Opera File Edit View History Bookmarks Developer Window Help

Gradient Descent Algorithm

— Find the model parameters that give minimum loss

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \quad + \beta_m x_m$$

$$\hat{y} = w_1 x_1 + w_2 x_2 + w_3 x_3 \quad + w_m x_m + b$$

x_1	x_2	x_m	y
1			
2			
3			

$$\text{Loss Fn} = fn(\text{error})$$

$$= (y - \hat{y})^2$$

$$\text{Loss} = \left[y - (\underline{w_1} x_1 + \underline{w_2} x_2 \quad + w_m x_m + b)\right]^2$$

$$\hat{y} = mx + c$$

$$\hat{y} = b + wx$$

$\hookrightarrow 0$

Task → Find the weights s.t the loss is minimum

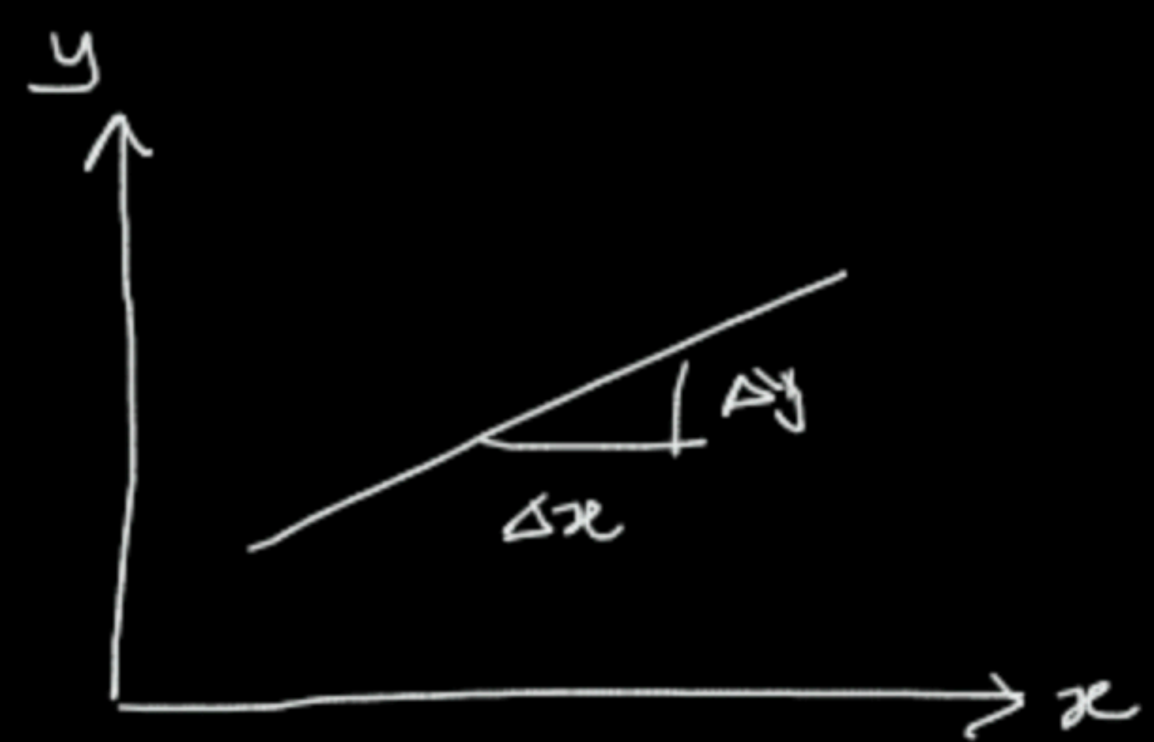
$$\hat{y} = \underline{w}x$$

$$(y - \hat{y})^2 = (y - wx)^2$$

$$\underset{(L)}{\text{Loss}} = (y - \underline{w}x)^2$$

$$\text{Gradient} \rightarrow \frac{dy}{dx} \rightarrow \frac{\Delta y}{\Delta x}$$

1/2



$\frac{dy}{dx} \rightarrow$ Slope

$y = x^3$

$\frac{dy}{dx} = 3x^2$

$y = 3x^2$

$\frac{dy}{dx} = 3(2x)$

$= 6x$

$y = 3$

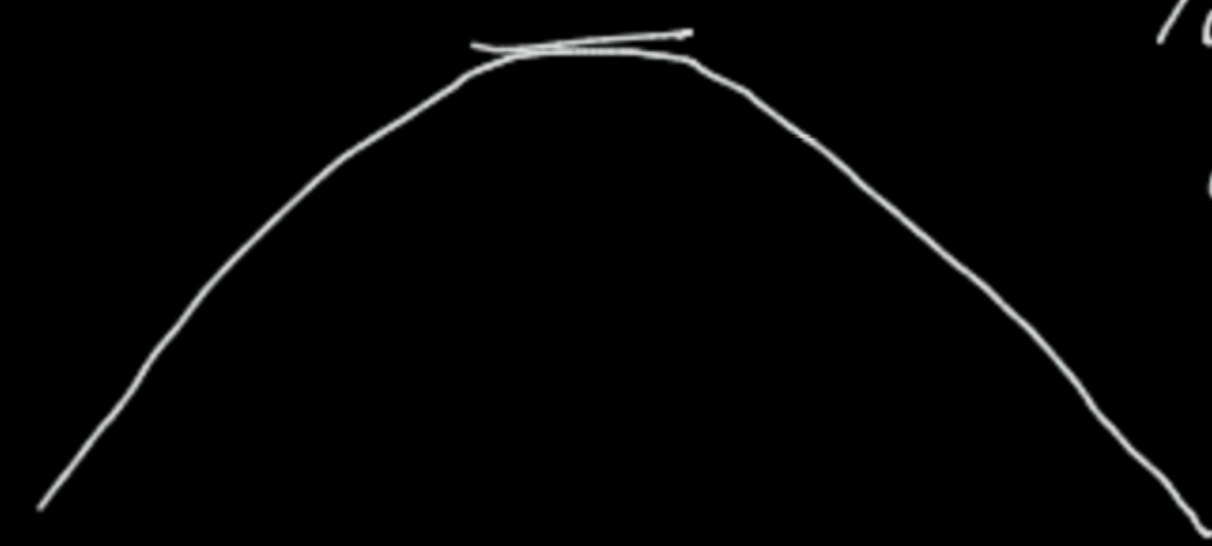
Slope $= 0$ $\quad \frac{dy}{dx} = 0$

minima

$\frac{d^2y}{dx^2} = +ve$

Maxima

slope $= 0$ $\quad \frac{dy}{dx} = 0$



$\frac{d^2y}{dx^2} = -ve$

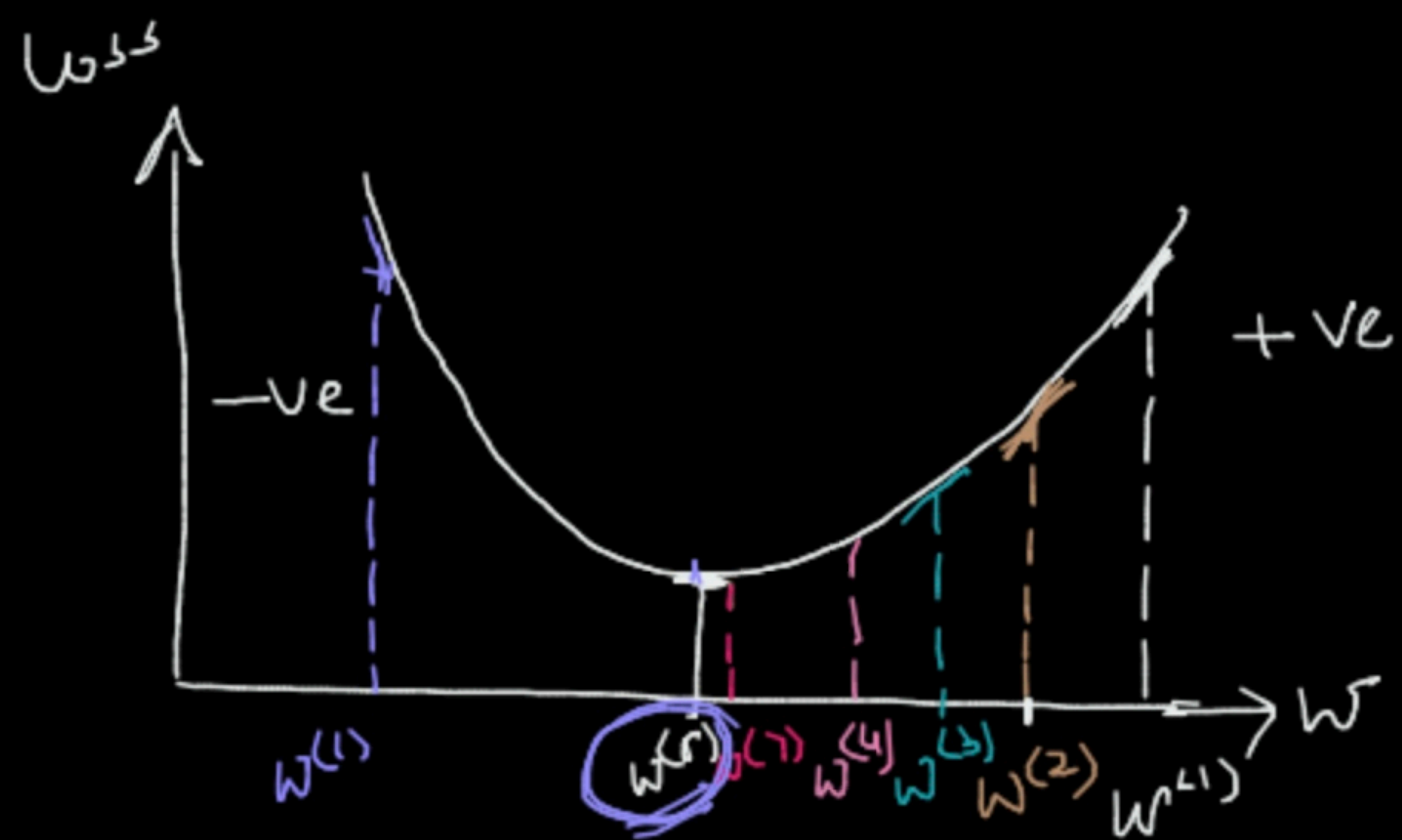
$y = x^2$

$\rightarrow$ parabola

$\hat{y} = wx$

loss $= [y - wx]^2$
(L)

$\frac{dL}{dW} \rightarrow$



Step 1 → Randomly Choose a wt

Step 2 → Find the gradient of the loss fn at w^{old}

$$\left[\frac{\partial L}{\partial w}\right]_{w^{(1)}}$$

Step 3 Weight update

$$w^{(2)} = w^{(1)} + \overset{+ve}{-}\left(\frac{\partial L}{\partial w}\right)_{w^{(1)}}$$

$$w^{new} = w^{old} + -\left(\frac{\partial L}{\partial w}\right)_{w^{old}}$$

Step 4 Repeat Steps 2 & 3 until Convergence

↳ $w^{new} = w^{old}$

$$\left[\frac{\partial L}{\partial w}\right]_{w^{(2)}} \qquad w^{(3)} = w^{(2)} + \left[-\frac{\partial L}{\partial w}\right]_{w^{(2)}}$$

$$\left[\frac{\partial L}{\partial w}\right]_{w^{(3)}} \qquad w^{(4)} = w^{(3)} + \overset{+ve}{\left[-\frac{\partial L}{\partial w}\right]}_{w^{(3)}}$$

$$w^{(8)} = w^{(7)} + \left[-\frac{\partial L}{\partial w}\right]_{w^{(7)}}$$

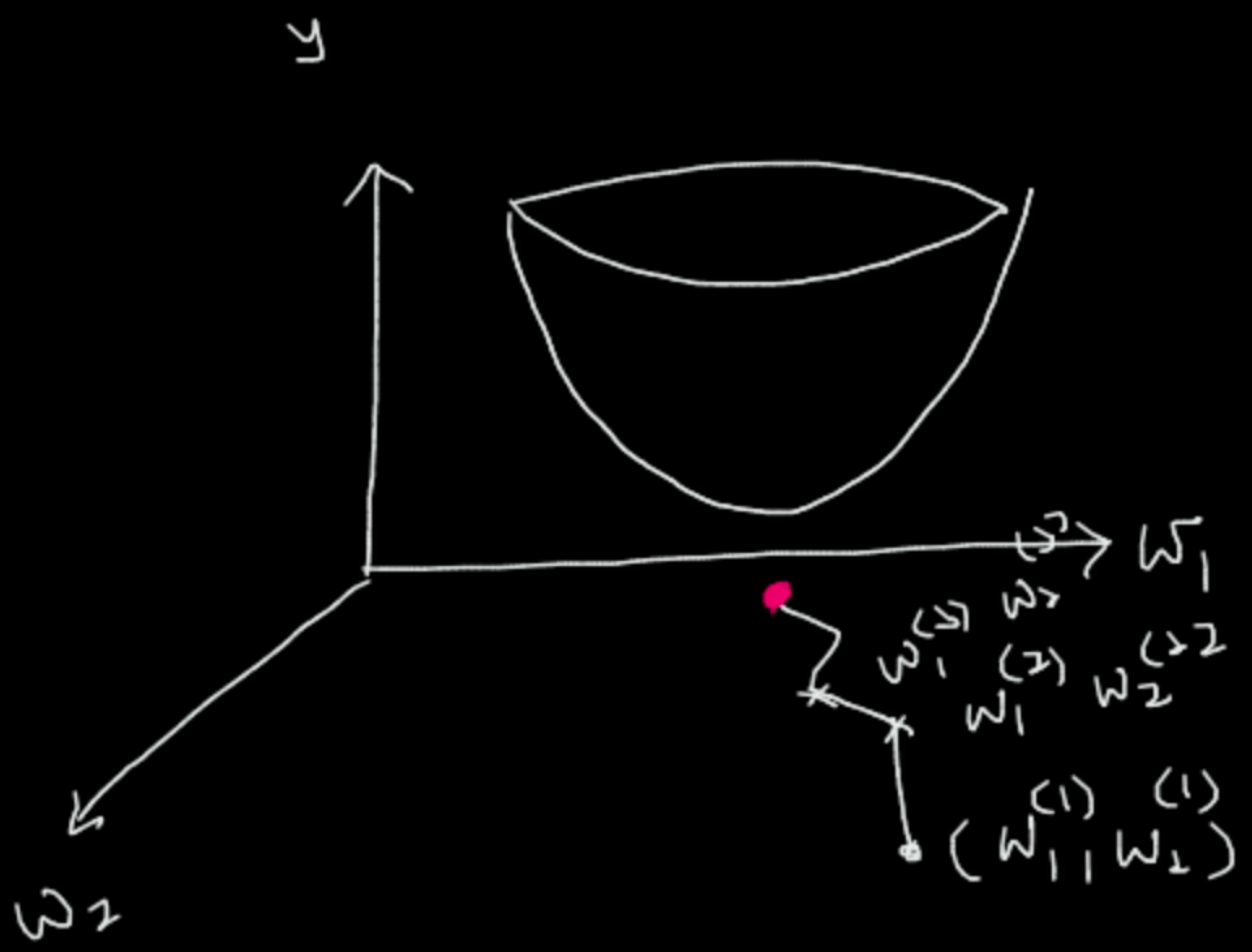
$$w^{(2)} = w^{(1)} + \overset{-ve}{\left[-\frac{\partial L}{\partial w}\right]}_{w^{(1)}} \qquad w^{(9)} = w^{(8)} + \left[-\frac{\partial L}{\partial w}\right]_{w^{(8)}}$$

$$w^{(2)} = w^{(1)} + \overset{+v}{-}$$

$$\hat{y} = \underline{w_1} x_1 + \underline{w_2} x_2$$

$$\text{Loss} = (y - \hat{y})^2$$

$$= \left[y - (\underline{w_1} x_1 + \underline{w_2} x_2) \right]^2$$



Step 1 Randomly choose 'm' weights

Step 2 Find the gradients of the loss w r t all the weights

$$\left[\frac{\partial L}{\partial w_1} \right]_{w_1^{(1)}} \quad \left[\frac{\partial L}{\partial w_2} \right]_{w_2^{(1)}}$$

Step 3 Update all weights simultaneously.

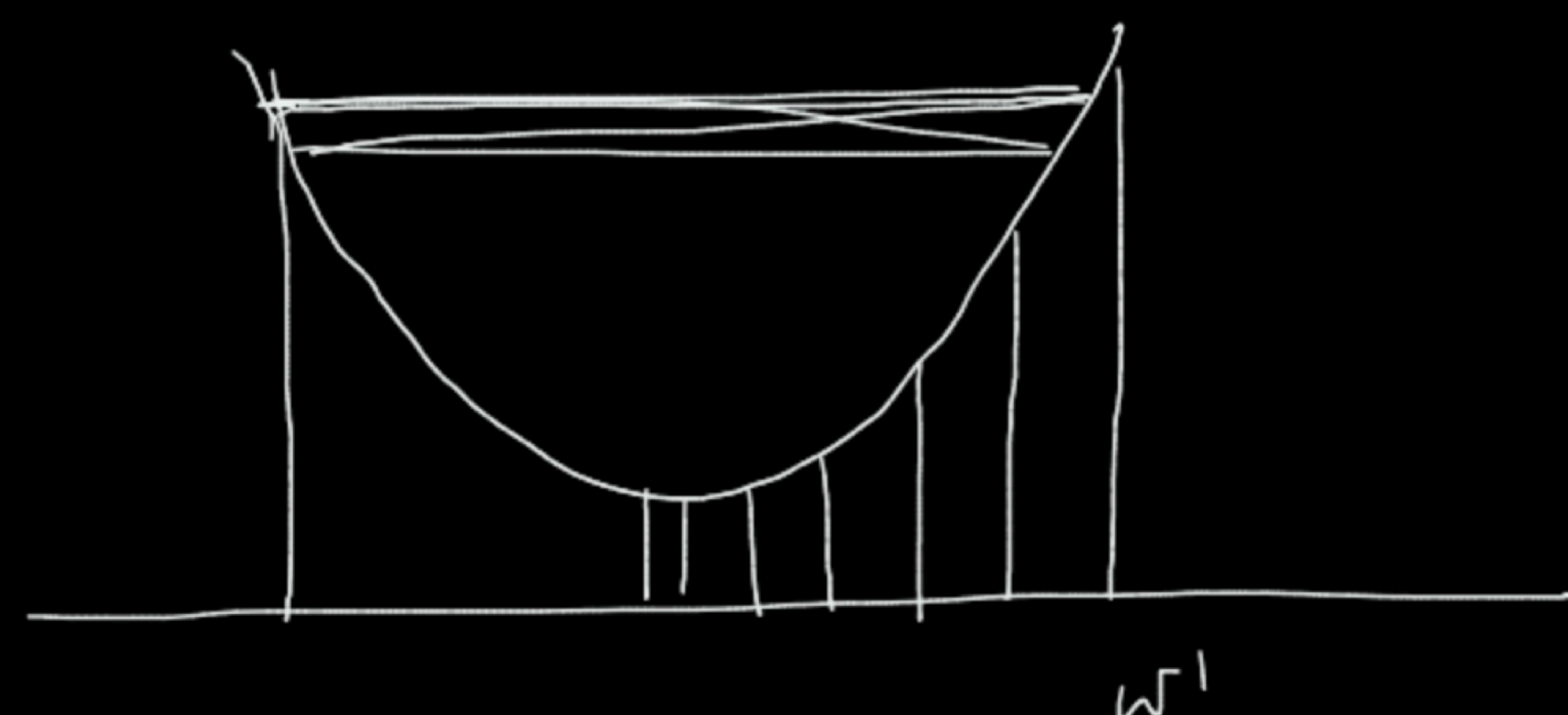
$$w_1^{(2)} = w_1^{(1)} + \left[-\frac{\partial L}{\partial w_1} \right]_{w_1^{(1)}}$$

$$w_2^{(2)} = w_2^{(1)} + \left[-\frac{\partial L}{\partial w_2} \right]_{w_2^{(1)}}$$

Step 4, Repeat steps 2 & 3 until Convergence $w_1^{old} = w_1^{new}$

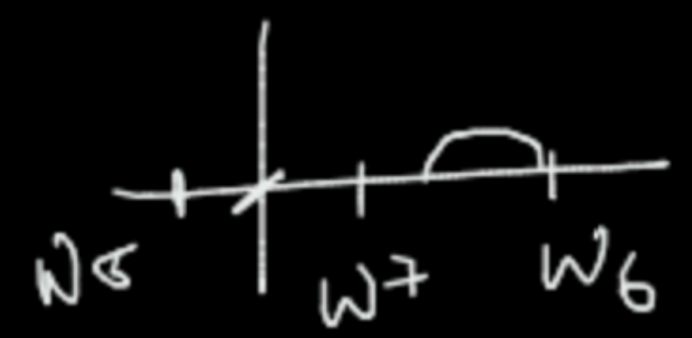
$$w^2 = w^1 + \lambda \underbrace{\overbrace{\left[-\frac{\partial L}{\partial w}\right]_{w^1}}^{\Delta w}}_{\text{step size}}$$

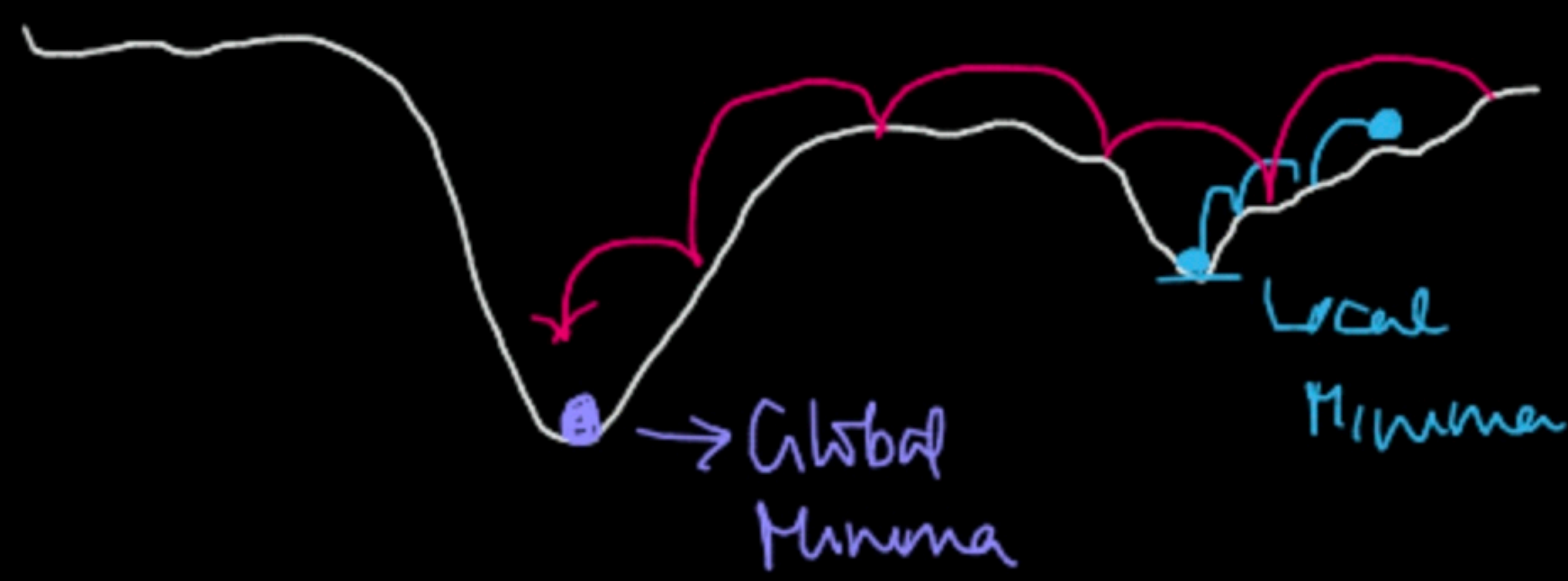
Learning rate ← λ

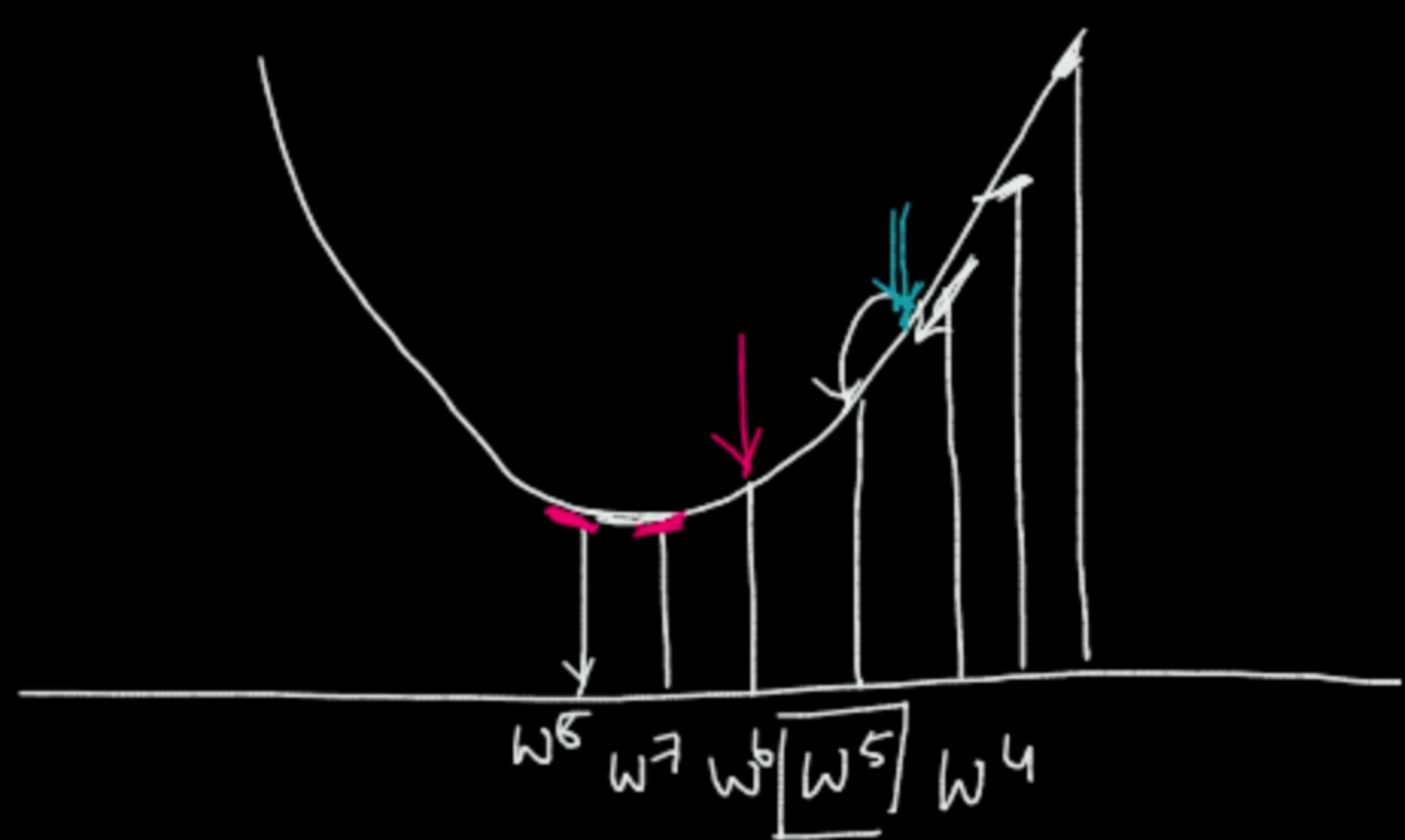


$$w^{new} = w^{old} + \Delta w$$

Early stopping







optimizers → The chances of landing in local minima will be reduced

1 W/o Momentum — It looks at the past

$$w^{new} = w^{old} + \lambda \underbrace{\left[-\frac{\partial L}{\partial w}\right]_{w^{old}}}_{\text{step size}}$$

$$w^5 = w^4 + -\left[\frac{\partial L}{\partial w}\right]_{w^4}$$

With momentum

$$w^5 = w^4 + -\lambda \underbrace{\left[\frac{\partial L}{\partial w}\Big|_{w^4} + \frac{\partial L}{\partial w}\Big|_{w^3} + \frac{\partial L}{\partial w}\Big|_{w^2}\right]}_{\text{step size}}$$

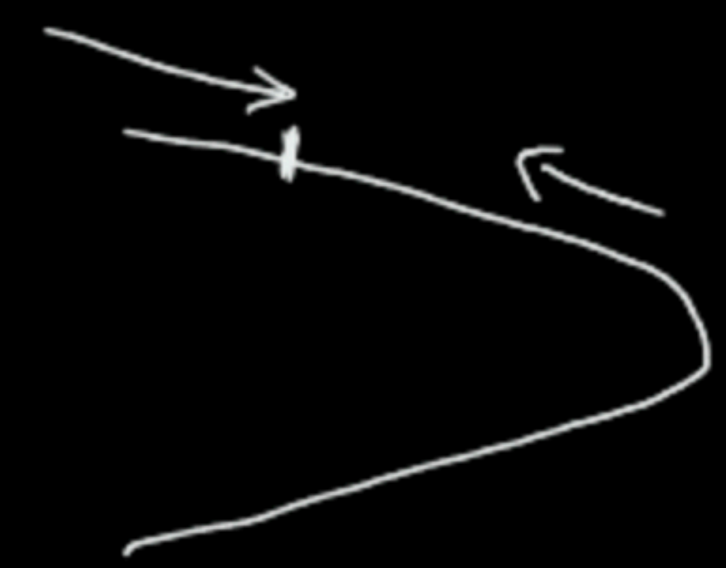
2. Nestrov Momentum ✓

— Look at the future

Adam —

RMSprop -

AdaGrad



gradient

$$-\lambda \left[\frac{\partial L}{\partial w}\right]$$

Types of Gradient descent