

Expt No: 7

Date : 08-09-25

Linear Regression

Aim :

To build a linear regression model using Python to predict the salary of an employee based on their Years of Experience using the dataset salary-data.csv.

Algorithm :

1. Start
2. Import required libraries (numpy, pandas and sklearn) for data handling and modeling.
3. Load the dataset and read the csv file.
4. Explore and clean the data
 - use info() and describe() to understand the dataset.
 - Handle missing values using dropna()
5. Prepare the data
6. Split the dataset
7. Train the model
8. Evaluate the model
9. Save and reload the model
10. Make predictions
11. Display result and save.

Program:

```
import numpy as np
import pandas as pd
df = pd.read_csv("Salary-data.csv")
df
df.info()
df.dropna(inplace=True)
df.info()
df.describe()
features = df.iloc[:, [0]].values
label = df.iloc[:, [1]].values
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(features, label, test_size=0.2, random_state=42)
from sklearn.linear_model import LinearRegression
model = LinearRegression()
model.fit(x_train, y_train)
model.score(x_train, y_train)
model.score(x_test, y_test)
model.coef_
model.intercept_
import pickle
pickle.dump(model, open('Salary Pred.model', 'wb'))
model = pickle.load(open('Salary Pred.model', 'rb'))
yr_of_exp = float(input("Enter Years of Experience:"))
yr_of_exp_NP = np.array([[yr_of_exp]])
```

```
Salary = model.predict(yr_of_exp_NP)
print ("Estimated salary for {} years of experience is: "
      .format(yr_of_exp_Salary))
```

Output:

```
<class 'pandas.core.frame.DataFrame'>
Range Index: 30 entries, 0 to 29
Data columns (total 2 columns):
 #   Column          Non-Null Count  Dtype  
--- 
 0   Years Experience    30 non-null   float64
 1   Salary            30 non-null   int64  
dtypes: float64(1), int64(1)
memory usage: 612.0 bytes
```

```
<class 'pandas.core.frame.DataFrame'>
Range Index: 30 entries, 0 to 29
Data columns (total 2 columns):
 #   Column          Non-Null Count  Dtype  
--- 
 0   Years Experience    30 non-null   float64
 1   Salary            30 non-null   int64
```

salary

The salary of an employee depends upon his/her experience. This is a linear relationship. We can implement a Linear Regression model to predict the salary based on years of experience.

We will use Python's scikit-learn library to implement this model. We will use the Boston Housing dataset which contains information about house prices in Boston. We will use the 'MedInc' column as the target variable and the 'RM' column as the feature variable.

We will first import the required libraries:

```
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn import metrics
```

We will then load the dataset and split it into training and testing sets:

```
df = pd.read_csv('BostonHousing.csv')
X = df['RM'].values.reshape(-1, 1)
y = df['MedInc'].values.reshape(-1, 1)

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

We will then create a Linear Regression model and fit it to the training data:

```
regressor = LinearRegression()
regressor.fit(X_train, y_train)
```

We will then make predictions on the testing data:

```
y_pred = regressor.predict(X_test)
```

We will then calculate the Mean Absolute Error (MAE) to evaluate the performance of the model:

```
mae = metrics.mean_absolute_error(y_test, y_pred)
```

The result of the prediction is:

Result: $\text{y} = 3.58 \times \text{RM} + 2.70$

The Linear Regression model was successfully implemented to predict an employee's salary based on years of experience.