①

$z_{h_1} = 2\times1 + 1\times4 + (-1)(-2) = 2+4+2 = 8 \qquad h_1 = 8.$

$z_{h_2} = -2\times1 + 3\times4 + 4\times-2 = -2+12-8 = 2 \qquad h_2 = 2.$

$z_{y_1} = 8\times0.5 + 2\times(-0.5) = 4-1 = 3 \qquad y_1 = 3.$

$z_{y_2} = 8\times-0.5 + 2\times0.5 = -4+1 = -3 \qquad y_2 = \max(0,-3) = 0.$

②

a) $y^i = \max(0, \sum_{j=0}^{p} w_j x_j^i).$

b)

i) $\frac{\partial}{\partial w_j} f(w) = \sum_{i=1}^{n} (\hat{y}_i - y^i) \cdot \sigma(\sum_{j=0}^{p} x_j^i)$

$= \sum_{i=1}^{n} (\hat{y}_i - y_i) \cdot \max(0, \sum_{j=0}^{p} x_j^i).$

(ii) (1) $w_0 \dots w_p =$ rand()

(2) randomly choose training sample i

(3) $\hat{y}_i = \max(0, \sum_{j=0}^{p} w_j x_j^i)$

(4) Gradient Descent update $w_0 \dots w_j$

(5) If no convergence, goto line (2).

(c) $\nabla f^{i_t}(w^{(t)}) = \delta^{i_t} x^{i_t} + 2\lambda w^{(t)}$

$w^{(t+1)} = w^{(t)} - \alpha_t \delta^{i_t} x^{i_t} - 2\alpha_i \lambda w^{(t)}.$