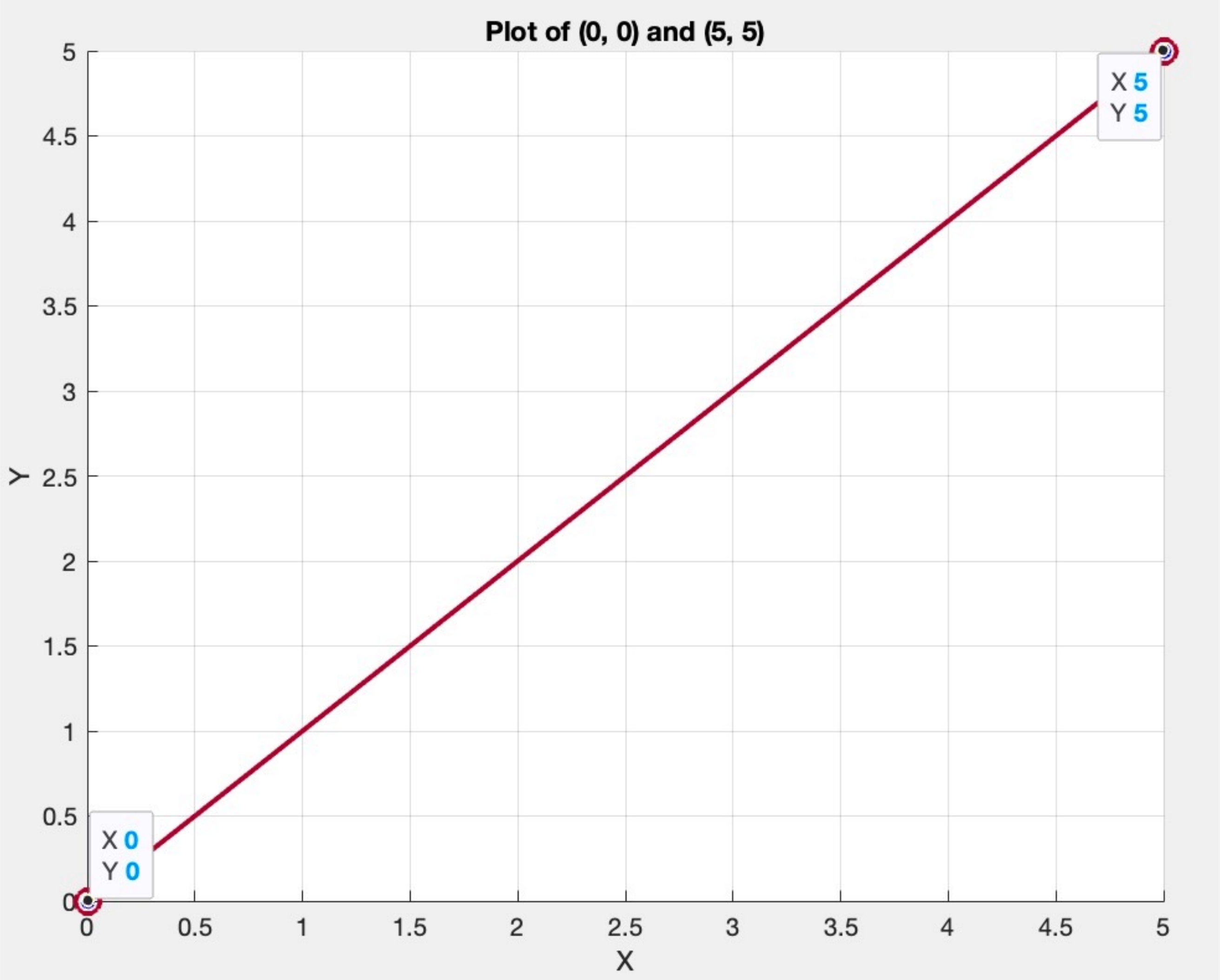
Weighted average of individual slopes

Edward Liu
006232687

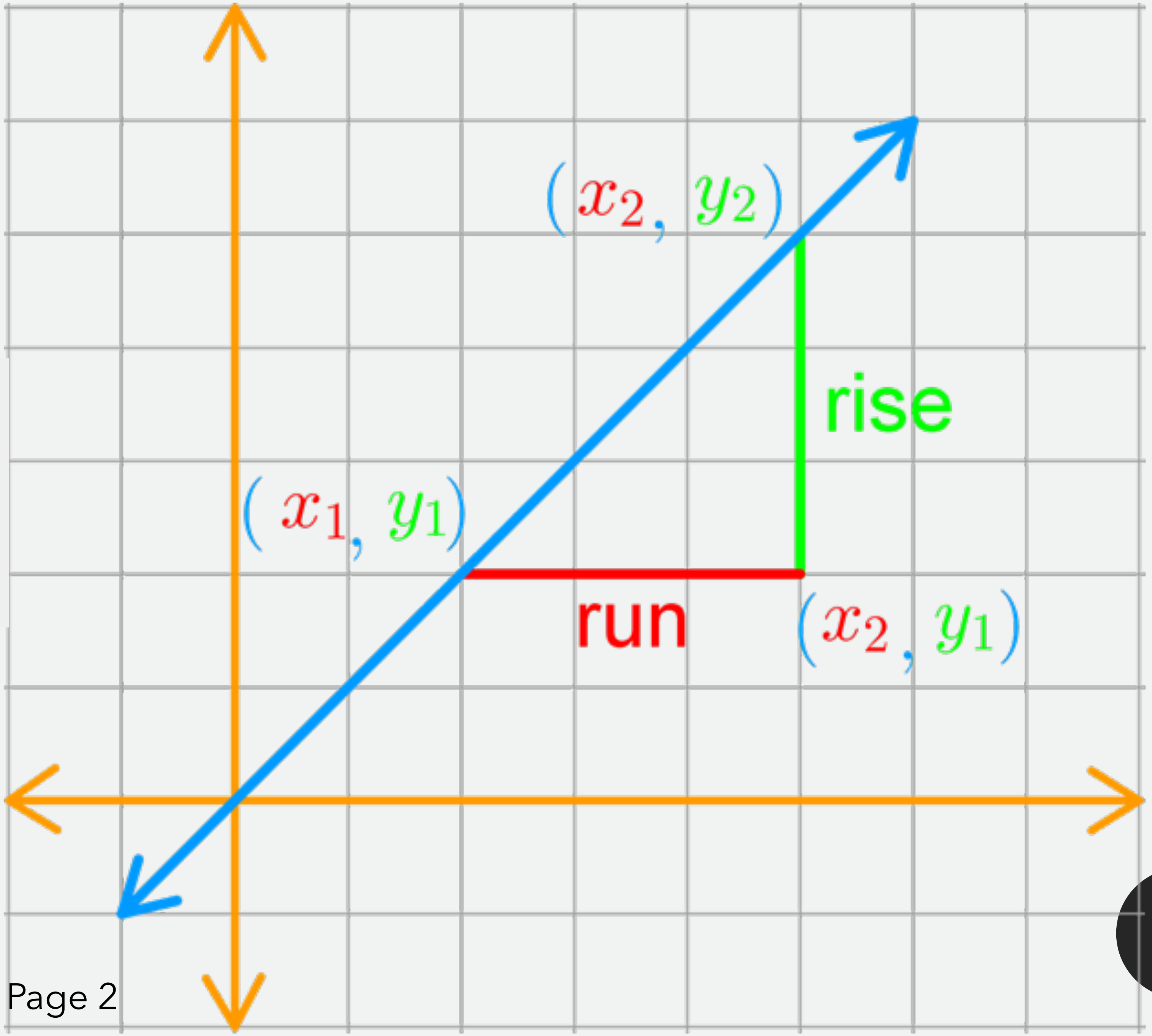
Does anyone know how to calculate slope?



Calculate slope

$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{rise}{run}$$

slope formula



- **Arithmetic mean**

$$\bar{x} = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right) = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

- **Weighted arithmetic mean**

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i},$$

which expands to:

$$\bar{x} = \frac{w_1 x_1 + w_2 x_2 + \cdots + w_n x_n}{w_1 + w_2 + \cdots + w_n}.$$

Can someone give me an example?

In what situations do we need to use weighted averages?

Least Squares Estimates (LSE)

- The least squares approach choose b_0 and b_1 that minimize the SS_{res}, i.e.,

$$(b_0, b_1) = \arg\min_{\alpha_0, \alpha_1} \sum_{i=1}^{n} (y_i - \alpha_0 - \alpha_1 x_i)^2$$

Take derivative w.r.t. α_0 and α_1, setting both equal to zero:

$$\left.\frac{\partial SS_{res}}{\partial \alpha_0}\right|_{b_0,b_1} = \sum_{i=1}^{n} \left.\frac{\partial (y_i - \alpha_0 - \alpha_1 x_i)^2}{\partial \alpha_0}\right|_{b_0,b_1} = -2\sum_{i=1}^{n} (y_i - b_0 - b_1 x_i) = 0$$

$$\left.\frac{\partial SS_{res}}{\partial \alpha_1}\right|_{b_0,b_1} = \sum_{i=1}^{n} \left.\frac{\partial (y_i - \alpha_0 - \alpha_1 x_i)^2}{\partial \alpha_1}\right|_{b_0,b_1} = -2\sum_{i=1}^{n} x_i (y_i - b_0 - b_1 x_i) = 0$$

The two equations are called the **normal equations**.

Least Squares Estimates: Solve for α_0 and α_1

- Solve for α_0 given b_1:

$$b_0 = \overline{y} - b_1 \overline{x}$$

- Solve for α_1 given $b_0 = \overline{y} - b_1 \overline{x}$:

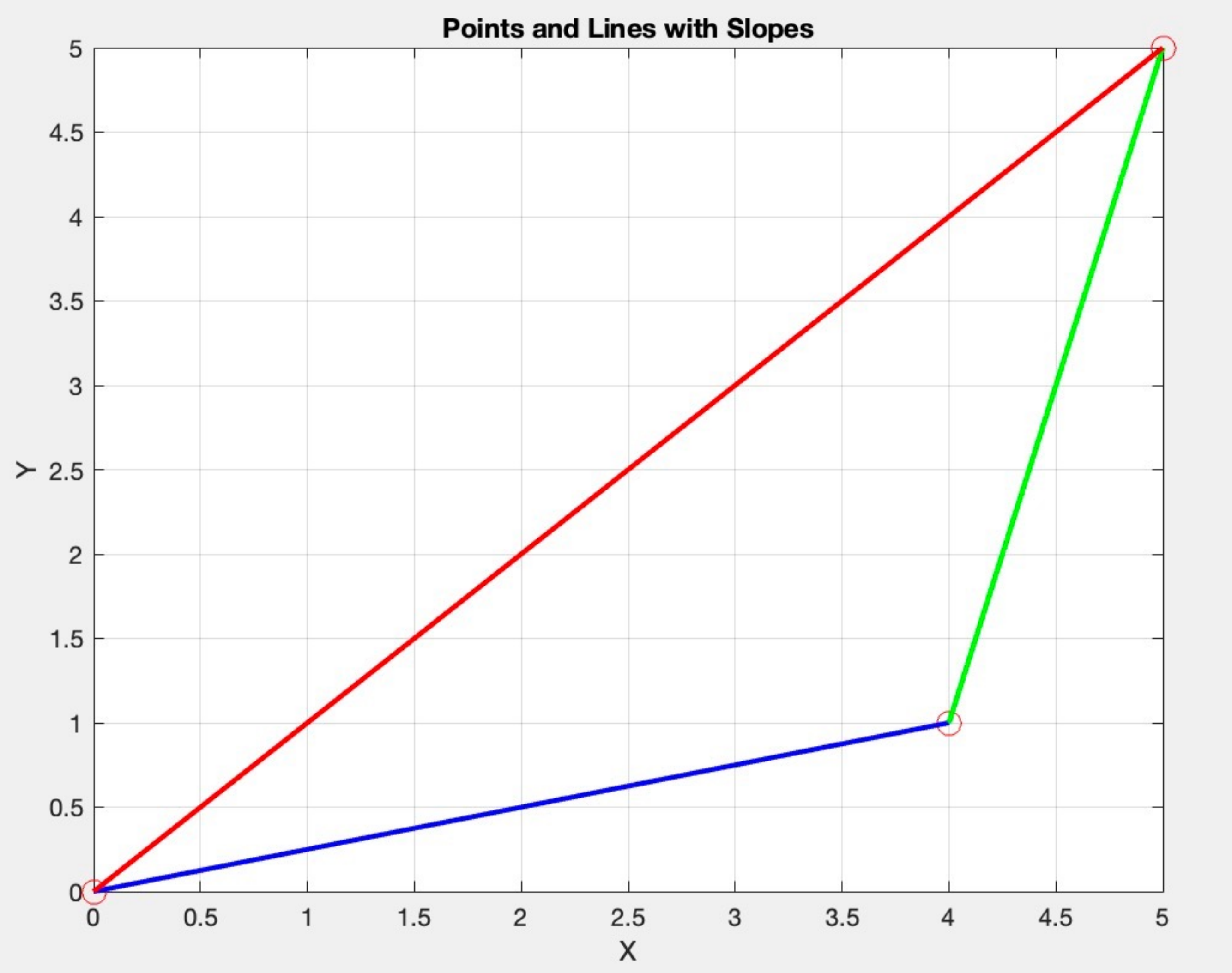
$$b_1 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} (x_i - \overline{x})^2} = \frac{S_{xy}}{S_{xx}} = r\frac{\sqrt{S_{yy}}}{\sqrt{S_{xx}}},$$

where $S_{xx} = \sum_{i=1}^{n} (x_i - \overline{x})^2$, $S_{yy} = \sum_{i=1}^{n} (y_i - \overline{y})^2$, $S_{xy} = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})$, and r is the sample correlation coefficient between x and y.

$f(X) = \beta_0 + \beta_1 X$ with unknown parameters β_0 and β_1.

In fact, the weighted average of slope$_{ij}$ weighted by the squared separation ***$(x_j - x_i)^2$*** *is* ***the least squares estimator b_1***

$$b_1 = \frac{\sum_{i,j}(x_j - x_i)^2 \text{slope}_{ij}}{\sum_{i,j}(x_j - x_i)^2}$$



- the weighted average of slopeij weighted by the squared separation $(xj - xi)$^2

```
% Define the points
x = [0, 4, 5];
y = [0, 1, 5];
z = b1 * x;
% Calculate slopes
slope1 = (y(2) - y(1)) / (x(2) - x(1)); % Slope for (0, 0) to (4, 1)
slope2 = (y(3) - y(2)) / (x(3) - x(2)); % Slope for (4, 1) to (5, 5)
slope3 = (y(3) - y(1)) / (x(3) - x(1)); % Slope for (0, 0) to (5, 5)


% Plot the points
figure;
plot(x, y, 'ro', 'MarkerSize', 10); % Red points


hold on;


% Plot the lines
plot(x, z, 'black', 'LineWidth', 2);
line([x(1), x(2)], [y(1), y(2)], 'Color', 'b', 'LineWidth', 2); % Line from (0, 0) to (4, 1)
line([x(2), x(3)], [y(2), y(3)], 'Color', 'g', 'LineWidth', 2); % Line from (4, 1) to (5, 5)
line([x(1), x(3)], [y(1), y(3)], 'Color', 'r', 'LineWidth', 2); % Line from (0, 0) to (5, 5)


xlabel('X');
ylabel('Y');
title('Points and Lines with Slopes');
grid on;
hold off;
```

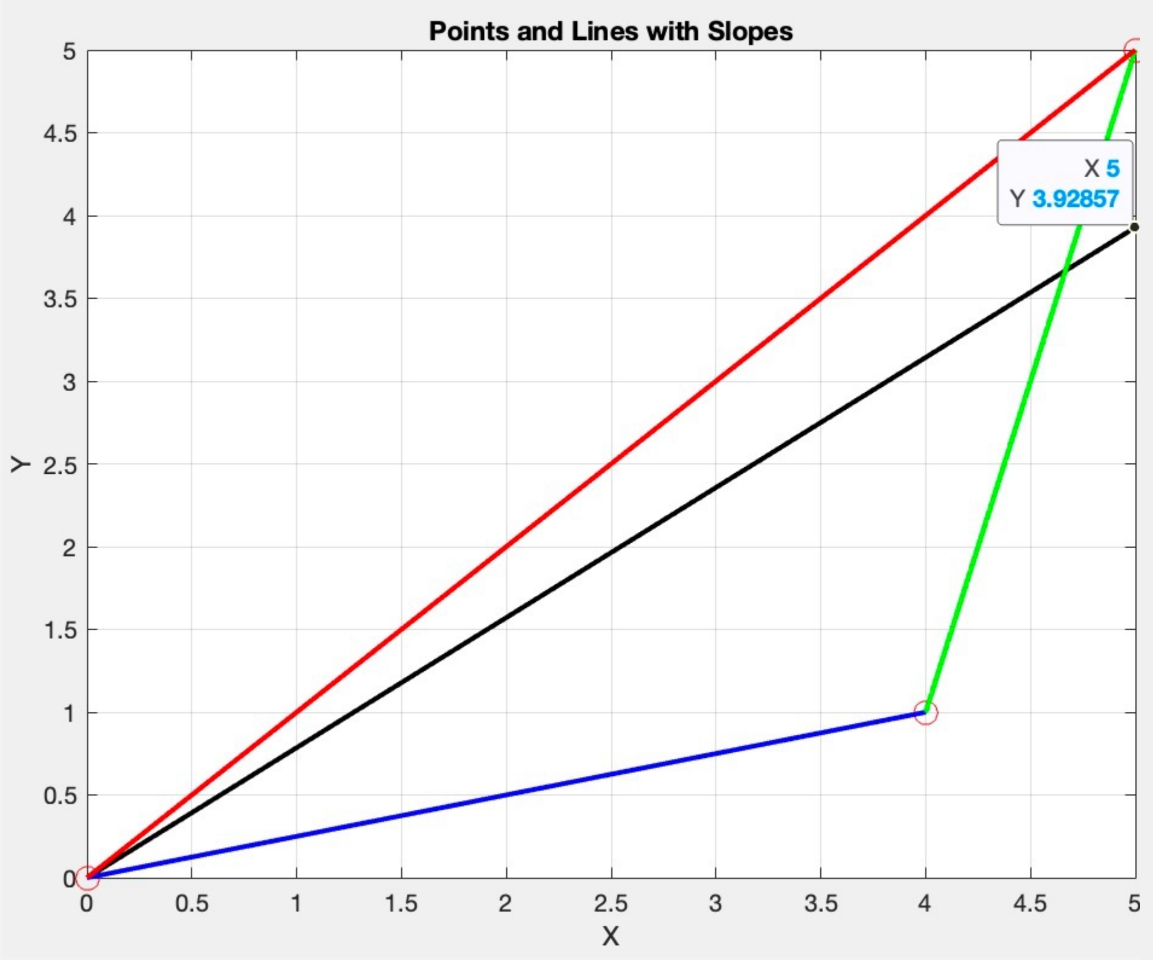

graphically

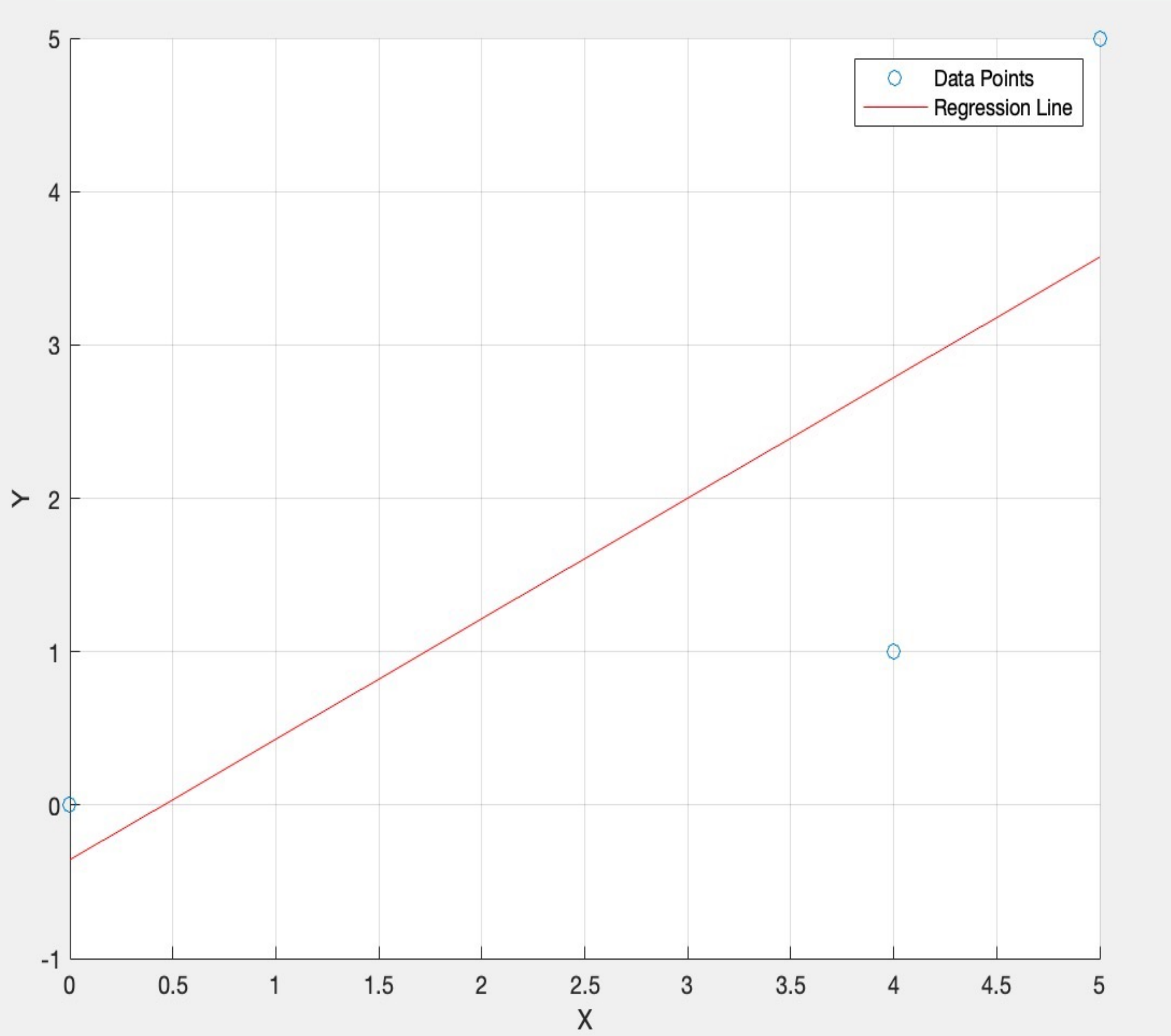
- the least squares estimator $b1$

```
% Define the slope
slope = 0.7857;


% Generate x-values (for example, from -10 to 10)
x = linspace(-10, 10, 100);


% Calculate corresponding y-values using the equation of the line
y = slope * x;


% Plot the line
plot(x, y, 'b', 'LineWidth', 2);
hold on;


% Mark the point (0, 0)
plot(0, 0, 'ro', 'MarkerSize', 10);


% Set axis labels and title
xlabel('x');
ylabel('y');
title('Line Through (0,0) with Slope 0.7857');


% Set grid
grid on;


% Display legend
legend('Line: y = 0.7857x', '(0, 0)', 'Location', 'NorthWest');


% Hold off to prevent further plotting on the same figure
hold off;
```

- the weighted average of slopeij weighted by the squared separation $(xj - xi)$^2

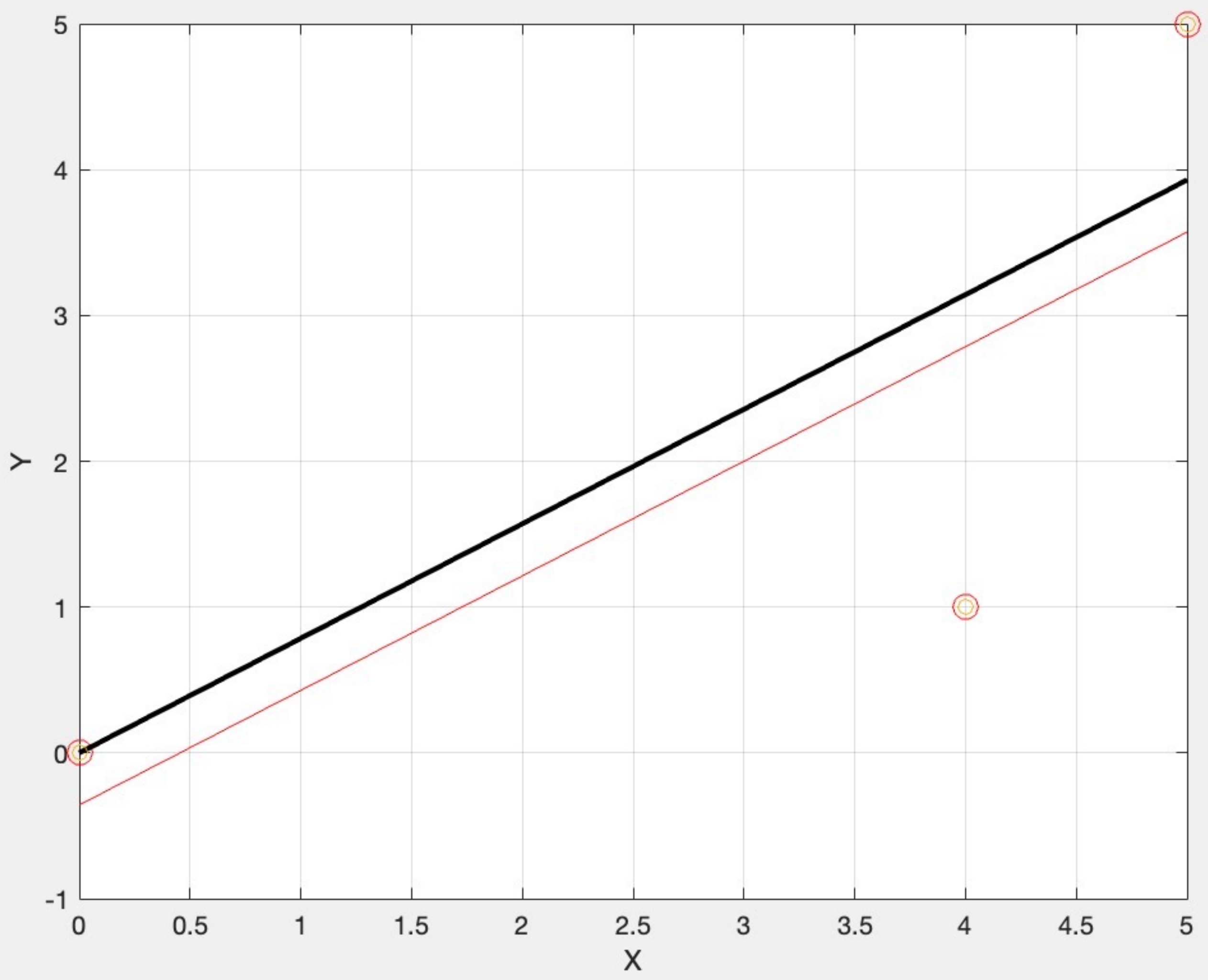
- the least squares estimator $b1$





 graphically

How about we put the two graphs together?



 graphically

- the weighted average of slopeij weighted by the squared separation $(xj - xi)$^2

1. **Calculate the Slopes (slope_{ij}) and Squared Separations ($(x_j - x_i)^2$):**

   For $(x_1, y_1) = (0, 0)$:
   - $\text{slope}_{12} = \frac{1-0}{4-0} = \frac{1}{4}$
   - $\text{slope}_{13} = \frac{5-0}{5-0} = 1$

   For $(x_2, y_2) = (4, 1)$:
   - $\text{slope}_{23} = \frac{5-1}{5-4} = 4$

   Squared Separations:
   - $(x_2 - x_1)^2 = (4 - 0)^2 = 16$
   - $(x_3 - x_1)^2 = (5 - 0)^2 = 25$
   - $(x_3 - x_2)^2 = (5 - 4)^2 = 1$

2. **Calculate the Weighted Average of Slopes (b_1):**

   $$b_1 = \frac{\sum_{i,j}(x_j - x_i)^2 \times \text{slope}_{ij}}{\sum_{i,j}(x_j - x_i)^2}$$

   Plugging in the values:

   $$b_1 = \frac{(16 \times \frac{1}{4}) + (25 \times 1) + (1 \times 4)}{16 + 25 + 1} = \frac{4 + 25 + 4}{42} = \frac{33}{42} = \frac{11}{14}$$

So, the least squares estimator b_1 for these data points is $\frac{11}{14}$ or approximately 0.7857.

- the least squares estimator $b1$

$$\bar{x} = \frac{\Sigma x_i}{n} = \frac{x_1 + x_2 + x_3}{n} = \frac{0 + 4 + 5}{3} = 3$$

$$\bar{y} = \frac{\Sigma y_i}{n} = \frac{y_1 + y_2 + y_3}{n} = \frac{0 + 1 + 5}{3} = 2$$

$$b_1 = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2}$$

$$= \frac{(0-3)(0-2) + (4-3)(1-2) + (5-3)(5-2)}{(0-3)^2 + (4-3)^2 + (5-3)^2}$$

$$= \frac{6 - 1 + 6}{9 + 1 + 4} = \frac{11}{14}$$

numerically

$$b_1 = \frac{\sum_{i,j}(x_j - x_i)^2 \text{slope}_{ij}}{\sum_{i,j}(x_j - x_i)^2}.$$

$$b_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$b = \frac{\sum_{i,j}(x_j - x_i)^2 \text{slope}_{ij}}{\sum_{i,j}(x_j - x_i)^2} = \frac{\sum_{i,j}(x_j - x_i)^2 \frac{y_j - y_i}{x_j - x_i}}{\sum_{i,j}(x_j - x_i)^2} = \frac{\sum_{ij}(x_j - x_i)(y_j - y_i)}{\sum_{i,j}(x_j - x_i)^2}$$

$$y_i = b_0 + b_1 X_i + \epsilon_i$$

where b_0, b_1 are the estimator of β_0, β_1, ϵ_i is the residual.

So

$$b = \frac{\sum_{ij}(x_j - x_i)[b_1(x_j - x_i) + \epsilon_j - \epsilon_i]}{\sum_{i,j}(x_j - x_i)^2}$$

$$= \frac{\sum_{ij}(x_j - x_i)^2 \cdot b_1 + \sum_{ij}(x_j - x_i)(\epsilon_j - \epsilon_i)}{\sum_{ij}(x_j - x_i)^2}$$

$$= b_1 + \frac{\sum(x_j\epsilon_j - x_j\epsilon_i - x_i\epsilon_j + x_i\epsilon_i)}{\sum(x_j - x_i)}$$

The inner product of residual and predictor is 0.

So that, $\sum x_j\epsilon_j = \sum x_j\epsilon_i = \sum x_i\epsilon_j = \sum x_i\epsilon_i = 0$

Therefore, $b = b_1$.

$$b = \frac{\sum_{ij}(x_j - x_i)(y_j - y_i)}{\sum_{i,j}(x_j - x_i)^2}$$

$$\sum_{ij}(x_j - x_i)(y_j - y_i)$$

$$= \sum_{ij}(x_j - x_i - \bar{x} + \bar{x})(y_j - y_i - \bar{y} + \bar{y})$$

$$= \sum_{ij}[(x_j - \bar{x}) - (x_i - \bar{x})][(y_j - \bar{y}) - (y_i - \bar{y})]$$

$$= \sum_{ij}\{(x_j - \bar{x})(y_j - \bar{y}) - (x_j - \bar{x})(y_i - \bar{y}) - (x_i - \bar{x})(y_j - \bar{y}) + (x_i - \bar{x})(y_i - \bar{y})\}$$

$$= \sum_{j}[(x_j - \bar{x})(y_j - \bar{y})] - \cancel{(y_j - \bar{y})\sum_i(x_i - \bar{x})} - \cancel{(x_j - \bar{x})\sum_i(y_i - \bar{y})} + \sum(x_i - \bar{x})(y_i - \bar{y})$$

$$= 2\sum(x_i - \bar{x})(y_i - \bar{y})$$

$$\sum_{ij}(x_j - x_i)^2$$

$$= \sum_{ij}(x_j - x_i - \bar{x} + \bar{x})^2$$

$$= \sum_{ij}[(x_j - \bar{x}) - (x_i - \bar{x})]^2$$

$$= \sum_{ij}[(x_j - \bar{x})^2 + (x_i - \bar{x})^2 - 2(x_j - \bar{x})(x_i - \bar{x})]$$

$$= \sum_{ij}(x_i - \bar{x})^2 + \sum_{ij}(x_j - \bar{x})^2 \cancel{- 2(x_i - \bar{x})\sum_j(x_j - \bar{x})}$$

$$= 2\sum(x_i - \bar{x})^2$$

So that $b = \frac{2\sum(x_i - \bar{x})(y_i - \bar{y})}{2\sum(x_i - \bar{x})^2} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} = b_1$

$$b_1 = \frac{\sum_{i,j}(x_j - x_i)^2 \text{slope}_{ij}}{\sum_{i,j}(x_j - x_i)^2}.$$

$$b_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$