```python
# Finding the regression line using Least Square Method (LSM)
def mse(Y, Y_hat, n):
    return sum([(i - j)**2 for i, j in zip(Y, Y_hat)]) / n

data = [(1, 2), (2, 4), (3, 5), (4, 4), (5, 5)]
X = [i[0] for i in data]  # data의 x 값
Y = [i[1] for i in data]  # data의 y 값
print(X, Y)

xm = sum(X) / len(X)      # X 데이터의 평균
ym = sum(Y) / len(Y)      # Y 데이터의 평균
print(f"(X 평균: {xm}, Y 평균: {ym})")   # 평균 x, y 값

X_diff = [i[0] - xm for i in data]      # x - xm 데이터
Y_diff = [i[1] - ym for i in data]      # y - ym 데이터
print(X_diff, Y_diff)

divisor = sum([i**2 for i in X_diff])   # 기울기 분모
dividend = sum([i*j for i, j in zip(X_diff, Y_diff)])   # 기울기 분자
b1 = dividend / divisor                          # 기울기 (slope)
print(dividend, divisor, b1)
b0 = 4 - b1 * 3                 # y-intercept (y-절편)
print(b0)
print(f"The regression line: y = {b1}*x + {b0}")

# Finding the Mean Squared Error (MSE)
Y_hat = [b1*i[0] + b0 for i in data]  # 회귀 직선을 이용한 예측값
mse_value = mse(Y, Y_hat, len(Y))     # mean squared error 값

print(f"MSE: {round(mse_value, 2)}")
```

```
[1, 2, 3, 4, 5] [2, 4, 5, 4, 5]
(X 평균: 3.0, Y 평균: 4.0)
[-2.0, -1.0, 0.0, 1.0, 2.0] [-2.0, 0.0, 1.0, 0.0, 1.0]
6.0 10.0 0.6
2.2
The regression line: y = 0.6*x + 2.2
MSE: 0.48
```

B

Can you find the regression line for the data [(1, 2), (2, 4), (3, 5), (4, 4), (5, 5)]?

Certainly! I'll go over the process of finding the regression line again for the given data set. The regression line is represented by the equation:

$$y = mx + c$$

Where:

- m is the slope.
- c is the y-intercept.

The formulas for m and c are:

$$m = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$$
$$c = \frac{\Sigma y - m\Sigma x}{n}$$

Using the given data:

$$[(1,2),(2,4),(3,5),(4,4),(5,5)]$$

1. Calculate the required sums:

- $\Sigma x = 1 + 2 + 3 + 4 + 5 = 15$
- $\Sigma y = 2 + 4 + 5 + 4 + 5 = 20$
- $\Sigma xy = 1 \times 2 + 2 \times 4 + 3 \times 5 + 4 \times 4 + 5 \times 5 = 66$
- $\Sigma x^2 = 1^2 + 2^2 + 3^2 + 4^2 + 5^2 = 55$

1. Use the formulas to calculate m and c:

$$m = \frac{5 \times 66 - 15 \times 20}{5 \times 55 - 15^2}$$
$$m = \frac{330 - 300}{275 - 225}$$
$$m = \frac{30}{50}$$
$$m = 0.6$$

$$c = \frac{20 - 0.6 \times 15}{5}$$
$$c = 2.2$$

So, the regression line for the given data set is:

Regenerate

$$y = 0.6x + 2.2$$

B

What is the MSE value ?

The Mean Squared Error (MSE) is calculated using the formula:

$$MSE = \frac{1}{n}\sum(y - y')^2$$

Where:

- n is the number of data points.
- y is the observed value.
- y' is the predicted value using our regression line.

Using the regression line:

$$y' = 0.6x + 2.2$$

For the given data points:

$$[(1, 2), (2, 4), (3, 5), (4, 4), (5, 5)]$$

We can calculate the predicted values:

1. $y'(1) = 0.6(1) + 2.2 = 2.8$
2. $y'(2) = 0.6(2) + 2.2 = 3.4$
3. $y'(3) = 0.6(3) + 2.2 = 4.0$
4. $y'(4) = 0.6(4) + 2.2 = 4.6$
5. $y'(5) = 0.6(5) + 2.2 = 5.2$

Compute the squared differences:

1. $(2 - 2.8)^2 = 0.64$
2. $(4 - 3.4)^2 = 0.36$
3. $(5 - 4.0)^2 = 1.0$
4. $(4 - 4.6)^2 = 0.36$
5. $(5 - 5.2)^2 = 0.04$

Now, compute the MSE:

$$MSE = \frac{0.64+0.36+1.0+0.36+0.04}{5} = \frac{2.4}{5} = 0.48$$

So, the MSE for the given data and regression line is 0.48.