x :
y
sa
x^i_j
x, s, a
y
$a =$
$(a\,bcdefghi) a^+ =$
$(1\,abc1def1ghi) a^- =$
$(b\,cefhi)$
$A \circ$
B
( )
x^0_0
$+1$
x^0_1
x^0_2
x^0_3
a^1_0
$+1$
$s^1_1 | a^1_1$
$s^1_2 | a^1_2$
$s^1_3 | a^1_3$
$s^1_4 | a^1_4$
$s^1_5 | a^1_2$
$s^1_6 | a^1_2$
$s^2_1 | a^2_1$
$s^2_2 | a^2_2$
$s^2_3 | a^2_3$
w^1_{01}
w^1_{11}
w^1_{21}
w^1_{31}
w^2_{01}
w^2_{11}
w^2_{21}
w^2_{31}
w^2_{41}
w^2_{51}
w^2_{61}
$\left(x^{(0)}\right)^+$
${W_b}^{(1)}$
$s^{(1)} | \left(a^{(1)}\right)^+$
${W_b}^{(2)}$
$s^{(2)} | a^{(2)}$

$\left(^{(0)}\right)^+ w^{(1)} =$
$s^{(1)}, \sigma(s^{(1)}) =$
$a^{(1)}$
$\left(a^{(1)}\right)^+ w^{(2)} =$
$s^{(2)}, \sigma(s^{(2)}) =$
$a^{(2)}$
$Loss L =$
$\frac{1}{2} \sum_{k=1}^{3} (a^{(2)}_k -$
$y_k)^2$
x
$1 \times$
3
x^+
$1 \times$
4
$w^{(1)}$
$4 \times$
6
$s^{(1)}$
$1 \times$
6
$a^{(1)}$
$1 \times$
6
$\left(a^{(1)}\right)^+$
$1 \times$

$\sigma(z)(1-\sigma(z))$

$\sigma(z) = \frac{e^z-e^{-z}}{e^z+e^{-z}} = 1-\frac{2e^{-z}}{e^z+e^{-z}}$

$\sigma'(z) = 1-\sigma(z)^2$

$\sigma(z) = \{ 0, z < 0$ $z, .$

$\sigma'(z) = \{ 0, z < 0$ $1, .$

$\sigma(z) = \{ z, z > 0$ $az, .$

$\sigma'(z) = \{ 1, z > 0$ $a, .$

$i =$

$a_i^{(N+1)} = \sigma(s_i^{(N+1)}) = \frac{e^{s_i^{(N+1)}}}{\sum_{j=1}^{K} e^{s_j^{(N+1)}}}$

$\partial p_i \frac{}{\partial s_j = \frac{\partial \frac{e^{s_i}}{\sum_{k=1}^{N} e^{s_k}}}{\partial s_j} f(x)=\frac{g(x)}{h(x)}, f'(x)=\frac{g'(x)h(x)-h'(x)g(x)}{h^2(x)} g(x)=e^{s_i}, h(x)=\sum_{k=1}^{N} e^{s_k}}$

$\neq j$

$\frac{\partial \frac{e^{s_i}}{\sum_{k=1}^{N} e^{s_k}}}{\partial s_j} = \frac{e^{s_i}\sum_{k=1}^{N} e^{s_k} - e^{s_j}e^{s_i}}{\left(\sum_{k=1}^{N} e^{s_k}\right)^2} \frac{\partial \frac{e^{s_i}}{\sum_{k=1}^{N}}}{\partial s_j} = \frac{0-e^{s_j}e^{s_i}}{\left(\sum_{k=1}^{N} e^{s_k}\right)^2}$

$= \frac{e^{s_i}\left(\sum_{k=1}^{N} e^{s_k} - e^{s_j}\right)}{\left(\sum_{k=1}^{N} e^{s_k}\right)^2} = -\frac{e^{s_i}}{\sum_{k=1}^{N} e^{s_k}} \frac{e^{s^j}}{\sum_{k=1}^{N} e^{s_k}}$

$= p_i(1-p_j) = -p_i p_j$

$\delta_{ij} = \{ 1\ i == j 0 i \neq j$

$\Rightarrow$

$$\frac{\partial p_i}{\partial s_j} = p_i(\delta_{ij} - p_j)$$

$\frac{}{2\sum_{i=1}^{K}(z_i-y_i)^2 \frac{\partial Loss}{\partial z} = (z-y)}$

$\sum_{i=1}^{K} y_i \log z_i = -y_k \log z_k (one-hot k 1)$

$\frac{\partial Loss}{\partial z} = -\frac{y}{z}$

$$\frac{\partial Loss(a^{(N+1)}, y)}{\partial s_j^{(N+1)}} = a_j^{(N+1)} - y_j$$

$\partial Loss(a^{(N+1)}, y) \frac{}{\partial s_j^{(N+1)} = \sum_{i=1}^{K} \frac{\partial Loss(a^{(N+1)},y)}{\partial a_i} \frac{\partial a_i}{\partial s_j} = -\sum_{i=1}^{K} \frac{y_i}{a_i} \frac{\partial a_i}{\partial s_j} = \left(-\frac{y_i}{a_i}\frac{\partial a_i}{\partial s_j}\right)_{i=j} - \sum_{i=1, i\neq j}^{K} \frac{y_i}{a_i}\frac{\partial a_i}{\partial s_j} = -\frac{y_j}{a_j} a_j(1-a_j) - \sum_{i=1, i\neq j}^{K} \frac{y_i}{a_i} \cdot -a_i a_j =}$

X	$N_T \times N_t$	$A^{(0)}$
X^+	$N_T \times (N_t+1)$	$\left(A^{(0)}\right)^+$
$S^{(i)}$	$N_T \times N^{(i)}$	$i = 1, \cdots, H, H+1$
$A^{(i)}$	$N_T \times N^{(i)}$	$i = 0, 1, \cdots, H, H+1$
$\left(A^{(i)}\right)^+$	$N_T \times (N^{(i)}+1)$	$i = 0, 1, \cdots, H$
$W^{(i)}$	$N^{(i-1)} \times N^{(i)}$	$i = 1, \cdots, H, H+1$
$W^{(H+1)}$	$N^{(N)} \times K$	
$W^{(i)}$	$(N^{(i-1)}+1) \times N^{(i)}$	$i = 1, \cdots, H, H+1$

$$
\begin{array}{cccc}
 & N_T \times (N_t + 1) & & \\
S^{(1)} & N_T \times N^{(1)} & \Delta^{(H+1)} & N_T \times K \\
(A^{(1)})^{+} & N_T \times (N^{(1)} + 1) & \nabla_{W^{(H+1)}} \mathcal{L} & (N^{(H)} + 1) \times K \\
S^{(2)} & N_T \times N^{(2)} & \Delta^{(H)} & N_T \times (N^{(H)} + 1) \\
(A^{(2)})^{+} & N_T \times (N^{(2)} + 1) & \nabla_{W^{(N)}} \mathcal{L} & (N^{(H-1)} + 1) \times N^{(H)} \\
 & & & \\
S^{(H)} & N_T \times N^{(H)} & \Delta^{(i)} & N_T \times (N^{(i)} + 1) \\
(A^{(H)})^{+} & N_T \times (N^{(H)} + 1) & \nabla_{W^{(i)}} \mathcal{L} & (N^{(i-1)} + 1) \times N^{(i)} \\
S^{(H+1)} & N_T \times K & \Delta^{(1)} & N_T \times (N^{(1)} + 1) \\
A^{(H+1)} & N_T \times K & \nabla_{W^{(1)}} \mathcal{L} & (N_t + 1) \times N^{(1)}
\end{array}
$$

$$
\frac{\Delta J\left(\theta_0, \theta_1\right)}{\Delta \theta_j} = \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta \left(x^{(i)} \right) - y^{(i)} \right) x_j^{(i)}
$$

$i = 1, 2, \cdots, m$
$j = 0, 1$
$x_0^{(i)} = 1$

$$
\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta \left(x^{(i)} \right) - y^{(i)} \right) x_j^{(i)}
$$

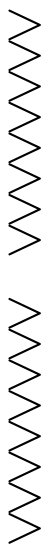
$repeat\{$
$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^{m} \left(h_\theta \left(x^{(i)} \right) - y^{(i)} \right) x_j^{(i)}$
$(for j = 0, 1)$
$\}$

$$
\frac{\Delta J^{(i)}\left(\theta_0, \theta_1\right)}{\theta_j} = \left(h_\theta \left(x^{(i)} \right) - y^{(i)} \right) x_j^{(i)}
$$

$$
\theta_j := \theta_j - \alpha \left(h_\theta \left(x^{(i)} \right) - y^{(i)} \right) x_j^{(i)}
$$

$repeat\{$
$for i = 1, \cdots, m\{$
$\theta_j := \theta_j - \alpha \left(h_\theta \left(x^{(i)} \right) - y^{(i)} \right) x_j^{(i)}$
$(for j = 0, 1)$
$\}$
$\}$
$repeat\{$
$for \mathrm{i} = 1, 11, 21, 31, \ldots, 991\{$
$\theta_j := \theta_j - \alpha \frac{1}{10} \sum_{k=i}^{(i+9)} \left(h_\theta \left(x^{(k)} \right) - y^{(k)} \right) x_j^{(k)}$
$(for j = 0, 1)$
$\}$
$\}$

$_feng12489/article/details/102856968$
$_feng12489/article/details/102913634 Xavier initialization He initialization$
$>$
$>$
$>$
$>$
$>$
$>$
$>$
$>$
$>$
$>$
$>$
$>$



□
□
□

$TrainSampleNum = BatchSize * BatchNum + remainder remainder = 0$

$BatchSize$

$v_t s_t$

$\beta_1 = 0.9, \beta_2 = 0.999, \eta = 10^{-3}, \epsilon = 10^{-8}$

$_t = \beta_1 v_{t-1} + (1 - \beta_1) g_t$

$s_t = \beta_2 s_{t-1} + (1 - \beta_2) g_t \odot g_t$

$\hat{v}_t = \frac{v_t}{1-\beta_1^t}$

$\hat{s}_t = \frac{s_t}{1-\beta_2^t}$

$g'_t = \frac{\eta \hat{v}_t}{\sqrt{\hat{s}_t} + \epsilon}$

$x_t = x_{t-1} - g'_t$