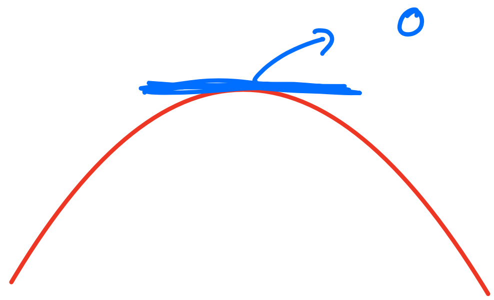
Towards Gradient Descent

$\vec{w}, w_0$



Recap

$$\max_{\vec{w}, w_0} \sum_{i=1}^{n} y_i \cdot \frac{w^T x_i + w_0}{||w||}$$

Gain Func ✗

Loss Function ✓

$$\boxed{\min_{\vec{w}, w_0} - \sum_{i=1}^{n} y_i \cdot \frac{w^T x_i + w_0}{||w||}}$$

$20, 50 \Rightarrow 50$

$-20, -50 \Rightarrow -50$

$$\min_{\vec{w}_1, \vec{w}_2, \vec{w}_3, \ldots} \quad f(w_1, w_2, \ldots, w_n)$$

$f(x) \quad 2x^2 + 4x + 3$

$f'(x) \quad 4x + 4 \quad \leftarrow$

$$y = f(x_1, x_2, \ldots, x_n)$$

$$\frac{d f(x_1, x_2, \ldots, x_n)}{dx} = ??$$

x_1	x_2	x_3		$\Rightarrow y$
Age	weight	height		salary

weight = 2 Age + 3 height

Ex

x_1, x_2

$x_2 = g(x_1)$

$$z = f(x_1, x_2)$$

$= f(x_1, g(x_1))$

Ex

$$z = f(x,y) = x^2 + y^2 \qquad y = g(x)$$

$$z = x^2 + [g(x)]^2$$

$$\frac{dz}{dx} = 2x + 2g(x) \cdot g'(x)$$

Partial Derivatives

$x_1, x_2, x_3 \rightarrow y$

$\Downarrow$

$x_4 \Rightarrow x_1 / x_2$

BMI

$$\frac{dz}{dz} \qquad \frac{\partial z}{\partial z}$$

$$z = f(x_1, x_2)$$

$$\frac{\partial z}{\partial x_1} = f'(x_1, x_2)$$

constant

$$f(x,y) = 2\underline{x}^2\underline{y} + \underline{3y}^3x^2 + \cancel{3y} \rightarrow 0$$

$$\frac{\partial f(x,y)}{\partial x} = 4xy + 6y^3x$$

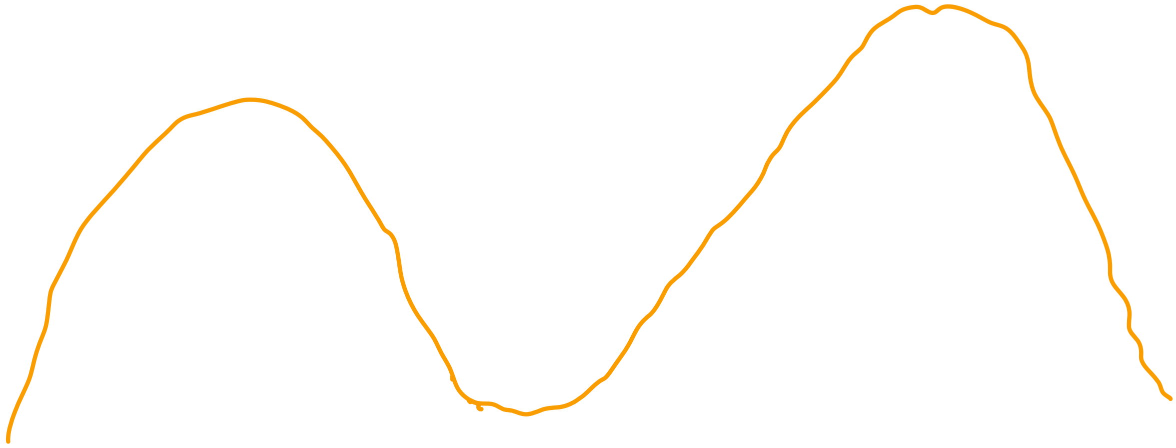
$$\frac{\partial f(x,y)}{\partial y} = 2x^2 + 9y^2x^2 + 3$$

Goal:

$$\min_{w_0, w_1, \dots, w_n} f(w_0, w_1, \dots, w_n)$$

$$\frac{\partial f}{\partial w_0}, \frac{\partial f}{\partial w_1}, \frac{\partial f}{\partial w_2}, \dots \frac{\partial f}{\partial w_n}$$

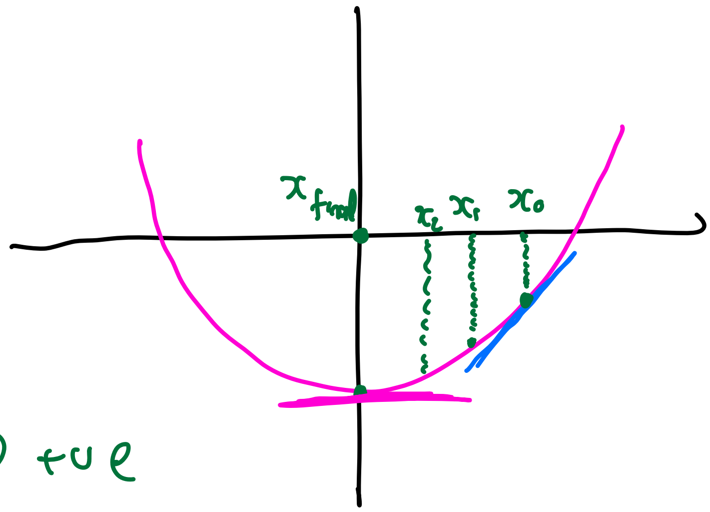
$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial w_0} \\ \frac{\partial f}{\partial w_1} \\ \vdots \\ \frac{\partial f}{\partial w_n} \end{bmatrix} \rightarrow \text{Gradient}$$



Gradient Descent Intution

$$y = x^2 - 30$$

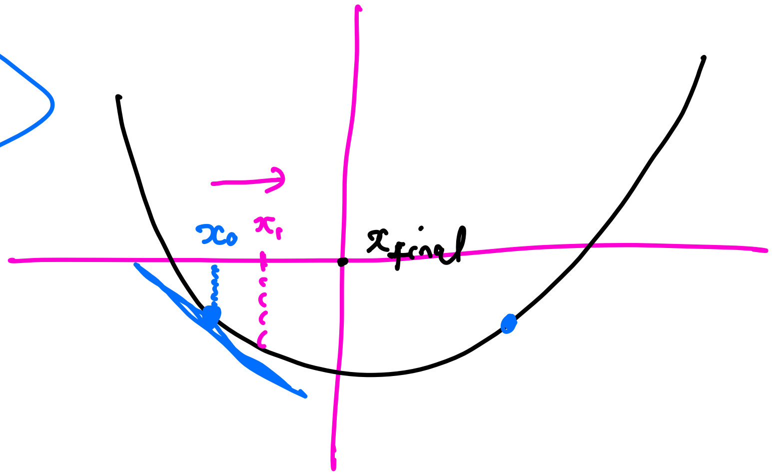
$$\frac{\partial y}{dx} = 2x$$



$$f'(x_0) \Rightarrow +ve$$

$$x_1 = x_0 - \underbrace{\eta}_{\text{learning rate}} f'(x_0)$$

$x_0 - f'(x)$



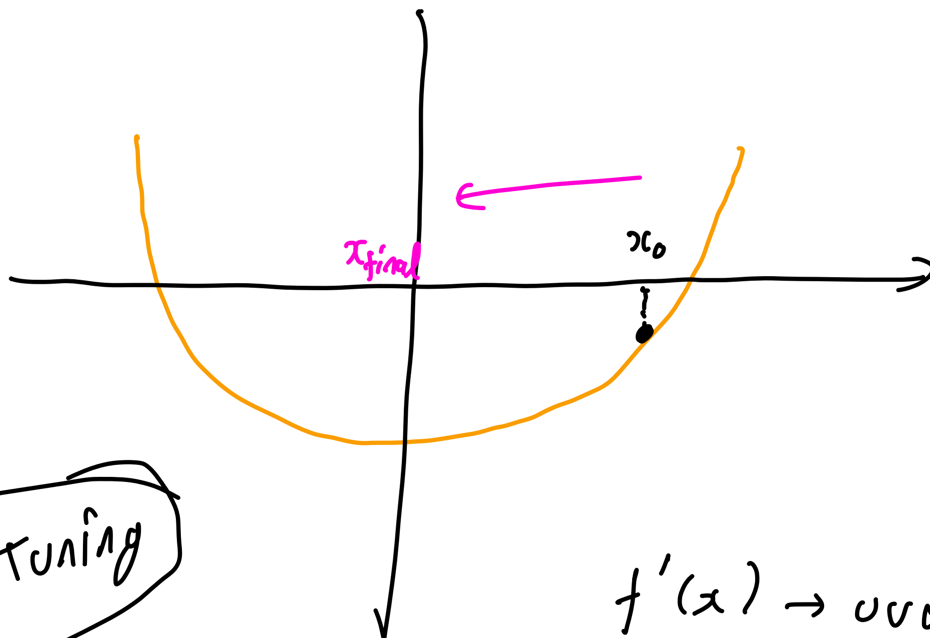
$f'(x) \Rightarrow -ve$

$x_0 - (-ve)$

$x_0 + ve$

$x_0 - \eta f'(x)$

$2 - 6$

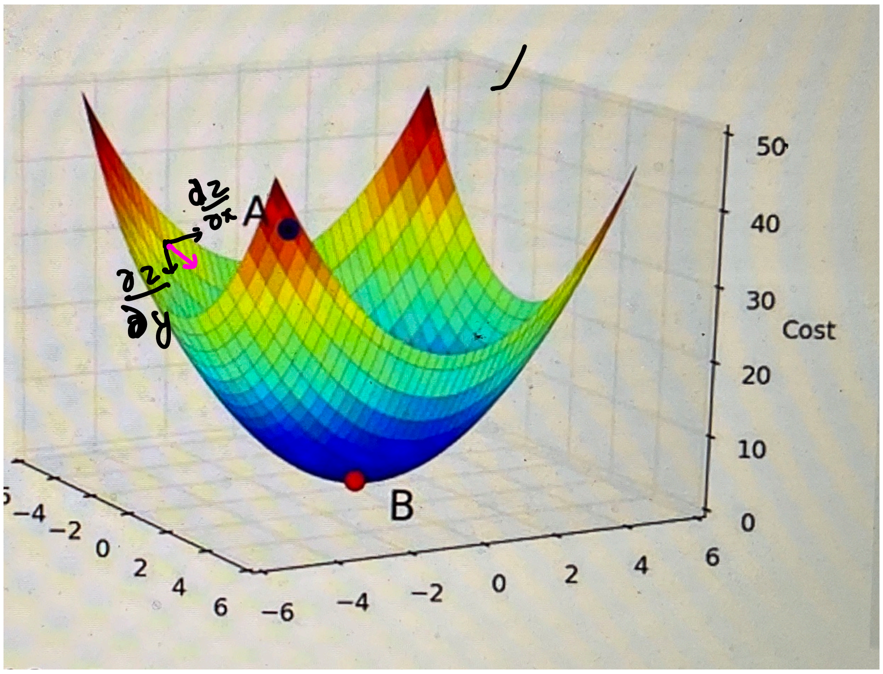


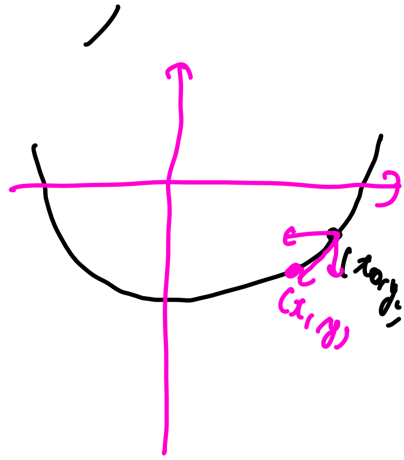
Hyperparameter Tuning

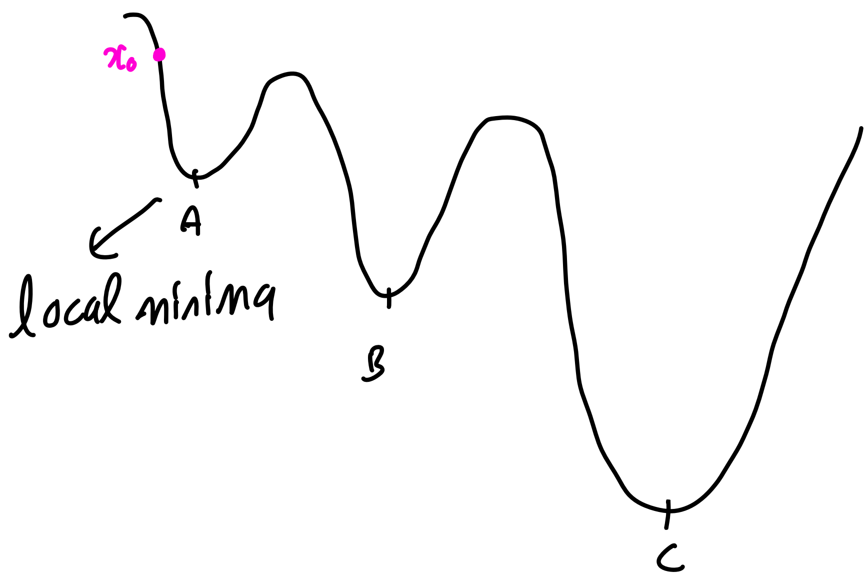
$f'(x) \to$ very small

$\eta \to 1$ or 2

Break : 8 : 05 AM



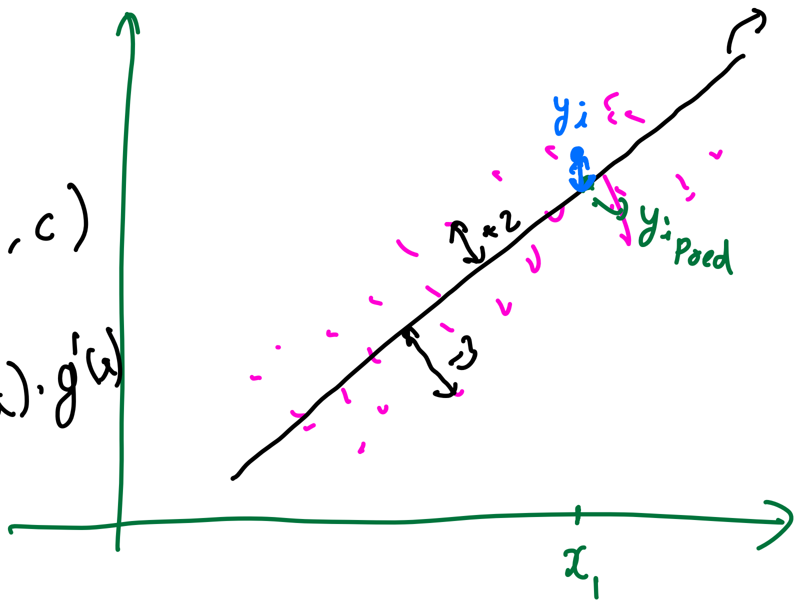




$$\hat{y}_i = mx_i + c$$

$$\sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2 = J(m, c)$$

$$g(x)^2 \Rightarrow 2g(x) \cdot g'(x)$$



$$\sum_{i=1}^{n} \left(y_i - (mx_i + c) \right)^2 = J(m, c)$$

$$\frac{\partial J}{\partial m} = \frac{2}{N} \sum \left(y_i - (mx_i + c) \right) \cdot (-x_i)$$

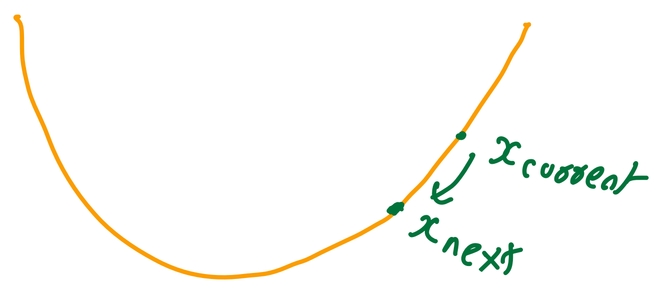
$$\frac{\partial J}{\partial c} = \frac{2}{N} \sum \left(y_i - (m x_i + c)\right) \cdot (-1)$$

$$\frac{\partial\, g(x)}{\partial\, m} \Rightarrow y_i - (m x_i + c)$$

$$\Downarrow$$

$$\frac{\partial\, m x_i}{\partial m} - \frac{\partial c}{\partial m}$$

$\Rightarrow$



$$w_1[next] = w_1[current] - \eta \frac{\partial f}{\partial w_1}$$

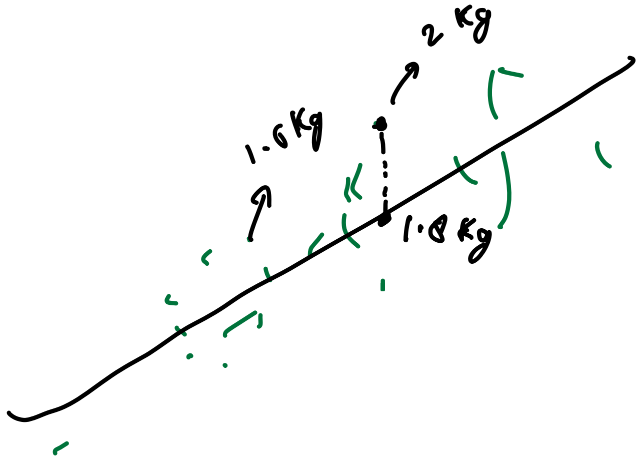
$$\vdots$$

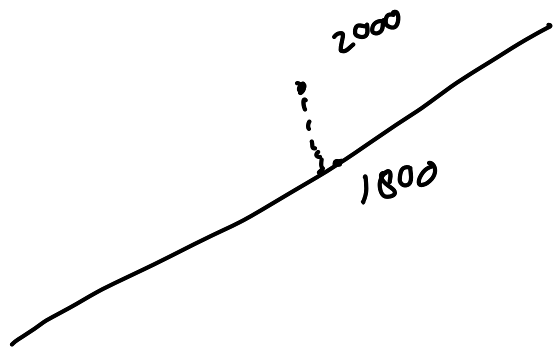
$$w_n[next] = w_n[curr] - \eta \frac{\partial f}{\partial w_n}$$

$(.2)^2$

$(200)^2$

Z-score





$w_1 x_1 + w_2 x_2 = y$

```python
# returns gradient at current m and c for each pair of m and c

def gradient(X, y, m_current=0, c_current=0, iters=1000, learning_rate=0
    N = float(len(y))
    gd_df = pd.DataFrame( columns = ['m_current', 'c_current','cost'])
    for i in range(iters):
        y_current = (m_current * X) + c_current
        cost = sum([data**2 for data in (y-y_current)]) / N
        m_gradient = -(2/N) * sum(X * (y - y_current))
        c_gradient = -(2/N) * sum(y - y_current)
        m_current = m_current - (learning_rate * m_gradient)
        c_current = c_current - (learning_rate * c_gradient)
        gd_df.loc[i] = [m_current,c_current,cost]
    return(gd_df)
```

$$\sum_{i=1}^{n} \left(y_i - (mx_i + c)\right)^2 = J(m, c)$$

$$\frac{\partial J}{\partial m} = \frac{2}{N} \sum \left(y_i - (mx_i + c)\right) \cdot (-x_i)$$