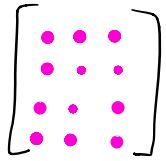
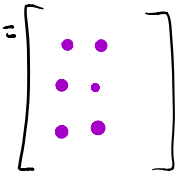
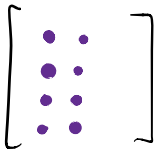
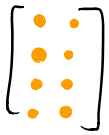
Forward (3 Layers)

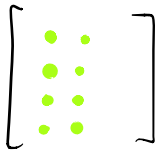
X = 
$n \times d$
4×3

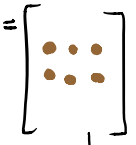
W_1 = 
$d \times h$
3×2

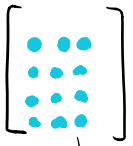
$X \cdot W_1$ = 
$n \times h$
4×2

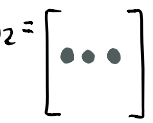
b_1 = 
$1 \times h$
1×2

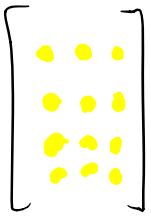
$H = X \cdot W_1 + b =$ 

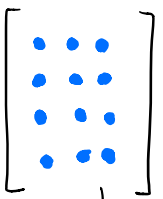
$A = \text{Relu}(X \cdot W_1 + b) =$ 
MLP input
$n \times h$
4×2

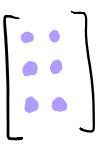
$\times$ W_2 = 
$h \times d$
2×3

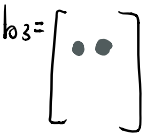
$=$ $A \cdot W_2 =$ 
$n \times d$
4×3

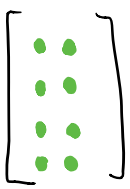
b_2 = 
$1 \times d$
1×3

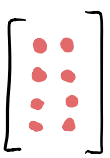
$O = A \cdot W_2 + b_2 =$ 
$n \times d$
4×3

$q = \text{Relu}(O) =$ 
$n \times d$
4×3

$\times$ W_3 = 
$d \times c$
3×2

b_3 = 
$1 \times c$
1×2

$S = q \cdot W_3 + b_3 =$ 
4×2
$n \times c$

$P = \text{softmax}(q \cdot W_3 + b_3) =$ 
$n \times c$
4×2

$L = -\text{Log likelihood}(P)$

Backward (3 Layers)

$\frac{\partial L}{\partial S} = P - T$ dp T는 레이블(label) $\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$ $n \times c$ 4×2

$\frac{\partial L}{\partial W_3} = \frac{\partial S}{\partial W_3} \cdot \frac{\partial L}{\partial S} = q^T \cdot (P - T)$ $\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$ $d \times c$ 3×2

q^T: $d \times n$ 3×4, $(P-T)$: $n \times c$ 4×2

$\frac{\partial L}{\partial b_3} = \frac{\partial S}{\partial b_3} \cdot \frac{\partial L}{\partial S} = 1 \cdot \frac{\partial L}{\partial S} = P - T$ $\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$ $n \times c$ 4×2

$\frac{\partial L}{\partial q} = \frac{\partial L}{\partial S} \cdot \frac{\partial S}{\partial q} = (P - T) \cdot W_3^T$ $\begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix}$ $n \times d$ 4×3

$(P-T)$: $n \times c$ 4×2, W_3^T: $c \times d$ 2×3

$\frac{\partial L}{\partial O} = \frac{\partial q}{\partial O} \cdot \frac{\partial L}{\partial q} = Relu'(O) \odot \frac{\partial L}{\partial q}$ $\begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix}$ $n \times d$ 4×3

$Relu'(O)$: $n \times d$ 4×3, $\frac{\partial L}{\partial q}$: $n \times d$ 4×3

$\frac{\partial L}{\partial W_2} = \frac{\partial O}{\partial W_2} \cdot \frac{\partial L}{\partial O} = A^T \cdot \frac{\partial L}{\partial O} = A^T \cdot Relu'(O) \odot \frac{\partial L}{\partial q} = \left[A^T \cdot \left[Relu'(O) \odot \left[(P-T) \cdot W_3^T \right]_{dp} \right]_{dq} \right]$

A^T: $h \times n$ 2×4, $\frac{\partial L}{\partial O}$: $n \times d$ 4×3 $\rightarrow$ $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ $h \times d$ 2×3

$\frac{\partial L}{\partial b_2} = \frac{\partial O}{\partial b_2} \cdot \frac{\partial L}{\partial O} = 1 \cdot \frac{\partial L}{\partial O} = Relu'(O) \odot \frac{\partial L}{\partial q} = \left[Relu'(O) \odot \left[(P-T) \cdot W_3^T \right]_{dp} \right]_{dq}$ $\begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix}$ $n \times d$ 4×3

$Relu'(O)$: $n \times d$ 4×3, $\frac{\partial L}{\partial q}$: $n \times d$ 4×3

$\frac{\partial L}{\partial A} = \frac{\partial L}{\partial O} \cdot \frac{\partial O}{\partial A} = \frac{\partial L}{\partial O} \cdot W_2^T = \left[Relu'(O) \odot \left[(P-T) \cdot W_3^T \right]_{dp} \right]_{dq} \cdot W_2^T$ $\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$ $n \times h$ 4×2

$\frac{\partial L}{\partial O}$: $n \times d$ 4×3, W_2^T: $d \times h$ 3×2

$\frac{\partial L}{\partial H} = \frac{\partial A}{\partial H} \cdot \frac{\partial L}{\partial A} = f'(H) \odot \frac{\partial L}{\partial A} = \left[f'(H)_{da} \odot \left[Relu'(O) \odot \left[(P-T) \cdot W_3^T \right]_{dp} \right]_{dq} \cdot W_2^T \right]$

$f'(H)$: $n \times h$ 4×2, $\frac{\partial L}{\partial A}$: $n \times h$ 4×2 $\rightarrow$ $\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$ $n \times h$ 4×2

※ $f'(H)_{da} = Relu'(H) = (A'(H)) = \frac{\partial A}{\partial H}$

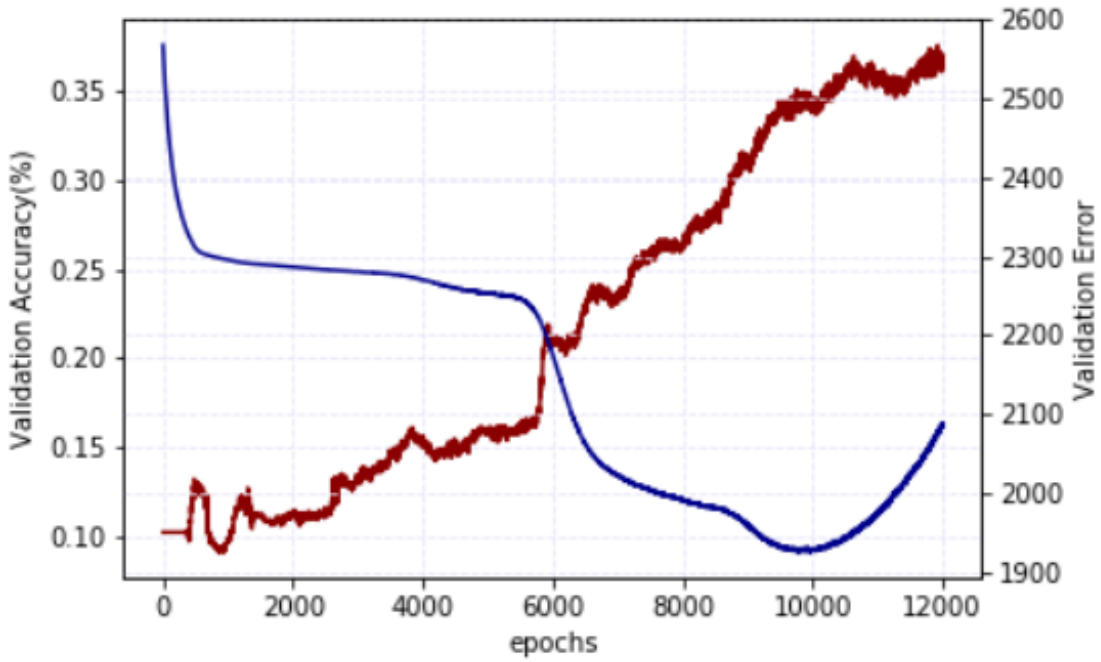
$\frac{\partial L}{\partial W_1} = \frac{\partial H}{\partial W_1} \cdot \frac{\partial L}{\partial H} = X^T \cdot \frac{\partial L}{\partial H} = X^T \cdot f'(H) \odot \frac{\partial L}{\partial A} = \left[X^T \cdot \left[f'(H)_{da} \odot \left[Relu'(O) \odot \left[(P-T) \cdot W_3^T \right]_{dp} \right]_{dq} \cdot W_2^T \right] \right]$

X^T: $d \times n$ 3×4, $\frac{\partial L}{\partial H}$: $n \times h$ 4×2 $\rightarrow$ $\begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$ $d \times h$ 3×2

$\frac{\partial L}{\partial b_1} = \frac{\partial H}{\partial b_1} \cdot \frac{\partial L}{\partial H} = 1 \cdot \frac{\partial L}{\partial H} = f'(H) \odot \frac{\partial L}{\partial A} = \left[f'(H)_{da} \odot \left[Relu'(O) \odot \left[(P-T) \cdot W_3^T \right]_{dp} \right]_{dq} \cdot W_2^T \right]$ $\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$ $n \times h$ 4×2

$\frac{\partial L}{\partial H}$: $n \times h$ 4×2

ⓒ 3 Layers 2번째 결과 그래프



나의 최고 validation accuracy : 0.376