

Birla Institute of Technology and Science, Pilani, Hyderabad
Campus

MACHINE LEARNING FOR ELECTRONICS ENGINEERS LAB

Experiment 4

Simple linear regression using Python

Q1. A company wants to analyze the relationship between Years of Experience and Salary (in lakhs) for entry-level data scientists. Given the data below, perform the following analysis

Years of Experience (x_i)	Salary in Lakhs
1	2
2	4
3	5
4	4
5	5

a) Write a Python code to generate the above given data (use NumPy) and import the Scikit-learn library to calculate the regression coefficients. Verify the coefficients with the hand calculation (you can use either formula or matrix-based approach)

Note: Since it is a small sample-sized data, there is no need to use a pandas dataframe

b) Write a code to calculate the residual standard error (RSE) and the R^2 statistic using the formula discussed in the lecture class. Verify your results with hand calculation, and verify the R^2 statistic you have calculated with the built-in `r2_score` command of the scikit learn library.

Note: For model predictions, you can use the `model.predict` command

c) Use the fitted model to predict the salary of the employees for the given years of experience using hand calculation (table below). Verify the hand-calculated results with the `model.predict` command.

Years of Experience (x_i)	Predicted Salary
6	
10	
12	

Q2. Load the California housing dataset from the scikit learn library (the same way you did it in the Experiment 3 lab) and perform simple linear regression (one feature at a time) for only the following three features: 'HouseAge', 'AveRooms', and 'AveBedrms' (do not worry about other features, consider only these three features). Note that each feature has approximately 2000 samples/data points. Fit a simple linear regression equation for each of the above three features separately (For example, Estimated Price = $\beta_0 + \beta_1 \times \text{HouseAge}$). Estimate/Predict the house price using the fitted model for the following feature values (you can use the `model.predict` command)

a) When HouseAge is 10 years (Using only Estimated Price = $\beta_0 + \beta_1 \times$

HouseAge)

b) When AveRooms is 8 rooms

c) When AveBedrms is 4 bedrooms

Note: If time permits, you can fit a simple linear regression model by exploring other features in the housing data, such as “Population”, “AveOccup”, “Longitude”, and “Latitude”, by considering them separately.