

Q4

```
In [ ]: from sklearn.linear_model import LinearRegression
        from sklearn.model_selection import train_test_split
        from sklearn.metrics import r2_score, mean_squared_error
        import numpy as np
        import pandas as pd
        from matplotlib import pyplot as plt
```

```
In [ ]: df = pd.read_csv("../data/salary.csv")
        y = df["Salary"]
        X = df["YearsExperience"]
        df.head()
```

```
Out [ ]:   YearsExperience  Salary
0           1.1    39343.0
1           1.3    46205.0
2           1.5    37731.0
3           2.0    43525.0
4           2.2    39891.0
```

```
In [ ]: X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=100, test_si
```

```
In [ ]: reg = LinearRegression()
        reg.fit(np.array(X_train).reshape(-1,1), y_train)
```

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

```
In [ ]: preds = reg.predict(np.array(X_test).reshape(-1,1))
        print(f"MSE : {mean_squared_error(y_test, preds)}")
        print(f"R2 score : {r2_score(y_test, preds)}")
```

```
MSE : 24477109.08965574
R2 score : 0.9720725422361338
```

```
In [ ]: plt.scatter(X_train, y_train, color='red', label="Train")
        plt.scatter(X_test, y_test, color='blue', label="Test")
        plt.xlabel("Years of Experience")
        plt.ylabel("Salary")
        plt.legend()

        intercept = reg.intercept_
```

```
slope = reg.coef_[0]  
plt.axline((0, intercept), slope=slope)
```

Out[]: <matplotlib.lines.AxLine at 0x297af11d6d0>

