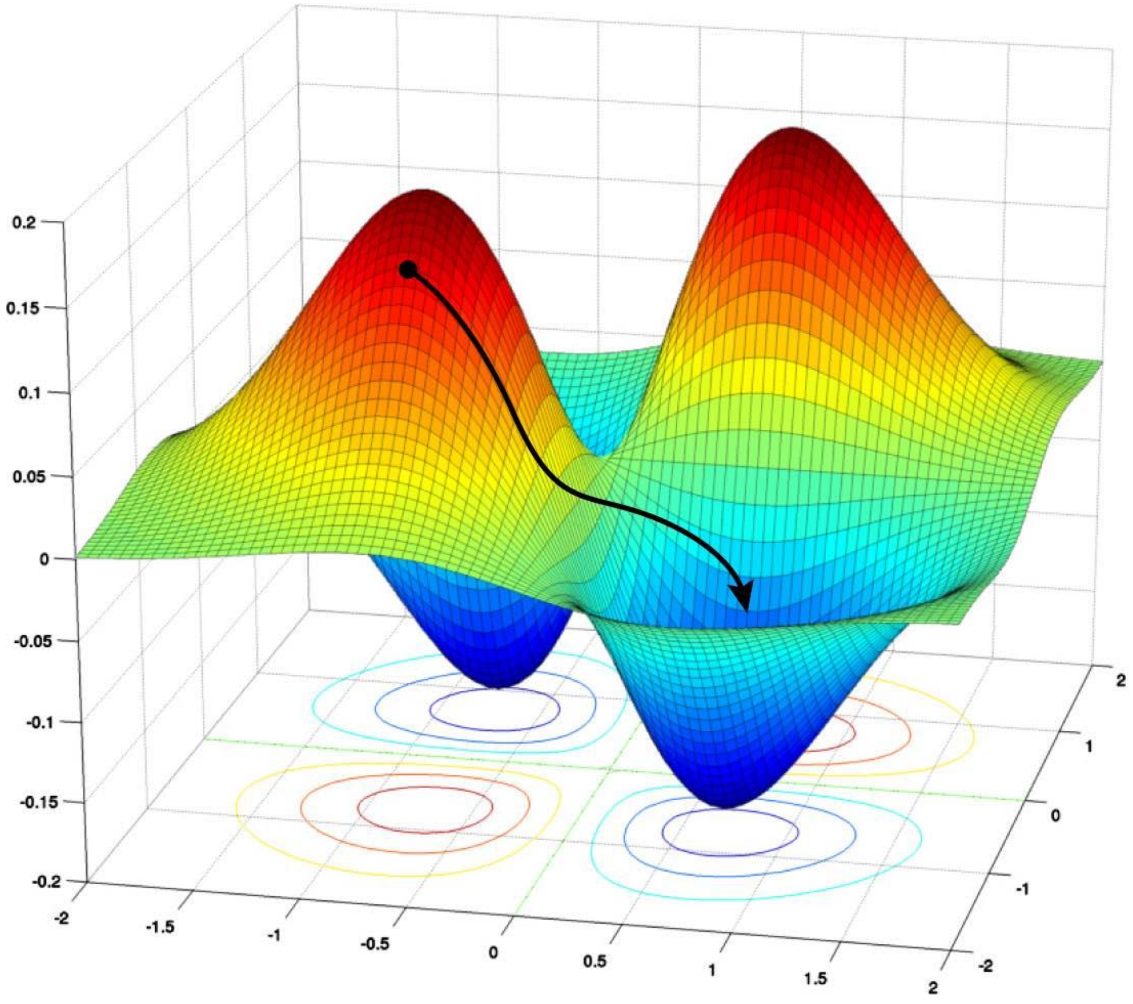
Gradient Descent



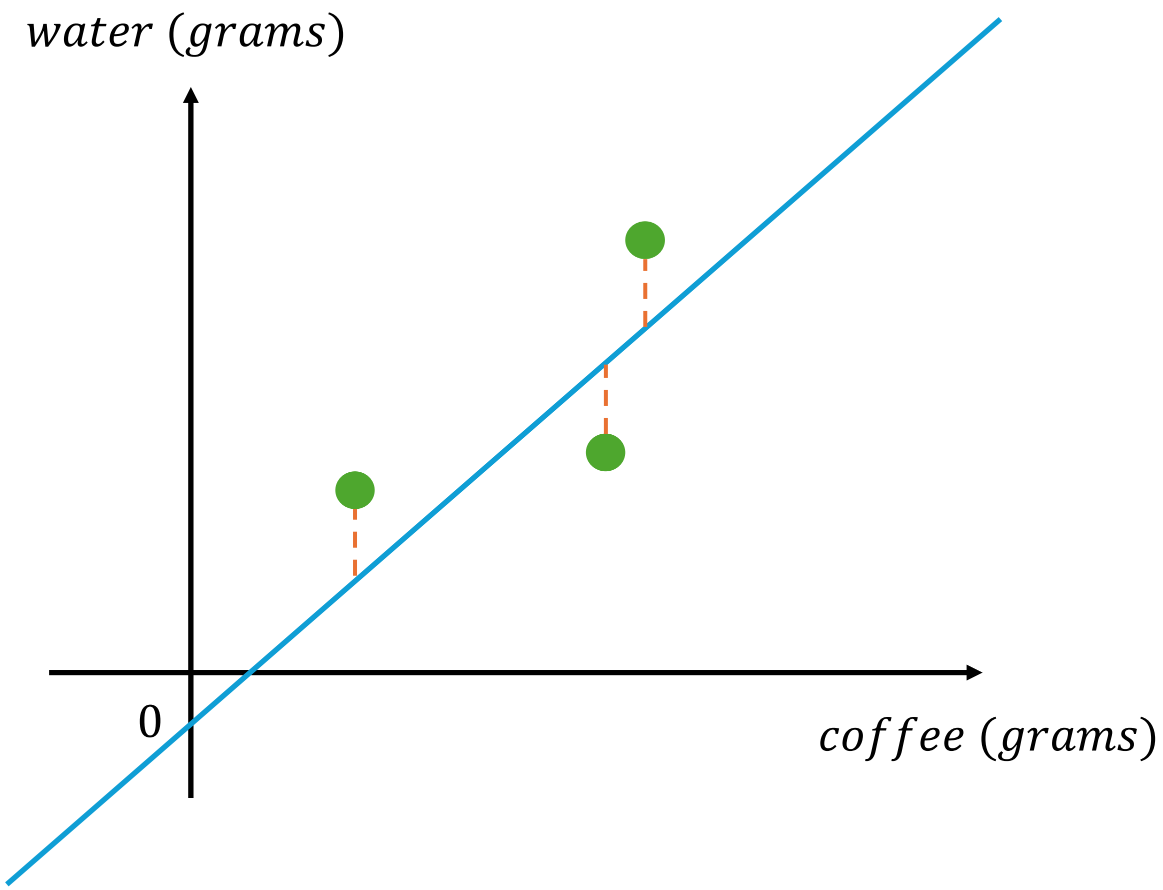
Coffee (grams)	Water (grams)
14	220
17	240
20	300

$$y = mx + c$$

$\boldsymbol{where}$

$\boldsymbol{m}$: $slope$

$\boldsymbol{c}$: $y - intercept$



$\boldsymbol{Linear\ Regression}$

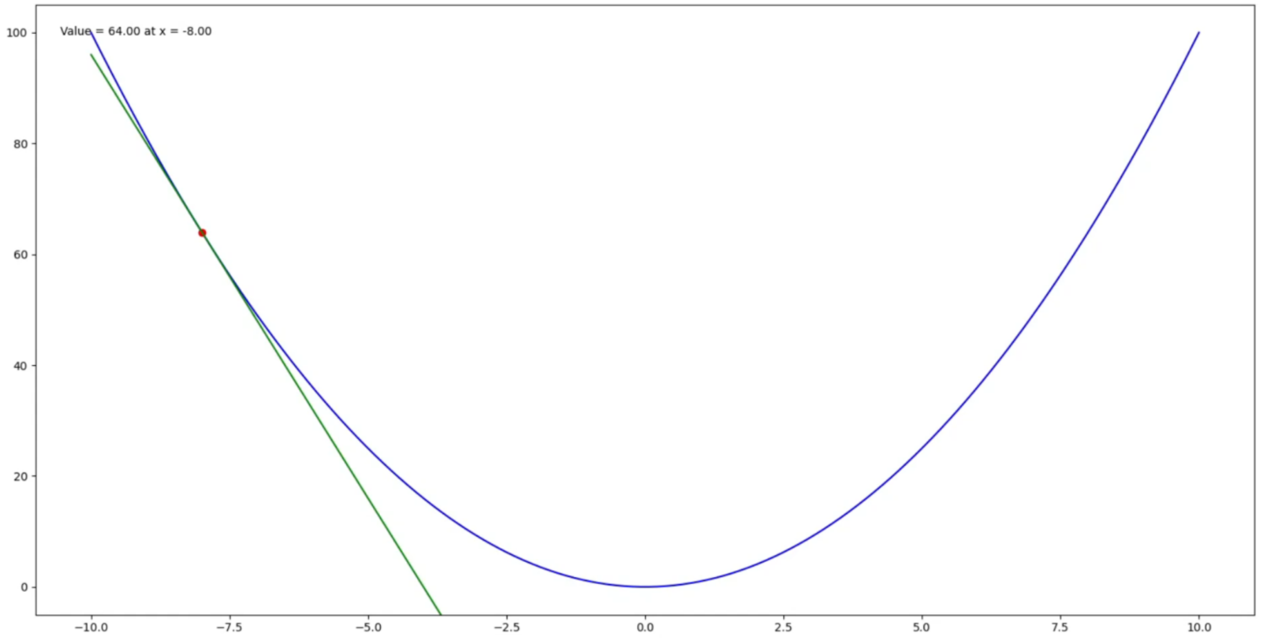
How to find

Slope and Y-Intercept ????

Using Gradient Descent to Optimize this.

Gradient Descent

$$\theta_{new} = \theta_{current} - \alpha \frac{\partial f}{\partial \theta}$$



Current Parameter

$$\theta_{new} = \theta_{current} - \alpha \frac{\partial \boxed{f}}{\partial \theta}$$

New Parameter

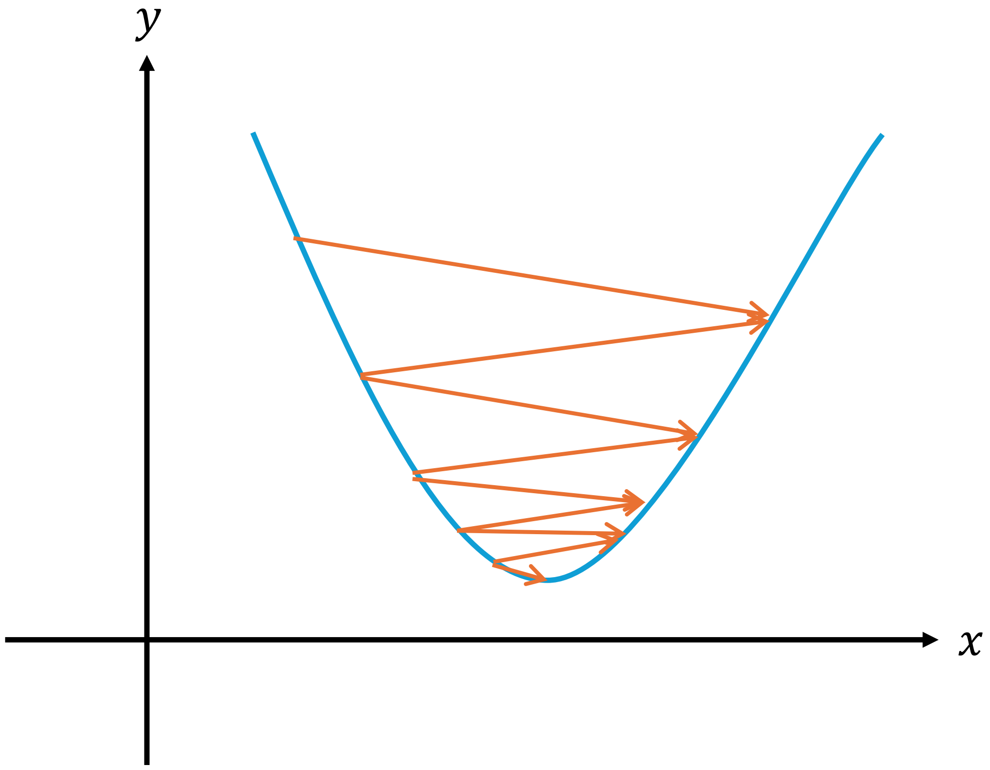
This is what we want!!!

Learning Rate

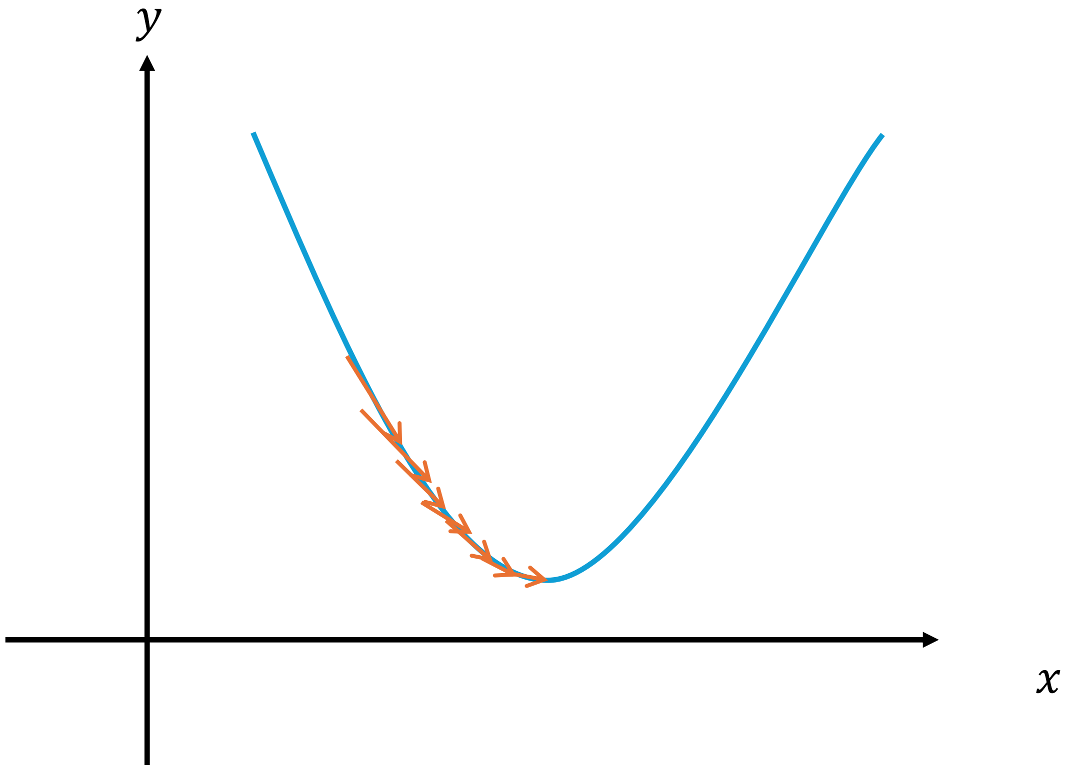
Interested Function

Learning Rate

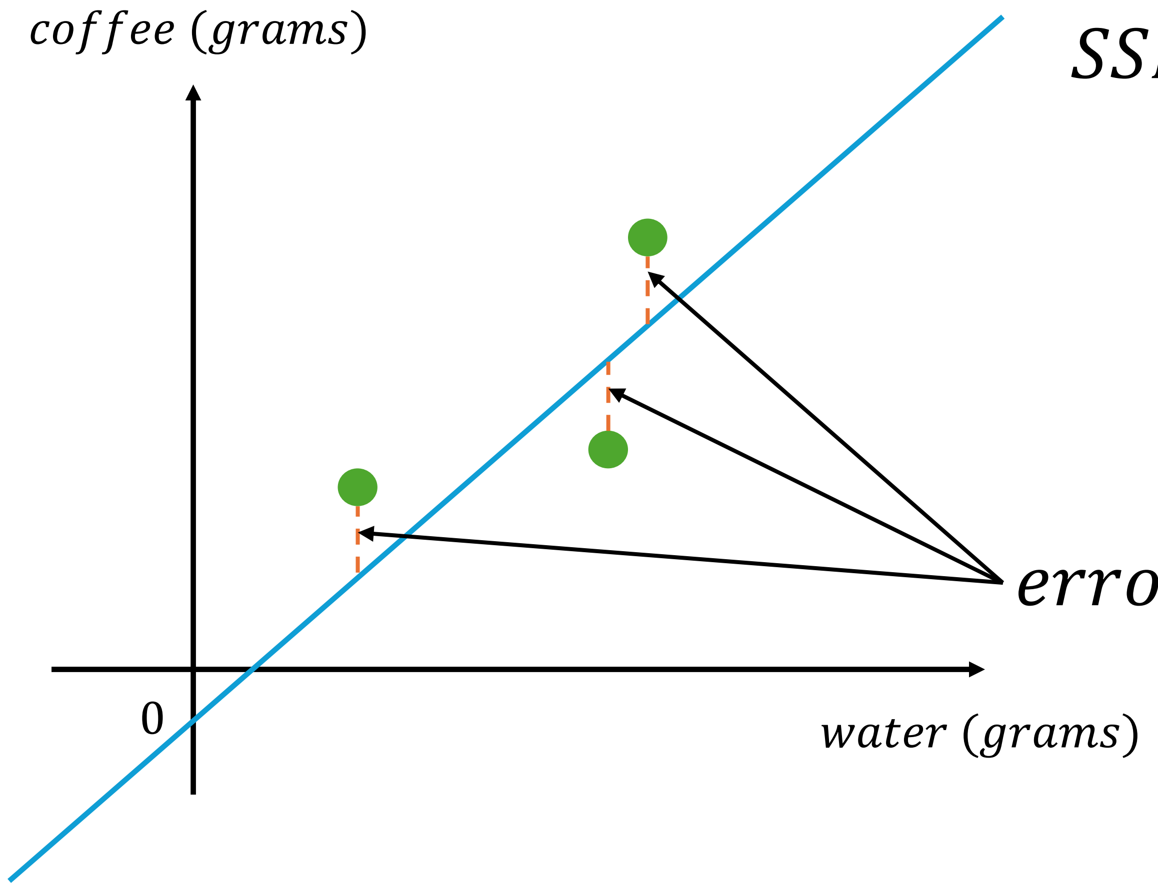
Huge Learning Rate



Small Learning Rate



$$SSE = \sum_{i=1}^{n} (\underbrace{y_i}_{y_{true}} - \underbrace{\hat{y}_i}_{y_{predicted}})^2$$



$$error = y_{true} - y_{predicted}$$

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$\theta_{new} = \theta_{current} - \alpha \frac{\partial \boxed{f}}{\partial \boxed{\theta}}$$

$$y = \boxed{m}x + \boxed{c}$$

Hey!, I didn't see this in SSE.

Chain Rule!!!

Slope

$$\frac{\partial SSE}{\partial m} = \frac{\partial SSE}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial m}$$

$$= \frac{\partial \sum_{i=1}^{n}(y-\hat{y})^2}{\partial \hat{y}} \cdot \frac{\partial(mx+\cancel{c})}{\partial m}$$

$$= \sum_{i=1}^{n} -2x_i(y_i - \hat{y}_i)$$

$$= -2x_1(y_1 - \hat{y}_1) - 2x_2(y_2 - \hat{y}_2) + \cdots - 2x_i(y_i - \hat{y}_i)$$

Y-Intercept

$$\frac{\partial SSE}{\partial c} = \frac{\partial SSE}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial c}$$

$$= \frac{\partial \sum_{i=1}^{n}(y-\hat{y})^2}{\partial \hat{y}} \cdot \frac{\partial(\cancel{mx}+c)}{\partial c}$$

$$= \sum_{i=1}^{n} -2(y_i - \hat{y}_i)$$

$$\theta_{new} = \theta_{current} - \alpha \frac{\partial f}{\partial \theta}$$

$$\begin{bmatrix} m \\ c \end{bmatrix}_{new} = \begin{bmatrix} m \\ c \end{bmatrix}_{current} - \alpha \begin{bmatrix} \sum_{i=1}^{n} \frac{\partial SSE}{\partial m} \\ \sum_{i=1}^{n} \frac{\partial SSE}{\partial c} \end{bmatrix}$$

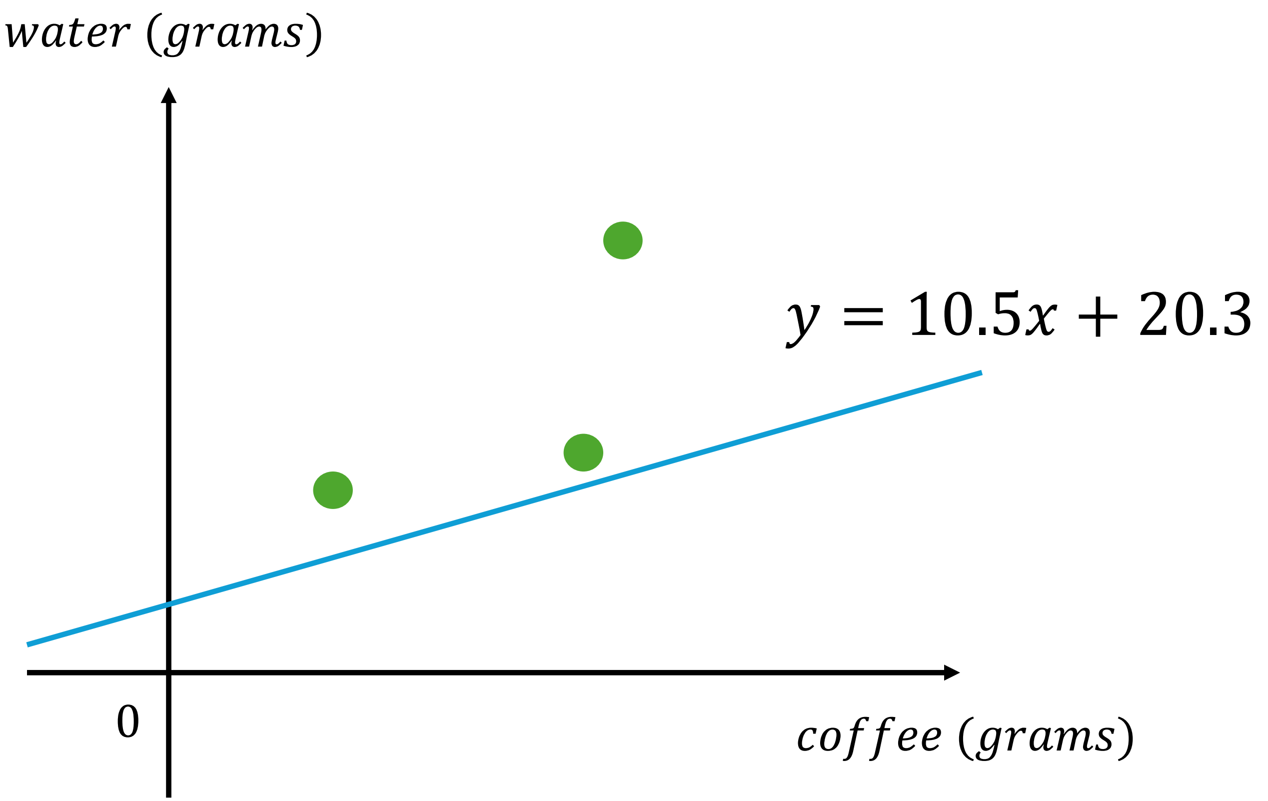
$$Define\!: \alpha = 0.001$$

Iteration 1:

Coffee (grams)	Water (grams)
14	220
17	240
20	300

$$\begin{bmatrix} m \\ c \end{bmatrix}_{new} = \begin{bmatrix} m \\ c \end{bmatrix}_{current} - \alpha \begin{bmatrix} \sum_{i=1}^{n} \frac{\partial SSE}{\partial m} \\ \sum_{i=1}^{n} \frac{\partial SSE}{\partial c} \end{bmatrix}$$

$$\begin{bmatrix} m \\ c \end{bmatrix}_{new} = \begin{bmatrix} 10.5 \\ 20.3 \end{bmatrix}_{current} - (0.001) \begin{bmatrix} \sum_{i=1}^{n} \frac{\partial SSE}{\partial m} \\ \sum_{i=1}^{n} \frac{\partial SSE}{\partial c} \end{bmatrix}$$



Iteration 1:

Coffee (grams)	Water (grams)
14	220
17	240
20	300

Slope

$$\sum_{i=1}^{n} \frac{\partial SSE}{\partial m}$$

$$= -2x_1(y_1 - \hat{y}_1) - 2x_2(y_2 - \hat{y}_2) - 2x_3(y_3 - \hat{y}_3)$$

$$= -2x_1(y_1 - (10.5x_1 + 20.3)) - 2x_2(y_2 - (10.5x_2 + 20.3)) - 2x_3(y_2 - (10.5x_3 + 20.3))$$

$$= -2(14)\big((220) - (10.5(14) + 20.3)\big) - 2(17)\big((240) - (10.5(17) + 20.3)\big) - 2(20)\big((300) - (10.5(20) + 20.3)\big)$$

$$= -5664.4\ \frac{grams}{grams}$$

Iteration 1:

Coffee (grams)	Water (grams)
14	220
17	240
20	300

Y-Intercept

$$\sum_{i=1}^{n} \frac{\partial SSE}{\partial c}$$

$$= -2(y_1 - \hat{y}_1) - 2(y_2 - \hat{y}_2) - 2(y_3 - \hat{y}_3)$$

$$= -2(y_1 - (10.5x_1 + 20.3)) - 2(y_2 - (10.5x_2 + 20.3)) - 2(y_2 - (10.5x_3 + 20.3))$$

$$= -2\big((220) - (10.5(14) + 20.3)\big) - 2\big((240) - (10.5(17) + 20.3)\big) - 2\big((300) - (10.5(20) + 20.3)\big)$$

$$= -327.2\ grams$$

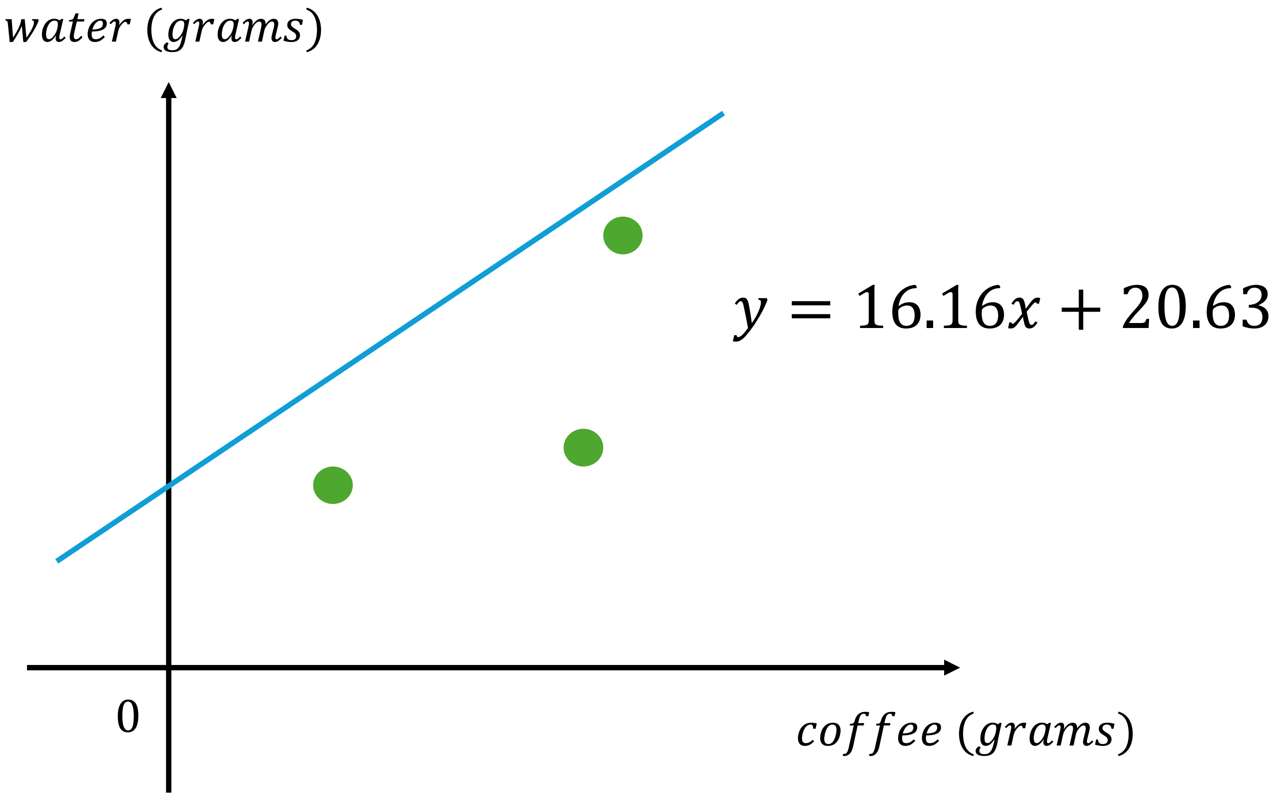
Iteration 1:

Coffee (grams)	Water (grams)
14	220
17	240
20	300

$$\begin{bmatrix} m \\ c \end{bmatrix}_{new} = \begin{bmatrix} 10.5 \\ 20.3 \end{bmatrix}_{current} - (0.001) \begin{bmatrix} \sum_{i=1}^{n} \frac{\partial SSE}{\partial m} \\ \sum_{i=1}^{n} \frac{\partial SSE}{\partial c} \end{bmatrix}$$

$$\begin{bmatrix} m \\ c \end{bmatrix}_{new} = \begin{bmatrix} 10.5 \\ 20.3 \end{bmatrix}_{current} - (0.001) \begin{bmatrix} -5664.4 \\ -327.2 \end{bmatrix}$$

$$\begin{bmatrix} m \\ c \end{bmatrix}_{new} = \begin{bmatrix} 16.16 \\ 20.63 \end{bmatrix}$$



Iteration 1:

Coffee (grams)	Water (grams)
14	220
17	240
20	300

$$y = 16.16x + 20.63$$

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$= (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2$$

$$= (y_1 - (16.16x_1 + 20.63))^2 + (y_2 - (16.16x_2 + 20.63))^2 + (y_3 - (16.16x_3 + 20.63))^2$$

$$= ((220) - (16.16(14) + 20.63))^2 + ((240) - (16.16(17) + 20.63))^2 + ((300) - (16.16(20) + 20.63))^2$$

$$= 5706.6883$$

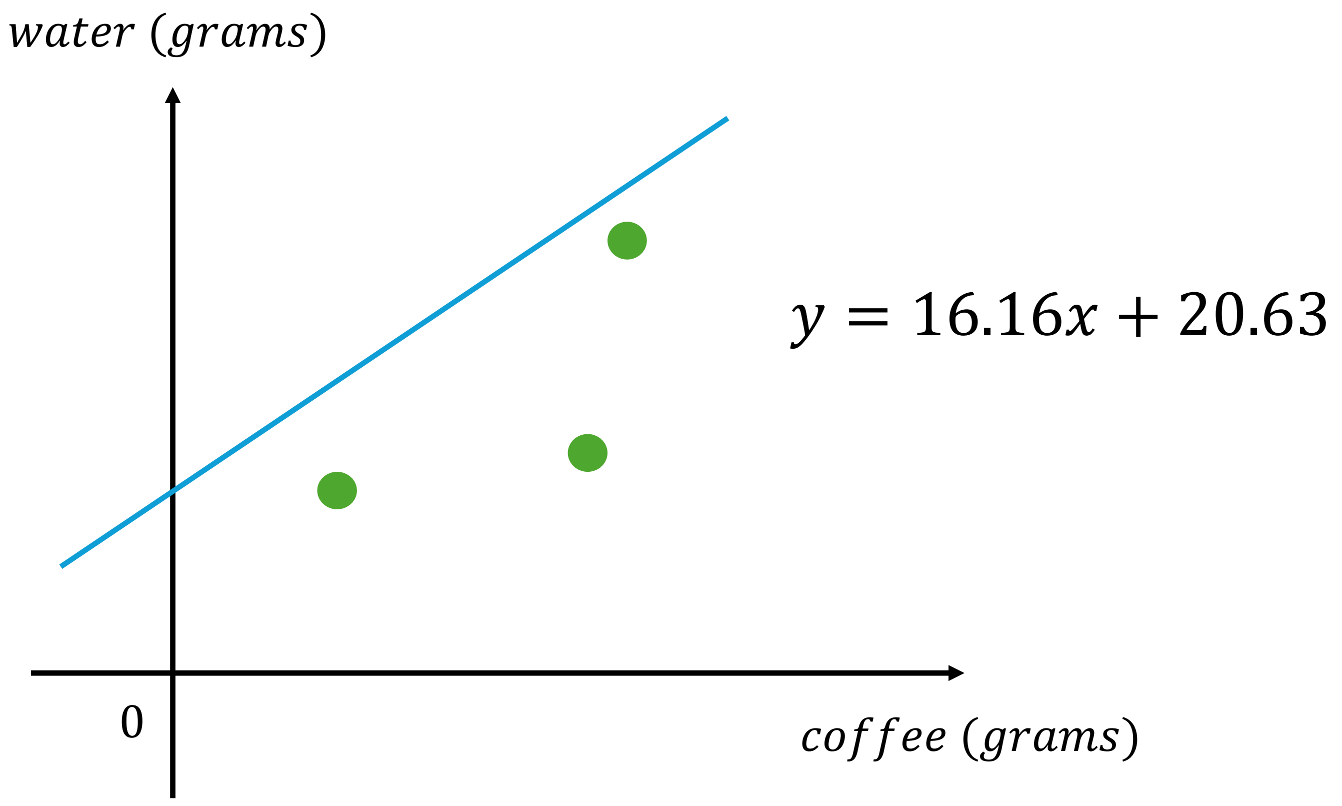
$$error = \sqrt{5706.6883} \approx 75.54$$

Iteration 2:

Coffee (grams)	Water (grams)
14	220
17	240
20	300

$$\begin{bmatrix} m \\ c \end{bmatrix}_{new} = \begin{bmatrix} m \\ c \end{bmatrix}_{current} - \alpha \begin{bmatrix} \sum_{i=1}^{n} \frac{\partial SSE}{\partial m} \\ \sum_{i=1}^{n} \frac{\partial SSE}{\partial c} \end{bmatrix}$$

$$\begin{bmatrix} m \\ c \end{bmatrix}_{new} = \begin{bmatrix} 16.16 \\ 20.63 \end{bmatrix}_{current} - (0.001) \begin{bmatrix} \sum_{i=1}^{n} \frac{\partial SSE}{\partial m} \\ \sum_{i=1}^{n} \frac{\partial SSE}{\partial c} \end{bmatrix}$$



Iteration 2:

Coffee (grams)	Water (grams)
14	220
17	240
20	300

Slope

$$\sum_{i=1}^{n} \frac{\partial SSE}{\partial m}$$

$$= -2x_1(y_1 - \hat{y}_1) - 2x_2(y_2 - \hat{y}_2) - 2x_3(y_3 - \hat{y}_3)$$

$$= -2x_1(y_1 - (16.16x_1 + 20.63)) - 2x_2(y_2 - (16.16x_2 + 20.63)) - 2x_3(y_2 - (16.16x_3 + 20.63))$$

$$= -2(14)\big((220) - (16.16(14) + 20.63)\big) - 2(17)\big((240) - (16.16(17) + 20.63)\big) - 2(20)\big((300) - (16.16(20) + 20.63)\big)$$

$$= 4387.46\ \frac{grams}{grams}$$

Iteration 2:

Coffee (grams)	Water (grams)
14	220
17	240
20	300

Y-Intercept

$$\sum_{i=1}^{n} \frac{\partial SSE}{\partial c}$$

$$= -2(y_1 - \hat{y}_1) - 2(y_2 - \hat{y}_2) - 2(y_3 - \hat{y}_3)$$

$$= -2(y_1 - (16.16x_1 + 20.63)) - 2(y_2 - (16.16x_2 + 20.63)) - 2(y_2 - (16.16x_3 + 20.63))$$

$$= -2\big((220) - (16.16(14) + 20.63)\big) - 2\big((240) - (16.16(17) + 20.63)\big) - 2\big((300) - (16.16(20) + 20.63)\big)$$

$$= 252.1\ grams$$

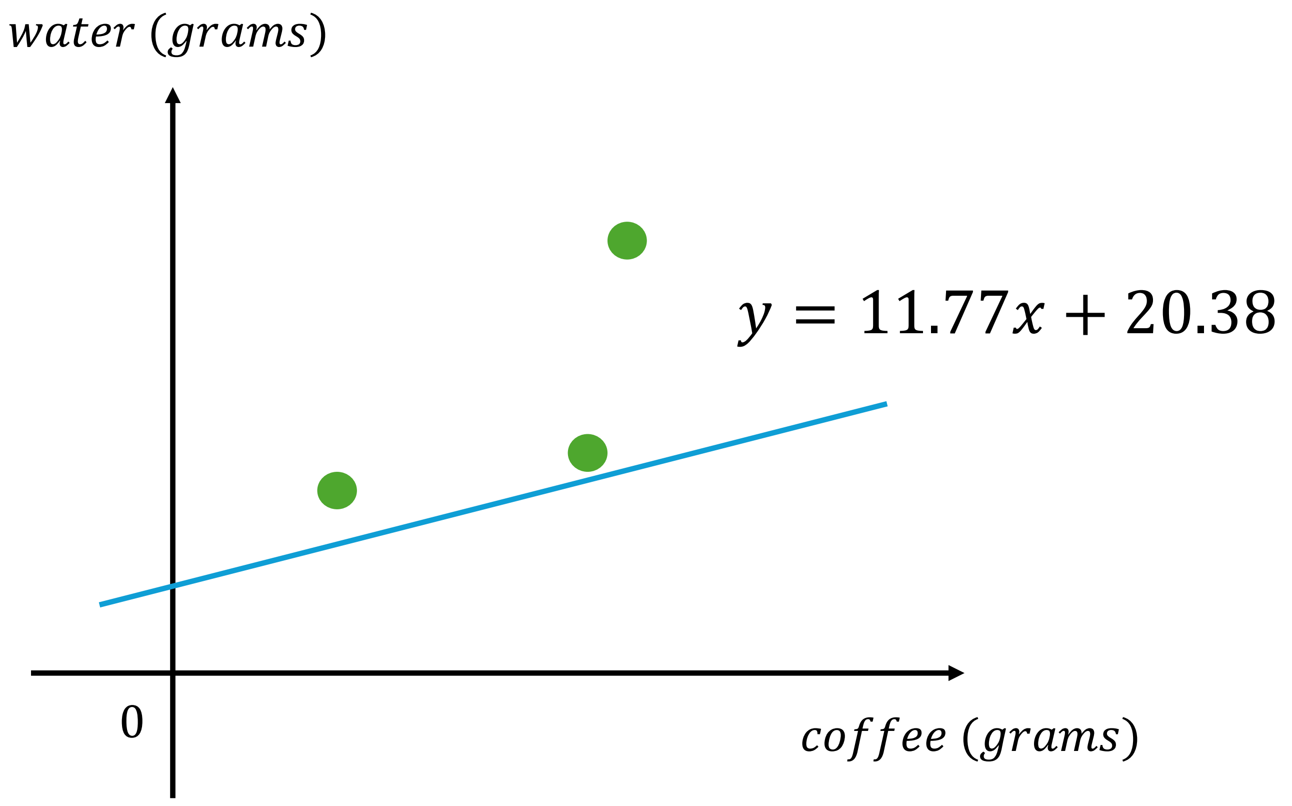
Iteration 2:

Coffee (grams)	Water (grams)
14	220
17	240
20	300

$$\begin{bmatrix} m \\ c \end{bmatrix}_{new} = \begin{bmatrix} 16.16 \\ 20.63 \end{bmatrix}_{current} - (0.001) \begin{bmatrix} \sum_{i=1}^{n} \frac{\partial SSE}{\partial m} \\ \sum_{i=1}^{n} \frac{\partial SSE}{\partial c} \end{bmatrix}$$

$$\begin{bmatrix} m \\ c \end{bmatrix}_{new} = \begin{bmatrix} 16.16 \\ 20.63 \end{bmatrix}_{current} - (0.001) \begin{bmatrix} 4387.46 \\ 252.1 \end{bmatrix}$$

$$\begin{bmatrix} m \\ c \end{bmatrix}_{new} = \begin{bmatrix} 11.77 \\ 20.38 \end{bmatrix}$$



Iteration 2:

Coffee (grams)	Water (grams)
14	220
17	240
20	300

$$y = 16.16x + 20.63$$

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$= (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + (y_3 - \hat{y}_3)^2$$

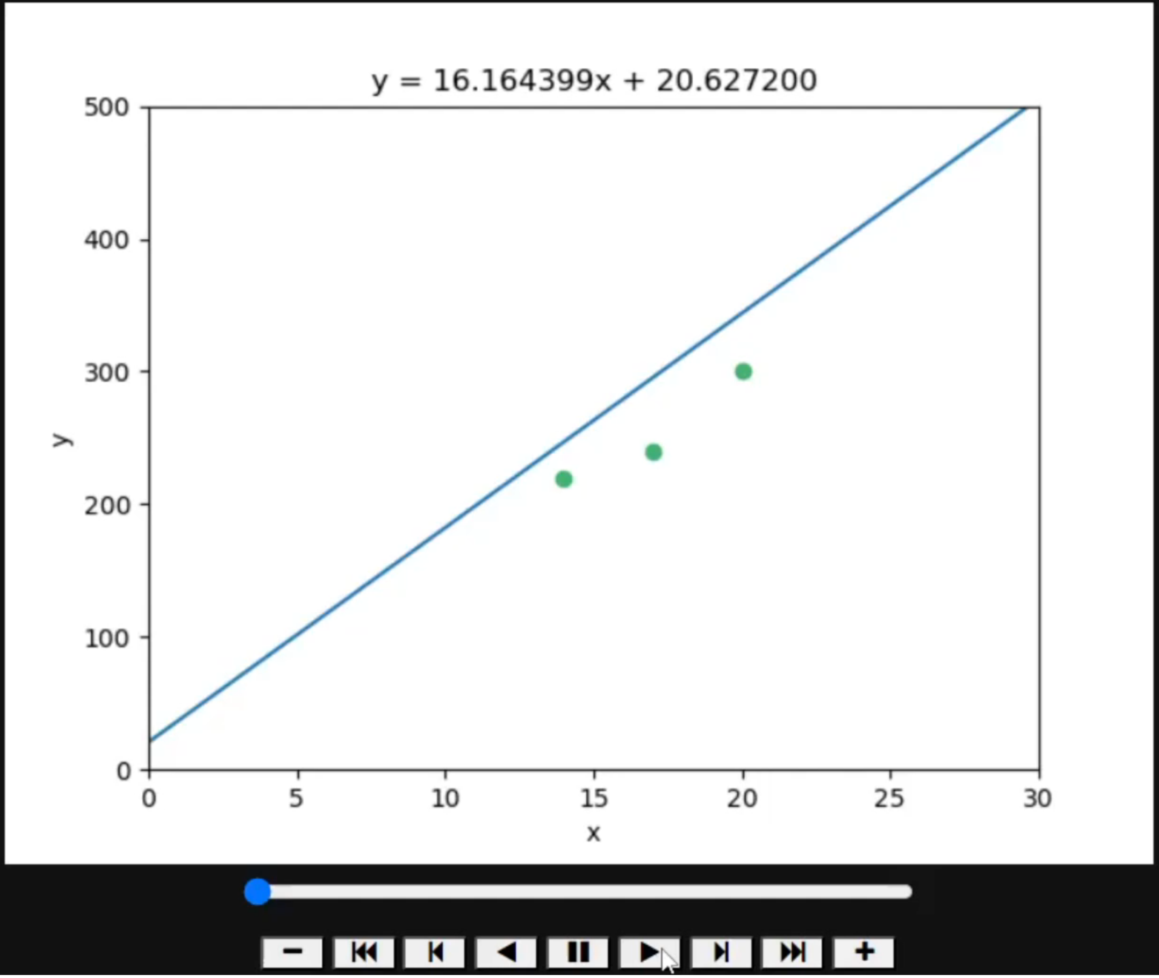
$$= (y_1 - (11.77x_1 + 20.38))^2 + (y_2 - (11.77x_2 + 20.38))^2 + (y_3 - (11.77x_3 + 20.38))^2$$

$$= ((220) - (11.77(14) + 20.38))^2 + ((240) - (11.77(17) + 20.38))^2 + ((300) - (11.77(20) + 20.38))^2$$

$$= 3550.65$$

$$error = \sqrt{3550.65} \approx 59.59$$

```
HTML(ani.to_jshtml())
```



Stochastic Gradient Descent

$$\begin{bmatrix} m \\ c \end{bmatrix}_{new} = \begin{bmatrix} m \\ c \end{bmatrix}_{current} - \alpha \begin{bmatrix} \cancel{\sum_{i=1}^{n}} \frac{\partial SSE}{\partial m} \\ \cancel{\sum_{i=1}^{n}} \frac{\partial SSE}{\partial c} \end{bmatrix}$$

Change from all train data to just random as a batch or just 1 sample/optimize

Taylor Series Approximation

HISTORICAL BIOGRAPHIES

Brook Taylor
(1685–1731)

Colin Maclaurin
(1698–1746)

DEFINITIONS Let f be a function with derivatives of all orders throughout some interval containing a as an interior point. Then the **Taylor series generated by f at $x = a$** is

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \cdots.$$

The **Maclaurin series generated by f** is

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!}x^k = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \cdots + \frac{f^{(n)}(0)}{n!}x^n + \cdots,$$

the Taylor series generated by f at $x = 0$.

Thomas' Calculus Early Transcendentals Instructor's Edition 12th Ed.

Derivative Definition

$$\frac{df(x)}{dx} = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

$$f'(x) = \frac{f(x) - f(a)}{(x-a)} = \frac{f(x) - f(x_{i+1})}{(x - x_{i+1})} = \frac{f(x) - f(x + \Delta x)}{(x - (x + \Delta x))}$$

$$\Delta x = 0.000001$$

$$f(x) = x^2$$

$$f'(x) = 2x$$

$$f'(2) = 2(2) = 4$$

Using Taylor Series Approximation

$$f'(x) = \frac{f(x) - f(x + \Delta x)}{(x - (x + \Delta x))}$$

$$f'(2) = \frac{f(2) - f(2 + 0.000001)}{(2 - (2 + 0.000001))} = 4.000001$$