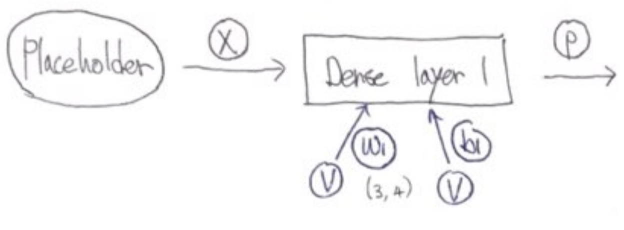
Practice.

$$X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \\ X_{31} & X_{32} \end{bmatrix} \qquad y = [\, y_1 \quad y_2 \,]$$

Dense 1.

- 

구하고 싶은 것   구하고 싶은 것.

- Initialization of W_1 & b_1 : $W_1 := \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \\ \theta_{31} & \theta_{32} & \theta_{33} \\ \theta_{41} & \theta_{42} & \theta_{43} \end{bmatrix}$, $b_1 := \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}$

- Then $P = I(W_1 X + b_1)$ where I is the identity mapping.

$$:= \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \\ p_{31} & p_{32} \\ p_{41} & p_{42} \end{bmatrix} \quad \text{where} \quad \begin{cases} p_{11} = \theta_{11}X_{11} + \theta_{12}X_{21} + \theta_{13}X_{31} + b_1 \\ p_{21} = \theta_{21}X_{11} + \theta_{22}X_{21} + \theta_{23}X_{31} + b_2 \\ \vdots \end{cases}$$

- Note).
  - p_{11} : (X_{11}, X_{21}, X_{31})에 대한 선형회귀 1 (가중치 : $(\theta_{11}, \theta_{12}, \theta_{13})$).
    : (X_{11}, X_{21}, X_{31})에 대한 정보를 축약한 뉴런.
  - p_{21} : (X_{11}, X_{21}, X_{31})에 대한 선형회귀 2 (가중치 : $(\theta_{21}, \theta_{22}, \theta_{23})$)
    : (X_{11}, X_{21}, X_{31})를 다른 관점으로 봤다고 할수 있음.

Dense 2.

- $$\begin{tikzcd} \boxed{\text{Dense layer 1}} \arrow[r, "P"] & \boxed{\text{Dense layer 2}} \arrow[r, "Q"] & {} \\ & W_2 \arrow[u] \quad b_2 \arrow[u] & \end{tikzcd}$$

  $V \quad U$

개별 설정값 개별 설정치

- Initialization of W_2 & b_2 : $W_2 := [\,\delta_{11}, \delta_{12}, \delta_{13}, \delta_{14}\,]$, $b_2 := [\,d_1\,]$

- Then $Q = I\left(W_2 P + b_2\right)$

$$= [\, q_{11} \quad q_{12} \,] \quad \text{where} \left[\begin{array}{l} q_{11} = \delta_{11} p_{11} + \delta_{12} p_{21} + \delta_{13} p_{31} + \delta_{14} p_{41} + d_1 \\ q_{12} = \delta_{11} p_{12} + \delta_{12} p_{22} + \delta_{13} p_{32} + \delta_{14} p_{42} + d_1 \end{array} \right.$$

Loss Function.

- $$\begin{tikzcd} {} \arrow[r, "Q"] & \boxed{\text{loss function}} \arrow[r, "L"] & {} \end{tikzcd}$$

- $$L = \frac{1}{2} \sum_{k=1}^{2} \left(y_k - q_{1k}\right)^2$$

$$= \frac{1}{2} \left\{ (y_1 - q_{11})^2 + (y_2 - q_{12})^2 \right\}.$$

Optimizer.

- $$\begin{tikzcd} {} \arrow[r, "L"] & \boxed{\text{Optimizer.}} \arrow[r, "\text{gradient}"] & {} \end{tikzcd}$$

- $$\begin{bmatrix} \frac{\partial L}{\partial \theta_{11}} \\ \frac{\partial L}{\partial \theta_{12}} \\ \vdots \\ \frac{\partial L}{\partial d_1} \end{bmatrix} = \text{gradient.}$$

- $$\frac{\partial L}{\partial \theta_{ij}} = -\left((y_1 - q_{11}) \delta_{1i} x_{j1} + (y_2 - q_{12}) \delta_{1i} x_{j2} \right)$$

- $$\frac{\partial L}{\partial b_i} = -\left((y_1 - q_{11}) \delta_{1i} + (y_2 - q_{12}) \delta_{1i} \right)$$

- $$\frac{\partial L}{\partial \delta_{ij}} = -\left((y_1 - q_{11}) p_{j1} + (y_2 - q_{12}) p_{j2} \right)$$

- $$\frac{\partial L}{\partial d_i} = -\left((y_1 - q_{11}) + (y_2 - q_{12}) \right)$$

$$\frac{\partial L}{\partial \theta_{ij}} = \frac{1}{2}\frac{\partial}{\partial \theta_{ij}}\left\{ (y_1 - q_{11})^2 + (y_2 - q_{12})^2 \right\}$$

$$= \frac{1}{2}\left\{ \frac{\partial}{\partial \theta_{ij}}(y_1 - q_{11})^2 + \frac{\partial}{\partial \theta_{ij}}(y_2 - q_{12})^2 \right\}$$

$$= \frac{1}{2}\left\{ -2(y_1 - q_{11})\frac{\partial}{\partial \theta_{ij}} q_{11} - 2(y_2 - q_{12})\frac{\partial}{\partial \theta_{ij}} q_{12} \right\}.$$

$$= \frac{1}{2}\left\{ -2(y_1 - q_{11})\frac{\partial}{\partial \theta_{ij}}\left(\delta_{11}p_{11} + \delta_{12}p_{21} + \delta_{13}p_{31} + \delta_{14}p_{41} + d_1 \right) - 2(y_2 - q_{12})\frac{\partial}{\partial \theta_{ij}}\left(\delta_{11}p_{12} + \delta_{12}p_{22} + \delta_{13}p_{32} + \delta_{14}p_{42} + d_1 \right) \right\}$$

$$= \frac{1}{2}\left\{ -2(y_1 - q_{11})\frac{\partial}{\partial \theta_{ij}} \delta_{1i}p_{i1} - 2(y_2 - q_{12})\frac{\partial}{\partial \theta_{ij}} \delta_{1i}p_{i2} \right\}.$$

$$= \frac{1}{2}\left\{ -2(y_1 - q_{11})\delta_{1i}X_{j1} - 2(y_2 - q_{12})\delta_{1i}X_{j2} \right\}$$

$$= -\delta_{1i}\left((y_1 - q_{11})X_{j1} + (y_2 - q_{12})X_{j2} \right)$$