

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
data = pd.read_csv("Salary_dataset.csv")
```

```
data.head()
```

	Unnamed: 0	YearsExperience	Salary
0	0	1.2	39344.0
1	1	1.4	46206.0
2	2	1.6	37732.0
3	3	2.1	43526.0
4	4	2.3	39892.0

Next steps: [Generate code with data](#) [New interactive sheet](#)

```
data.tail()
```

	Unnamed: 0	YearsExperience	Salary
25	25	9.1	105583.0
26	26	9.6	116970.0
27	27	9.7	112636.0
28	28	10.4	122392.0
29	29	10.6	121873.0

```
print(data.columns)
```

```
Index(['Unnamed: 0', 'YearsExperience', 'Salary'], dtype='object')
```

```
X = data['YearsExperience']
y = data['Salary']
```

```
X = np.array(X)
y = np.array(y)
```

```
X = X.reshape(-1, 1)
y = y.reshape(-1, 1)
```

```
X = X.reshape(-1, 1)
y = y.reshape(-1, 1)
```

```
print(X.shape, y.shape)
```

```
(30, 1) (30, 1)
```

```
m = 0 # slope
c = 0 # intercept
```

```
def predict(X, m, c):
    return m * X + c
```

```
def cost_function(X, y, m, c):
    n = len(y)
    y_pred = predict(X, m, c)
    cost = (1/n) * np.sum((y - y_pred) ** 2)
    return cost
```

```
cost = cost_function(X, y, m, c)
print("Initial Cost:", cost)
```

Initial Cost: 6503107277.733334

```
learning_rate = 0.01
epochs = 1000
n = len(y)

for _ in range(epochs):
    y_pred = predict(X, m, c)

    dm = (-2/n) * np.sum(X * (y - y_pred))
    dc = (-2/n) * np.sum(y - y_pred)

    m = m - learning_rate * dm
    c = c - learning_rate * dc
```

```
print("Slope (m):", m)
print("Intercept (c):", c)
```

Slope (m): 9450.854505842613
 Intercept (c): 24842.125054036598

```
y_pred = predict(X, m, c)
```

```
mse = cost_function(X, y, m, c)
print("Mean Squared Error:", mse)
```

Mean Squared Error: 31270959.47976325

```
plt.scatter(X, y, label="Actual Data")
plt.plot(X, y_pred, color='red', label="Regression Line")
plt.xlabel("Years of Experience")
plt.ylabel("Salary")
plt.title("Salary vs Experience (Linear Regression)")
plt.legend()
plt.show()
```



