



Machine Learning

Experiment No – 01

Aim: Regression Analysis and Plot Interpretation.

Theory: Regression analysis is a statistical method to model the relationship between a dependent (target) and independent (predictor) variables with one or more independent variables. More specifically, Regression analysis helps us to understand how the value of the dependent variable is changing corresponding to an independent variable when other independent variables are held fixed.

A Practical approach to Simple Linear Regression :

Simple Linear Regression is a statistical method that allows us to summarize and study relationships between two continuous (quantitative) variables. One variable denoted x is regarded as an independent variable and other one denoted y is regarded as a dependent variable. It is assumed that the two variables are linearly related. Hence, we try to find a linear function that predicts the response value (y) as accurately as possible as a function of the feature or independent variable (x):

The simplest form of the regression equation with one dependent and one independent variable is defined by the formula:

$$Y = ax + b$$

Here

Y :- is Dependent Variable

X :- is Independent Variable

a:- is Slope of the line

b:- is y intercept of the Line

Dependent Variable is also called as outcome variable, criterion variable, endogenous variable, or regress and.

Independent Variable is also called as exogenous variables, predictor variables, or regressors.

Implementation of Simple Linear Regression Algorithm

Problem Statement example for Simple Linear Regression:

Here we are taking a dataset that has two variables: salary (dependent variable) and experience (Independent variable). The goals of this problem is:

- We want to find out if there is any correlation between these two variables
- We will find the best fit line for the dataset.
- How the dependent variable is changing by changing the dependent variable.

Write a program to implement the Simple Linear regression model in machine learning and Take the printout and Attached with output .

(Refer Salary_Data .csv dataset)

Let's see the complete code in Python as follows –

Simple Linear Regression

Importing the libraries

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

Importing the dataset

```
dataset = pd.read_csv('Salary_Data.csv')
X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, 1].values
```

Splitting the dataset into the Training set and Test set

```
from sklearn.cross_validation import train_test_split
```



```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 1/3, random_state = 0)
```

Fitting Simple Linear Regression to the Training set

```
from sklearn.linear_model import LinearRegression  
regressor = LinearRegression()  
regressor.fit(X_train, y_train)
```

Predicting the Test set results

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

Visualising the Training set results

```
plt.scatter(X_train, y_train, color = 'red')  
plt.plot(X_train, regressor.predict(X_train), color = 'blue')  
plt.title('Salary vs Experience (Training set)')  
plt.xlabel('Years of Experience')  
plt.ylabel('Salary')  
plt.show()
```

Visualising the Test set results

```
plt.scatter(X_test, y_test, color = 'red')  
plt.plot(X_train, regressor.predict(X_train), color = 'blue')  
plt.title('Salary vs Experience (Test set)')  
plt.xlabel('Years of Experience')  
plt.ylabel('Salary')  
plt.show()
```

Questions:

1. Define Machine Learning and Explain type of Machine Learning algorithms
2. Define Cross Validation Technique? State type of Cross validation Technique. And Explain K-fold Cross Validation Technique.

(Subject In-charge)

(Prof.S.B.Mehta)