Linear Regression Example: Salary vs Experience

Aymen Negadi

1 Linear Regression in Machine Learning

Linear Regression is a supervised Machine Learning algorithm used to model the relationship between a dependent variable and one independent variable. It assumes that this relationship can be approximated by a linear function.

The objective of linear regression is to find the best-fitting straight line that minimizes the difference between the predicted values and the actual observed data. This line is generally expressed by the equation:

$$y = ax + b$$

where a represents the slope of the line and b represents the intercept.

2 Example Description and Dataset

In this example, linear regression is applied to study the relationship between an employee's years of experience and their salary.

The dataset consists of two variables:

- x: Years of experience
- y: Salary measured in thousands of DA

The data used in this experiment is shown below:

Experience (Years)	Salary (DA)
1	30
2	35
3	45
4	50
5	60

Each data point corresponds to one employee. The goal is to learn a linear model that describes the general trend between experience and salary.

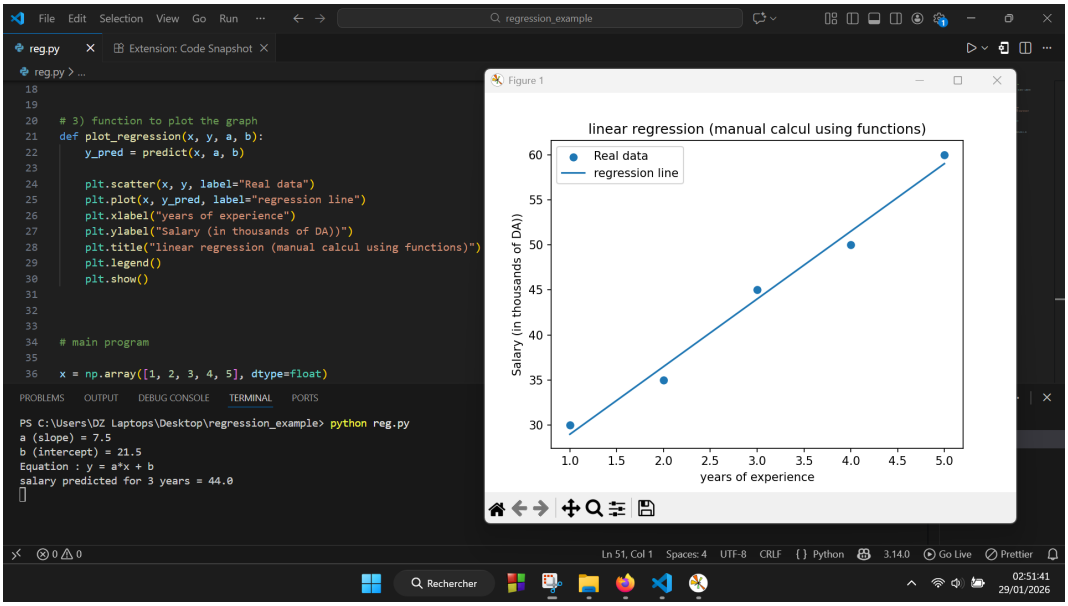

Figure 1: Linear regression result: Salary as a function of experience

3 Regression Result Visualization

Figure 1 illustrates the result of the linear regression model.

The blue points represent the original observed data, while the blue line represents the regression line obtained from the model.

4 Result Interpretation

The regression line represents the best linear approximation of the relationship between years of experience and salary. Although the line does not pass exactly through all data points, it minimizes the overall squared error between predicted and observed values.

The positive slope of the line indicates that salary increases as experience increases. From the model, the salary grows by approximately 7.5 thousand DA for each additional year of experience. This confirms the existence of a strong positive linear relationship between the two variables.

5 Conclusion

This example demonstrates how linear regression can be used to model and analyze a real-world relationship between two variables. By visualizing the regression line alongside the original data points, it becomes easier to interpret the model and understand the underlying data trend.