Date 17/3/25
Page

Lab-03

To Solve the following linear Regression Problem using matrix Approach

To find linear regression of data of week & product sales

x_i (week)	y_j (Sales in thousands)
1	2
2	H
3.	5
H	9

① matrix X:

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & H \end{bmatrix} \qquad Y = \begin{bmatrix} 2 \\ H \\ 5 \\ 9 \end{bmatrix}$$

② Computing matrix $X^T X$ & $X^T y$

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & H \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & H \end{bmatrix}$$

$$= \begin{bmatrix} H & 10 \\ 10 & 30 \end{bmatrix}$$

Date 4/3/22
Page

$$X^T y = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \\ 5 \\ 9 \end{bmatrix}$$

$$= \begin{bmatrix} 20 \\ 54 \end{bmatrix}$$

③. Inverse of $X^T X$

$$(X^T X)^{-1} = \frac{1}{20} \begin{bmatrix} 30 & -10 \\ -10 & 4 \end{bmatrix}$$

④ We know that,

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

i.e $$\beta = (X^T X)^{-1} X^T y$$

$$= \frac{1}{20} \begin{bmatrix} 30 & -10 \\ -10 & 4 \end{bmatrix} \begin{bmatrix} 20 \\ 54 \end{bmatrix}$$

$$\beta = \begin{bmatrix} 1 \\ 1.6 \end{bmatrix}$$

⑤. Regression line equation

$$y = \beta_0 + \beta_1 x$$

$$y = 1 + 1.6x$$

Date 21/2/25
Page

1. Considering the data file: Did you perform any data preprocessing steps? If yes, how?

→ canada_per_capita_income.csv: No preprocessing steps were required as the data file has no missing values or categorical variables.

→ salary.csv: There were missing values present in "Years of Experience" & "Salary" which was processed & filled using the "mean".

→ Hiring.csv: "experience" values were missing which were filled as "zero" & missing "test_score" were filled using mean.

2. For canada_per_capita_income.csv, did you visualize the regression line along with the data points? What does the plot tell you about the relationship between year & per capita income?

Yes, the regression line was plotted along with data points, the plot shows a strong positive linear relationship between year & per-capita-income, indicating that as year increases the per-capita-income also increase consistently over time.

Date 19/3/25
Page

3. For hiring.csv, what is the predicted salary for candidate with 12 years of experience, 10 test score & 10 interview score?

For candidate, the predicted salary is $92268.0422

4. For 1000_companies.csv, did you encode categorical variables (eg: state)? If yes, how? Did you scale the features? If yes, why?

→ categorical encoding -
The 'state' column was encoded using OneHotEncoder to convert categorical state to binary column

→ Feature Scaling: Scaling was not applied, while scaling can improve the model performance sometimes for features with vastly different scales

A 19/3/25