21MIS1152- Rajeev Sekar

Importing necessary modules and loading the dataset

```
# Making the imports
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
data = pd.read_csv('data.csv')
data.head()
```

```
   32.502345269453031  31.70700584656992
0           53.426804          68.777596
1           61.530358          62.562382
2           47.475640          71.546632
3           59.813208          87.230925
4           55.142188          78.211518
```

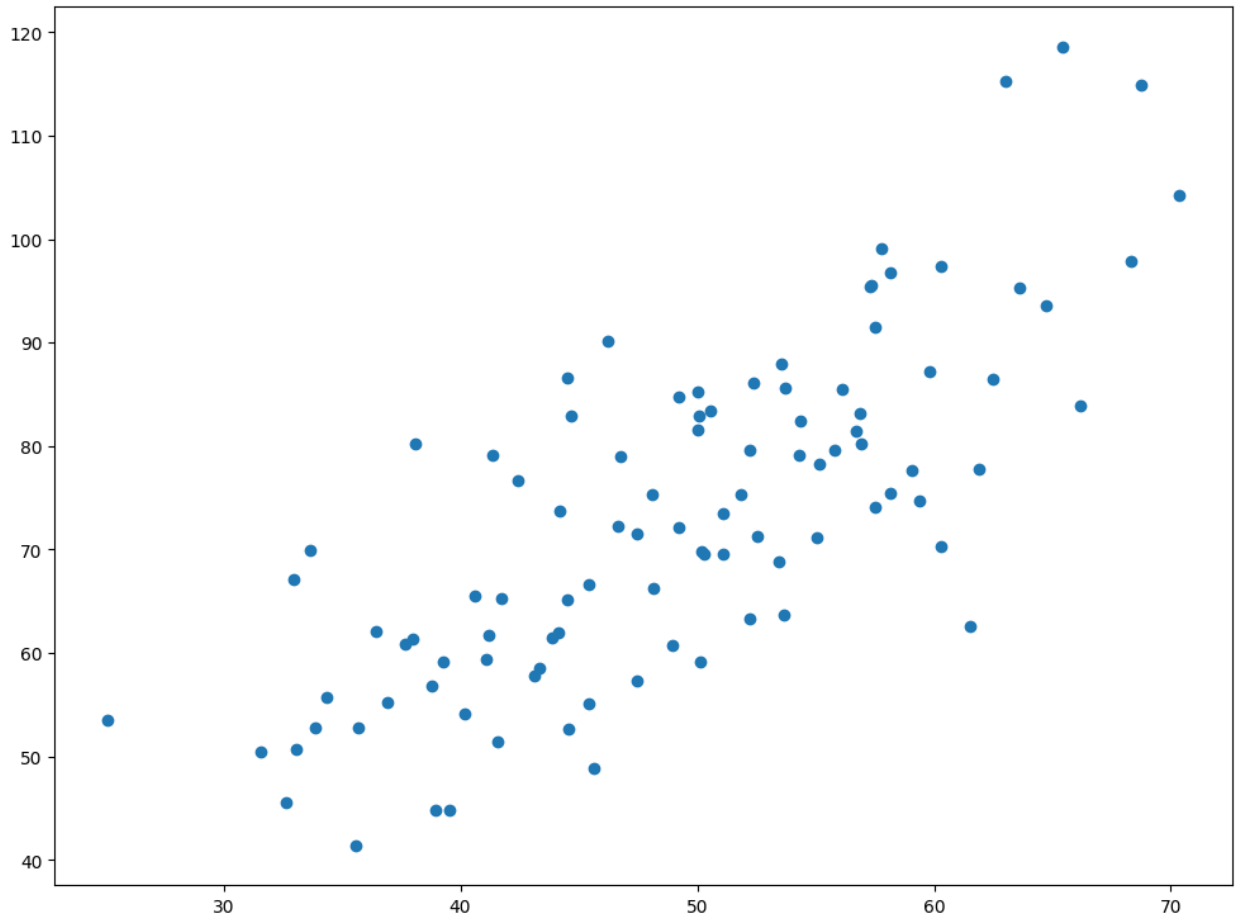
DATA PREPROCESSING: Splitting feature and target variable and finding the correlation b/w them

```
X = data.iloc[:, 0]
Y = data.iloc[:, 1]

from scipy.stats import spearmanr
corr, _ = spearmanr(X, Y)
print('correlation: %.3f' % corr)
```

```
correlation: 0.753
```

the correlation is greater that 0.6 => linear regresion can be applied

```
#plotting data
plt.scatter(X, Y)
plt.show()
```



MODEL BUILDING

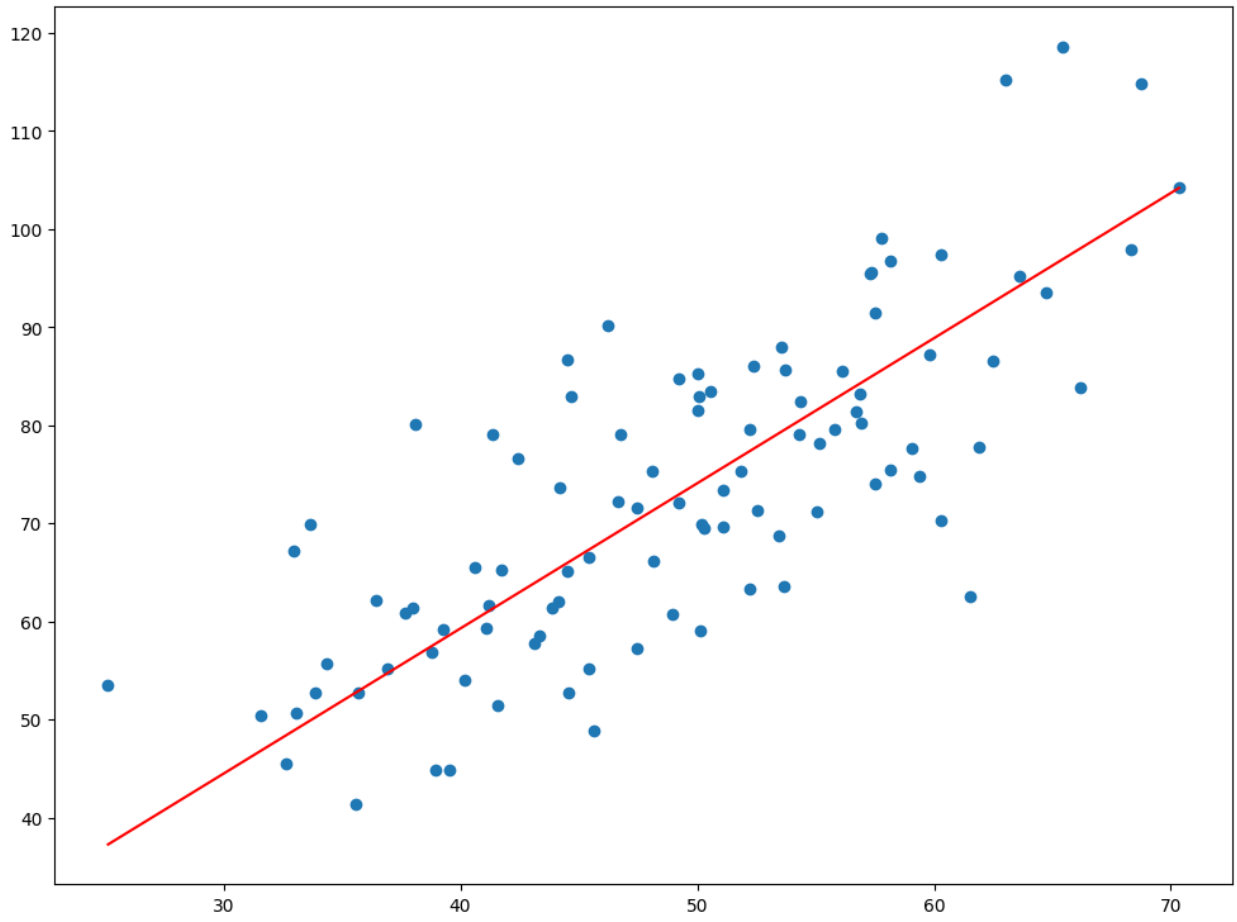
- first we initialize m(slope) and c(intercept) as 0
- Learning rate is initialised as 0.0001 and we iterate 1000 times to update m and c to minimize the cost function

```
m = 0
c = 0

L = 0.0001
iters = 1000  # The number of iterations to perform gradient descent
n = float(len(X))

# Performing Gradient Descent
for i in range(iters):
    Y_pred = m*X + c  # The current predicted value of Y
    D_m = (-2/n) * sum(X * (Y - Y_pred))  # Derivative wrt m
    D_c = (-2/n) * sum(Y - Y_pred)  # Derivative wrt c
    m = m - L * D_m  # Update m
    c = c - L * D_c  # Update c

print (m, c)
```

```
1.4796491688889395 0.10148121494753734
```

VISUALISATION of the regression line after gradient descent

```
# Making predictions
Y_pred = m*X + c

plt.scatter(X, Y)
plt.plot([min(X), max(X)], [min(Y_pred), max(Y_pred)], color='red')  # regression line
plt.show()
```



MODEL EVALUATION using R-square

```
from sklearn.metrics import r2_score
r2 = r2_score(Y, Y_pred)
print('r2 score: ', r2)

r2 score:  0.5735604046292567
```

MODEL INTERPRETATION

=> slope of the equation is 1.47 which means that increase in 1 unit of the feature variable increases the target by 1.47 times