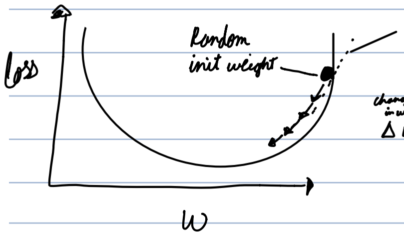
How can we find the set of weights that min loss

Mathematical expressed as $\arg\min_{w} \text{loss}(w)$

As w grows in size it is more difficult / becomes impos to compute manually

Gradient Descent helps:



Gradient $\frac{\partial \text{loss}}{\partial w}$

$\frac{d\,\text{loss}}{d\,\text{weights}}$

$$\underset{\text{change in weight}}{\Delta w} = \underset{\text{new weight}}{w} - \underset{\text{Learning Rate}}{\alpha} \frac{\partial \text{loss}}{\partial w}$$

loss : $(\hat{y} - y)^2 = (x * w - y)^2$

d loss : $2(x * w - y)$

$$\frac{d\,\text{loss}}{dw} : 2(x*w - y) * \frac{d}{dw}(x*w - y)$$

$$\left(2x \cdot \overset{1}{\frac{d}{dw}[w]} + \cancel{\overset{0}{\frac{d}{dw}[y]}}\right)(x*w - y)$$

$$2x(x*w - y)$$

Exercise

$$\hat{y} = x^2 w_2 + x w_1 + b$$

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

$$\frac{d\,\text{loss}}{dw_1} = (\hat{y} - y)^2 = 2(\hat{y} - y) = 2(x^2 w_2 + x w_1 + b)$$

$$= 2(x^2 w_2 + x w_1 + b) \cdot \frac{d}{dw_1}(x^2 w_2 + x w_1 + b)$$

$$= 2\left(\cancel{2x \overset{0}{\frac{d}{dw_1}[w_2]}} + x \overset{1}{\frac{d}{dw_1}[w_1]} + \cancel{\overset{0}{\frac{d}{dw_1}[b]}}\right) \cdot \frac{d}{dw_1}(x^2 w_2 + x w_1 + b)$$

$$\frac{d\,\text{loss}}{dw_1} \quad 2x(x)$$

$$\frac{d\,\text{loss}}{dw_2} = (\hat{y} - y)^2 = 2(\hat{y} - y) = 2(x^2 w_2 + x w_1 + b)$$

$$= 2(x^2 w_2 + x w_1 + b) \cdot \frac{d}{dw_2}(x^2 w_2 + x w_1 + b)$$

$$= 2\left(2x \overset{1}{\frac{d}{}[w_2]} + \cancel{x \overset{0}{\frac{d}{}[w_1]}} + \cancel{\overset{0}{\frac{d}{}[b]}}\right) \cdot \frac{d}{}\left(x^2 \overset{1}{\frac{d}{}[w_2]} + \cancel{\overset{0}{x w_1}} + \cancel{\overset{0}{b}}\right)$$

$$\frac{d\,loss}{dw} \quad 2x\,(x^2) = 3x^2$$