Simple Linear Regression

Import Libraries

In [1]:
```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

Import Dataset

In [2]:
```
dataset = pd.read_csv("Salary_Data.csv")
X = dataset.iloc[ : ,  : -1].values
y = dataset.iloc[ : , -1].values
```

In [3]:
```
print(X)
print("\n\n\n")
print(y)
```

```
[[ 1.1]
 [ 1.3]
 [ 1.5]
 [ 2. ]
 [ 2.2]
 [ 2.9]
 [ 3. ]
 [ 3.2]
 [ 3.2]
 [ 3.7]
 [ 3.9]
 [ 4. ]
 [ 4. ]
 [ 4.1]
 [ 4.5]
 [ 4.9]
 [ 5.1]
 [ 5.3]
 [ 5.9]
 [ 6. ]
 [ 6.8]
 [ 7.1]
 [ 7.9]
 [ 8.2]
 [ 8.7]
 [ 9. ]
 [ 9.5]
 [ 9.6]
 [10.3]
 [10.5]]
```

```
[ 39343.  46205.  37731.  43525.  39891.  56642.  60150.  54445.  64445.
  57189.  63218.  55794.  56957.  57081.  61111.  67938.  66029.  83088.
  81363.  93940.  91738.  98273. 101302. 113812. 109431. 105582. 116969.
 112635. 122391. 121872.]
```

Splitting data into training and testing datasets

In [4]:
```
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 1/4, random_stat
```

In [5]:
```
print(X_train)
print("\n\n\n")
print(X_test)
print("\n\n\n")
print(y_train)
print("\n\n\n")
print(y_test)
```

```
[[ 5.3]
 [ 7.9]
 [ 2.9]
 [ 5.1]
 [ 3.2]
 [ 4.5]
 [ 8.2]
 [ 6.8]
 [ 1.3]
 [10.5]
 [ 3. ]
 [ 2.2]
 [ 5.9]
 [ 6. ]
 [ 3.7]
 [ 3.2]
 [ 9. ]
 [ 2. ]
 [ 1.1]
 [ 7.1]
 [ 4.9]
 [ 4. ]]
```

```
[[ 1.5]
 [10.3]
 [ 4.1]
 [ 3.9]
 [ 9.5]
 [ 8.7]
 [ 9.6]
 [ 4. ]]
```

```
[ 83088. 101302.  56642.  66029.  64445.  61111. 113812.  91738.  46205.
 121872.  60150.  39891.  81363.  93940.  57189.  54445. 105582.  43525.
  39343.  98273.  67938.  56957.]
```

```
[ 37731. 122391.  57081.  63218. 116969. 109431. 112635.  55794.]
```

Training the model

In [6]:
```
from sklearn.linear_model import LinearRegression

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

Out[6]:
```
LinearRegression()
```

Predicting y_train and y_test

In [7]:
```
y_train_pred = linear_regression.predict(X_train)
y_test_pred = linear_regression.predict(X_test)
```

Visualise Training dataset

In [8]:
```
plt.scatter(X_train, y_train, color = "green", marker = "+", label = "Observed data")
plt.plot(X_train, y_train_pred, color = "red", label = "Predicted data")
plt.xlabel("Years of experience")
plt.ylabel("Salary")
plt.title("Years of experience v/s Salary (Training dataset)")
plt.legend()
plt.show()
```

Visualise Testing dataset

In [9]:
```
plt.scatter(X_test, y_test, color = "green", marker = "+", label = "Observed data")
plt.plot(X_test, y_test_pred, color = "red", label = "Predicted data")
plt.xlabel("Years of experience")
plt.ylabel("Salary")
plt.title("Years of experience v/s Salary (Testing dataset)")
plt.legend()
plt.show()
```


Years of experience v/s Salary (Testing dataset)
Predicted data
Observed data
120000
100000
80000
60000
40000
Salary
2
4
6
8
10
Years of experience