Slide 31~33

$$\boxed{\tilde{J} = \frac{1}{N}\sum L_i}$$ → calculated using x_i & w_i $\boxed{\& b_i}$ for now. incorporated

$$\boxed{\frac{\partial L}{\partial w_j} = (y-t)x^j}$$

$$\nabla_W J = \frac{\partial J}{\partial W} = \left(\frac{\partial J}{\partial w_1}, \frac{\partial J}{\partial w_2}, \dots, \frac{\partial J}{\partial w_D}\right)$$

weight matrix, not a scalar

$$W \leftarrow W - \underbrace{a}_{\text{learning rate}}\frac{\partial J}{\partial W} = W - a\sum_{i=0}^{N}\frac{\partial L_i}{\partial w_i}$$

$$\boxed{W \leftarrow W - \frac{a}{N}\sum_{i=0}^{N}(y^i - t^i)x^i}$$ how GD update the weight matrix, after each iteration epoch

optimal $= \frac{\partial J}{\partial W} = 0$. $\frac{\partial J}{\partial w} = 0$

$\| [[1,1],[1,2]] \|_2^2$

→ Calculate the euclidean distance given a vector — Euclid. dist

$R = \frac{1}{2}\|W\|_2^2$ ⟶ L_2 regularization

constant — a tunable hyperparameter

$$W \leftarrow W - a\frac{\partial (J + \lambda R)}{\partial W}$$

$$W - a\left(\frac{\partial J}{\partial W} + \frac{\partial \lambda R}{\partial W}\right) \Longleftrightarrow W - a\left(\frac{\partial J}{\partial W} + \boxed{\lambda\frac{\partial R}{\partial W}}\right)$$ → w, $= \lambda w$

$$W \leftarrow W - a\left(\frac{\partial J}{\partial w} + \lambda w\right)$$

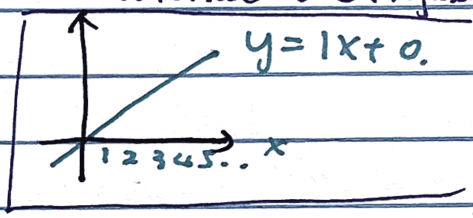
$$W \leftarrow (1 - a\lambda)w - a\boxed{\frac{\partial J}{\partial w}}$$

θ = Config. of parameters = (W, b) Slide 36

$$\mathcal{J}(\theta) = \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}^{(i)} = \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}(y(x^{(i)}, \theta), t^{(i)})$$

$$\frac{\partial \mathcal{J}}{\partial \theta} = \frac{1}{N}\sum^{N} \frac{\partial \mathcal{L}^{(i)}}{\partial \theta}$$ impractical.

tutorial: linear Regressor



θ ↓ (b

$W = [1, W]$

↓

dummy index

$$W \approx \theta - \alpha \frac{\partial \mathcal{L}^{i}}{\partial \theta}$$