**Name: Sahil Vishwas Shendkar**

**Roll No: TYITB118**

**Assignment No. 2**

**AIM:** Assignment on Linear Regression on Salary Dataset.

**OBJECTIVE: To apply Linear Regression on the Salary Dataset to analyze the relationship between variables and predict outcomes based on input features.**

**PREREQUISITE:** Python programming

**THEORY:**

**Linear Regression is a basic and essential technique in statistics and machine learning that helps us predict an output (or target) value using one or more input (or predictor) variables. It tries to find a straight-line relationship between the inputs and the output. This method is commonly used for forecasting and estimating relationships in various fields like business, healthcare, economics, and engineering.**

**Linear Regression can be of two types:**

* **Simple Linear Regression, which uses one independent variable to predict the outcome.**
* **Multiple Linear Regression, which uses two or more independent variables.**

**Its simplicity, efficiency, and interpretability make it a good starting point for anyone learning about predictive modeling and machine learning.**

**DATABASE:**

**In this assignment, we use the Train Dataset, a structured file containing past records or observations used to train the model. The dataset is usually in the form of rows and columns, with each row representing one data entry and each column representing a feature or attribute.**

**Typical features in the Train Dataset include:**

* **Inputs like age, experience, house size, etc.**
* **Output (target) like salary, sales, house price, or score.**

**This dataset is used to teach the model how different input values relate to the output. Once the model learns this relationship, it can be used to make predictions on new or unseen data. Data is usually split into a training set (for learning) and a testing set (for evaluating performance).**

**Simple Linear Regression**

**Simple Linear Regression focuses on predicting one outcome using one input. For example, predicting salary based on years of experience. The method assumes a straight-line relationship where one variable has a direct and measurable effect on the other.**

**Key components of this model include:**

* **Dependent Variable (Target): The value we want to predict.**
* **Independent Variable (Feature): The input used for prediction.**
* **Intercept: The point where the prediction starts when the input is zero.**
* **Slope: How much the output increases or decreases when the input increases by one unit.**

**Estimating Coefficients**

**To make the linear regression model useful, we calculate values that define the line – the intercept and the slope. These values are chosen based on the dataset in such a way that the predicted values are as close as possible to the actual values. This process involves comparing the differences between the actual and predicted outcomes and adjusting the line until the smallest possible error is achieved.**

**Making Predictions**

**Once the model is trained and the coefficients are known, it can be used to predict outcomes for new inputs. For example, if we input the years of experience of a new employee, the model will predict the expected salary. These predictions can be visualized by drawing a straight line through a set of data points, showing how the values are expected to change.**

**Assessing Model Performance**

**To check how good the model is, we use performance metrics. One commonly used metric is the Root Mean Squared Error (RMSE), which tells us how far off, on average, our predictions are from the actual values. A smaller value means the model is more accurate.**

**Another important metric is R-squared, which tells us how well the model explains the variation in the data. A higher R-squared means the model fits the data better.**

**These tools help us know whether our model is reliable for making predictions on new data.**

**Importance of Linear Regression**

**Linear Regression is highly valuable because of its:**

* **Simplicity and Interpretability: Easy to understand and explain.**
* **Speed and Efficiency: Works quickly even with large datasets.**
* **Widespread Use: Applied in forecasting trends, estimating prices, predicting outcomes, and more.**

**It is commonly used in:**

* **Economics for predicting inflation or demand.**
* **Healthcare for estimating patient outcomes.**
* **Marketing for forecasting sales.**
* **Finance for stock or loan default prediction.**

**Linear Regression is also a stepping stone for learning more complex models in machine learning.**

**CONCLUSION:**

**Linear Regression is one of the most fundamental techniques in predictive modeling. It helps us understand and quantify the relationship between variables and use that relationship to make future predictions. With the use of a well-structured Train Dataset, it becomes possible to analyze trends, estimate outcomes, and guide data-driven decision-making. Even though it is simple, it provides a solid foundation for building and understanding more advanced algorithms.**