \ldots$), also the learning rate.

MSE → Average of squared difference b/w actual and predicted values.

03 Compute Prediction:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon$$

04 Calculate MSE: {Mean Squared Error}

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2$$

N → observations
$\hat{y}_i$ → predicted values
y_i → actual values

☆ Applications:

- Analyzing risk in financial systems
- Forecasting sales or revenue
- Estimating trends in data
- Predicting student satisfaction

☆ Pseudocode for Linear Regression:

```
Function LinearRegression(x, y):
  # step 1: Add a column of ones to x for
  the intercept term
  x = AddColumnOfOnes(x)

  # step 2: Compute the coefficients using
  the OLS formula
  # beta = (X^T * X)^-1 * X^T * Y
  X_transpose = Transpose(x)
  XTX = multiply(X_transpose, x)
  XTX_inverse = Inverse(XTX)
```

Date / /
Page

XTY = multiply (x_transpose, y)
beta = multiply (XTX_inverse, XTY)

step3: Return the co-efficients
Return beta.

Multiple Linear Regression :-

Multiple independent variable $(x_1, x_2, \ldots x_n)$ and a single dependent variable (y).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_n x_n$$

y is dependent variable.
$x_1, x_2, \ldots x_n$ are independent variables
β_0 is intercept.
$\beta_1, \beta_2 \ldots \beta_n$ are the slopes

Let datapoints be $(x_1, \ldots x_n, y_1 \ldots)$
$\forall_i (0, \ldots m)$ where $x_i \forall i \in \{0, 0\}$.

represents independent variables as y values are dependent variables.

In the form of matrix

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}^{y} = \begin{bmatrix} 1 + x_1 + x_2 + \ldots x_{n_1} \\ 1 + x_{1 2} + \ldots + x_{n 2} \\ 1 + x_{1 n} + \ldots x_m \end{bmatrix}$$

where $\beta = ((x^T x)^{-1} x^T) y$

The above values can be used to plot the best fit line and can be used to predict future values.

☆ Logistic Regression :-

- Logistic regression approach operates on Sigmoid curve rather than best fit line, we got a value $\in [0,1]$ (Binary Classification) and then classify into +ve or -ve by comparing with median.

Let data points be $(x_i, y_i)\ \forall\ i \in [0, n]$, finding but fit line through previously mentioned methods

$$V = \frac{1}{1+e^{-(mx+c)}} = \frac{1}{1+e^{-(b_1 m + b x)}}$$

Classification will be based on the obtained value V.

* If $V < 0.5 \rightarrow$ then "no"
* If $V > 0.5 \rightarrow$ then "yes".