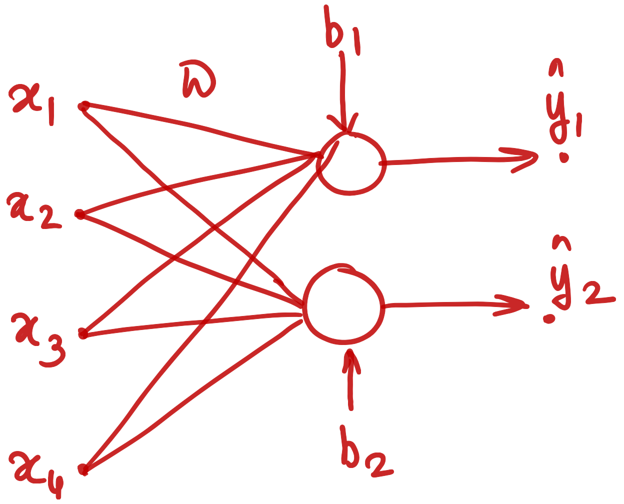
Linear Layer

$x = [x_1 \; x_2 \; x_3 \; x_4]$

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



$$X = [\text{---} x \text{---}]_{n \times 4}, \; n = \text{samples}$$

$$w = [\quad]_{4 \times 2}$$

$$\hat{y} = [\quad]_{n \times 2}$$

single layer perceptron

nn.linear(4, 2, bias = True)

$$X = \begin{bmatrix} \text{---} x_1 \text{---} & 1 \\ \text{---} x_2 \text{---} & 1 \\ \vdots & \\ \text{---} x_n \text{---} & 1 \end{bmatrix}_{n \times 5} \qquad \theta = \begin{bmatrix} w \\ b \end{bmatrix}_{5 \times 2}$$

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

$$\therefore MSE = J(\theta) = \frac{1}{2n} \sum_i (\hat{y}_i - y_i)^2$$

$$= \frac{1}{2n} \sum_{i=1}^{n} (x_i\theta - y_i)^2 = \frac{1}{2n} \left((x_0\theta_0 - y_0)^2 + (x_1\theta_1 - y_1)^2 + \ldots \right)$$

$$\therefore \nabla_\theta J(\theta) = \begin{bmatrix} \frac{\partial J}{\partial \theta_0} & \frac{\partial J}{\partial \theta_1} & \ldots \end{bmatrix}$$

$$\nabla_\theta J(\theta) = \frac{1}{n} \sum_i (x_i\theta - y_i) . x_i$$

$$\underset{5 \times 2}{\nabla_\theta J(\theta)} = \frac{1}{n} \underset{5 \times n}{X^T} . \underset{n \times 2}{[X\theta - Y]}$$

$$\therefore \quad \theta = \theta - \alpha \nabla_{\theta} J(\theta)$$

$$\therefore \quad \boxed{\theta = \theta - \frac{\alpha}{n} X^T \cdot [X\theta - Y]}$$

Gradient Descent