SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)
Accredited "A" Grade by NAAC | 12B Status by UGC | Approved by AICTE
www.sathyabama.ac.in

SCSA2601-Machine Learning and Data Analytics Lab

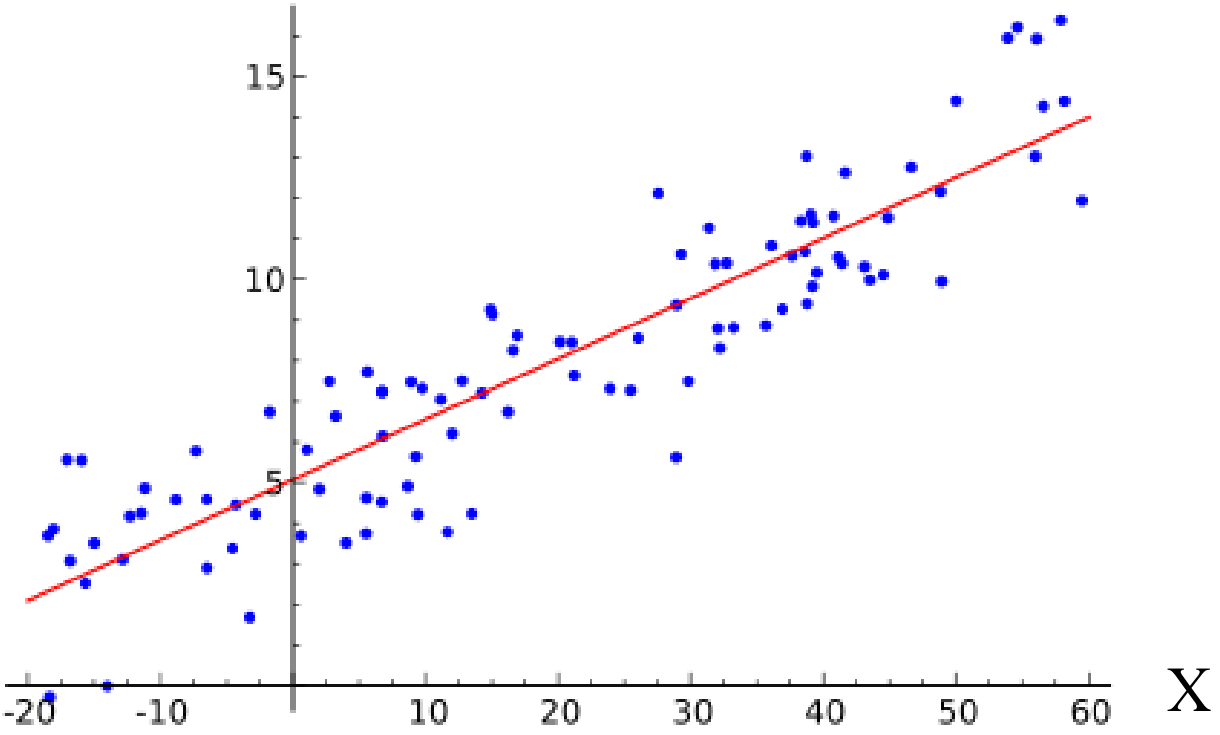
Cycle-2

Mrs. M. VANATHI

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Linear Regression

- We want to find the best line (linear function y=f(X)) to explain the data.



Linear Regression

- The predicted value of y is given by:

$$\hat{y} = \hat{\beta}_0 + \sum_{j=1}^{p} X_j \hat{\beta}_j$$

- The vector of coefficients $\hat{\beta}$ is the regression model.

- If $X_0 = 1$, the formula becomes a matrix product:

$$\hat{y} = \mathrm{X}\,\hat{\beta}$$

Regression

Example: Price of a used car

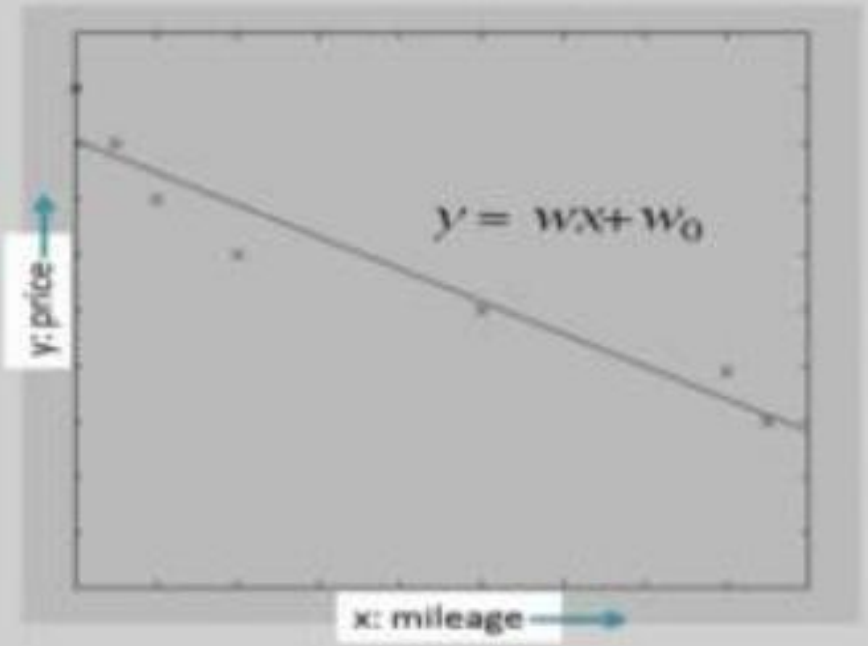
x : car attributes

y : price

$$y = g(x, \theta)$$

$g(\ )$ model,

θ parameters



Python Code to implement Linear Regression

Input

A Dataset and the X value to predict future Y.

Apply Regression Algorithm

Output

Scatter Plot and Best Regression Line and Predicted Y Value

Program

```
import pandas as pd
import numpy as np
dict={"Experience":[1,2,3,4,5],
"Salary":[20000,40000,50000,40000,
50000]}
df=pd.DataFrame(dict)
df
x=df.iloc[:,0].values
y=df.iloc[:,1].values
def LinearRegression(x,y):
  N = len(x)
  x_mean = x.mean()
  y_mean = y.mean()
  B1_num = ((x - x_mean) * (y -
y_mean)).sum()
  B1_den = ((x - x_mean)**2).sum()
  B1 = B1_num / B1_den
  B0 = y_mean - (B1*x_mean)
  reg_line = 'y = {} + {}β'.format(B0,
round(B1, 3))
  return (B0, B1, reg_line)
def predict(B0, B1, new_x):
  y = B0 + B1 * new_x
  return y
  B0, B1, reg_line
=LinearRegression(x,y)
  pred=predict(B0,B1,8)
  pred
```

Program Contd.,

```
import matplotlib.pyplot as plt
plt.figure(figsize=(12,5))
plt.scatter(x, y, s=300, linewidths=1,
edgecolor='black')
text = '''X : {} Years
Y : ${}
y = {} +
{}X'''.format(round(x.mean(),
2),round(y.mean(), 2),round(B0,
3),round(B1, 3))
plt.text(x=1, y=10, s=text, fontsize=12,
bbox={'facecolor': 'grey', 'alpha': 0.2,
'pad': 10})
plt.title('Salary over Experience')
plt.xlabel('Years of Experience',
fontsize=15)
plt.ylabel('Salary', fontsize=15)
plt.plot(x, B0 + B1*x, c = 'r', linewidth=5,
alpha=.5, solid_capstyle='round')
plt.scatter(x=x.mean(), y=y.mean(),
marker='', s=10*2.5, c='r')
```

Output

X : 3.0 Years
Y : $40000.0
y = 22000.0 + 6000.0X