$$\sum_{i=1}^{N=2} (\varepsilon_i)^2$$

$$\varepsilon^T = [\cdot \ \cdot \ \cdot \ \cdot \ \cdot]$$

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix} \in \mathbb{R}^{n\times 1}$$

$$\varepsilon^T \in \mathbb{R}^{\cancel{1}\times n = 1}$$

$$\varepsilon^T \cdot \varepsilon \in \underbrace{\mathbb{R}^{1\times n} \times \mathbb{R}^{n\times 1}}$$

$$\mathbb{R}^{1\times 1} = \mathbb{R}$$

$$\varepsilon^T \varepsilon$$

$$[\cdot \ \cdot \ \cdots] \begin{bmatrix} \cdot \\ \cdot \\ \vdots \end{bmatrix} = \sum_{i=1}^{N} \varepsilon_i^2$$

$$\varepsilon \varepsilon^T \in \mathbb{R}^{n\times n}$$

$$tr(\varepsilon \varepsilon^T)$$

$$l(\theta) \qquad \nabla l = 0$$

$f(\cdot)$ ∇_θ

θ^*

$$\nabla_\theta f \circ g(\theta) = g'(\theta) \cdot f'(g(\theta))$$

$$f(x) = x^2 \qquad f'(x) = 2x$$

$$\boxed{\nabla_x f(g(x))} = \nabla_x (g(x))^2$$

$$g'(x)\, f'(g(x))$$

$$2\, g'(x)\, g(x)$$

$$\nabla_x f(\underbrace{g(x)}_{y}) = \frac{\partial f}{\partial g} \cdot \frac{\partial g}{\partial x}$$

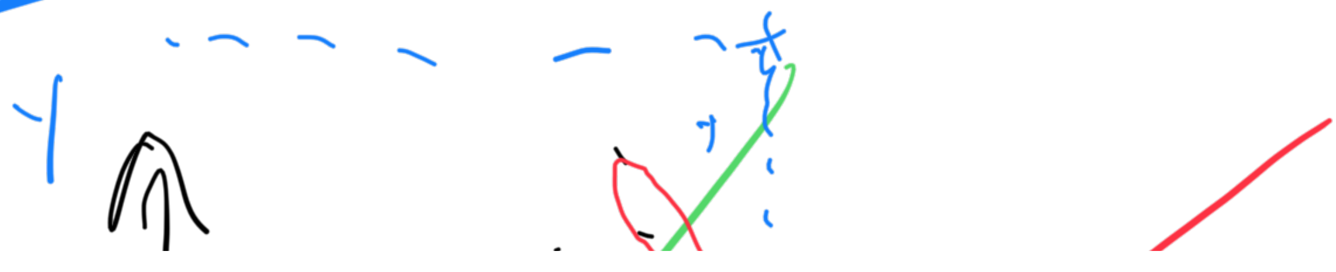
$$= 2g \cdot g'(x)$$

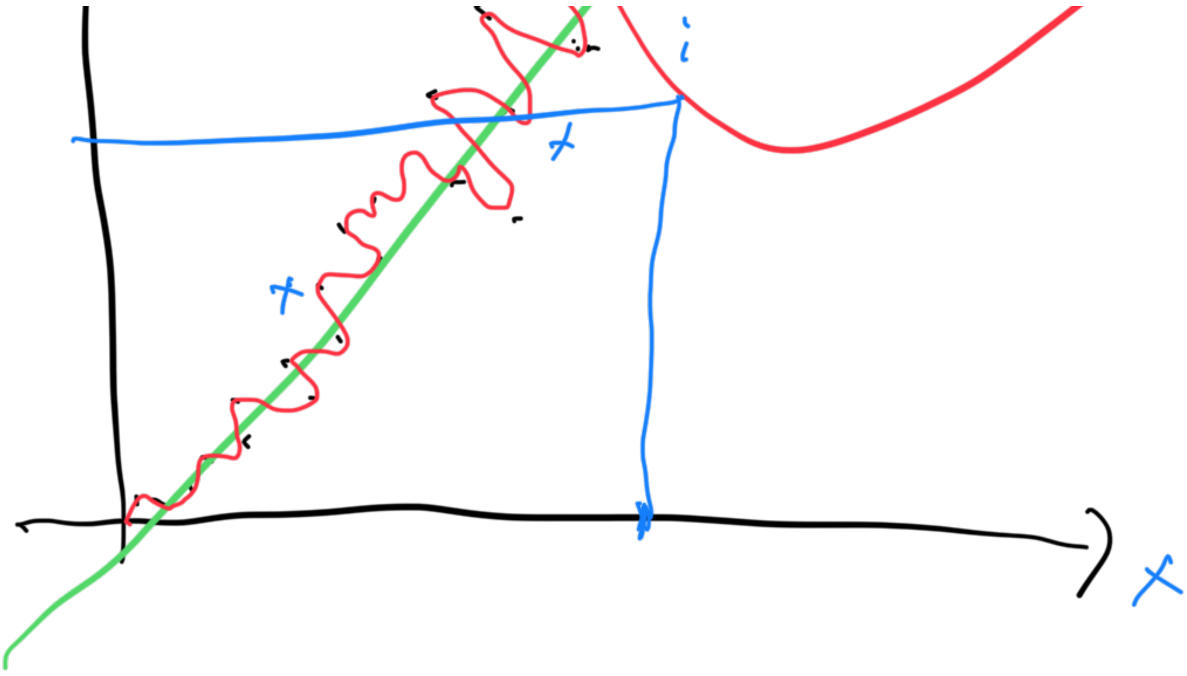
$$=$$

x^n

$\boxed{n x^{n-1}}$

$\frac{1}{x^4}$ x^{-1}





$$\hat{y} = w_1 x$$

$$y' = w_2 x$$

$$\|w_1\| \gg \|w_2\|$$

$$w_1(x+\varepsilon) = w_1 x + w_1 \varepsilon$$

$$\mathcal{L}(w) = MSE(w) + \underset{\underset{1/2}{=}}{\lambda} \|w\|$$

$$\min \mathcal{L}(w)$$