Task Details:

Prediction using Supervised ML.

Predict the percentage of an student based on the no. of study hours.

This is a **simple linear regression** task as it involves just 2 variables.

Dataset - [http://bit.ly/w-data (http://bit.ly/w-data)](http://bit.ly/w-data)

Problem Statement:

What will be predicted score if a student studies for 9.25 hrs/day?

Prediction:

In [4]:
```
# importing the required libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, mean_squared_error
```

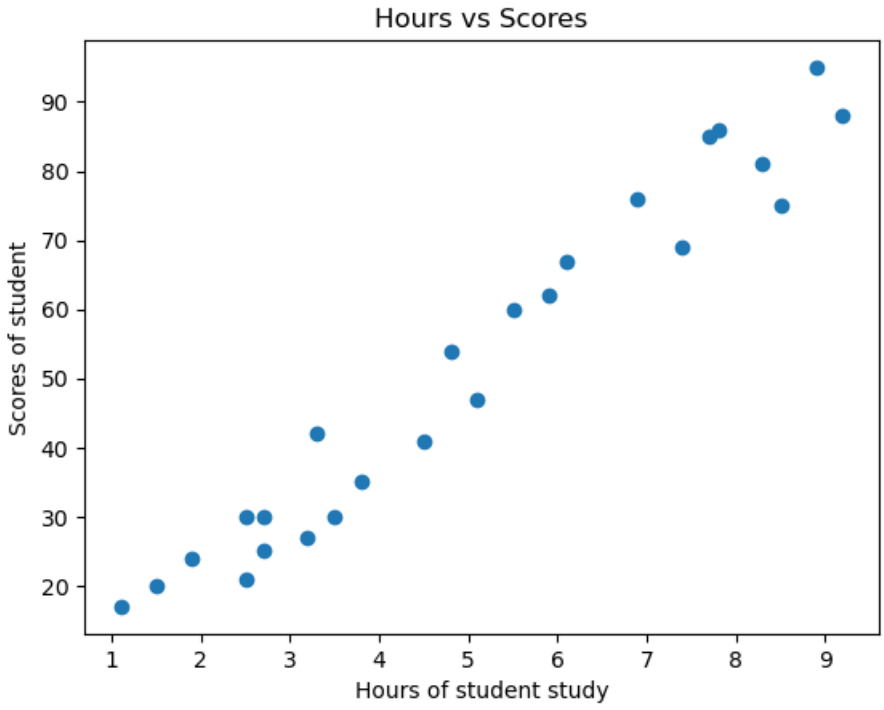
In [5]:
```
# data collection and exploration
dataset_url = "http://bit.ly/w-data"
dataset = pd.read_csv(dataset_url)
dataset.head()
```

Out[5]:

	Hours	Scores
0	2.5	21
1	5.1	47
2	3.2	27
3	8.5	75
4	3.5	30

In [6]:
```
# clearing the warning of linear regression
feature_names = ['Hours']
```

In [7]:
```
# data visualization
plt.scatter(dataset['Hours'], dataset['Scores'])
plt.title('Hours vs Scores')
plt.xlabel('Hours of student study')
plt.ylabel('Scores of student')
plt.show()
```



In [8]:
```
# spliting the data into training and test data
X = dataset[feature_names]
y = dataset['Scores']
X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=0.2,random_state=0)
```

In [9]:
```
# model selection and training for 2 variables
model = LinearRegression()
model.fit(X_train, y_train)
```

Out[9]:
```
LinearRegression()
```

In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

In [10]:
```
# model evaluating
y_pred = model.predict(X_test)
ma_error = mean_absolute_error(y_test, y_pred)
ms_error = mean_squared_error(y_test, y_pred)
root_ms_error = np.sqrt(ms_error)

print(f"X_test prediction value is {y_pred}")
print(ma_error)
print(ms_error)
print(root_ms_error)
```

```
X_test prediction value is [16.88414476 33.73226078 75.357018   26.79480124 60.49103328]
4.183859899002975
21.5987693072174
4.6474476121003665
```

In [11]:
```
# making prediction
hours_predict = 9.25
prediction_value = model.predict(np.array([[hours_predict]]))
```

```
C:\Users\baps\anaconda3\lib\site-packages\sklearn\base.py:464: UserWarning: X does not have valid
feature names, but LinearRegression was fitted with feature names
  warnings.warn(
```

In [12]:
```
# printing the predicted scores
print(f"Predicted score for {hours_predict} hours of study: {prediction_value[0]}")
```

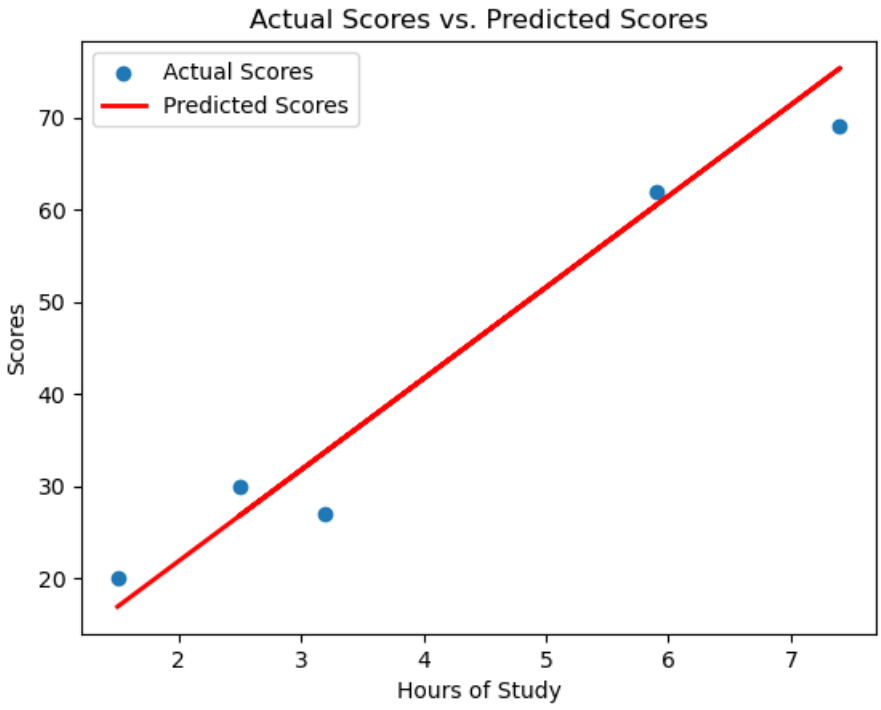
```
Predicted score for 9.25 hours of study: 93.69173248737538
```

Model Evaluation details:

In [13]:
```
# summarizing the results and visualization
print("Model Evaluation Metrics:")
print(f"Mean Absolute Error (MAE): {ma_error}")
print(f"Mean Squared Error (MSE): {ms_error}")
print(f"Root Mean Squared Error (RMSE): {root_ms_error}")
print(f"Predicted score for {hours_predict} hours of study: {prediction_value[0]}")

# Visualization of Predictions vs. Actual Data
plt.scatter(X_test, y_test, label='Actual Scores')
plt.plot(X_test, y_pred, color='red', linewidth=2, label='Predicted Scores')
plt.xlabel('Hours of Study')
plt.ylabel('Scores')
plt.title('Actual Scores vs. Predicted Scores')
plt.legend()
plt.show()
```

```
Model Evaluation Metrics:
Mean Absolute Error (MAE): 4.183859899002975
Mean Squared Error (MSE): 21.5987693072174
Root Mean Squared Error (RMSE): 4.6474476121003665
Predicted score for 9.25 hours of study: 93.69173248737538
```



Thank You!

https://github.com/anujtiwari21?tab=repositories (https://github.com/anujtiwari21?tab=repositories)