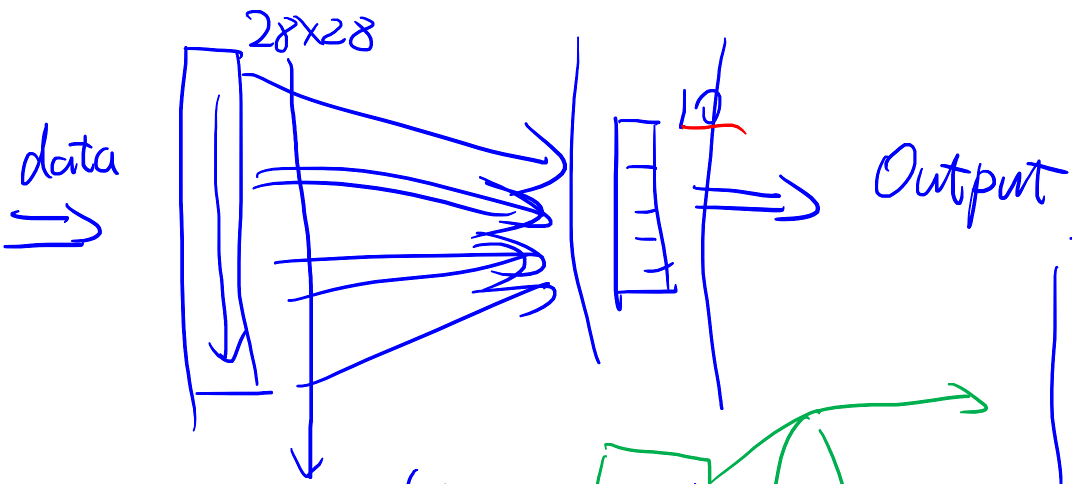


$$\underset{784}{l_0} = A(\underset{784}{data} + \underset{784}{\boxed{b_0}})$$

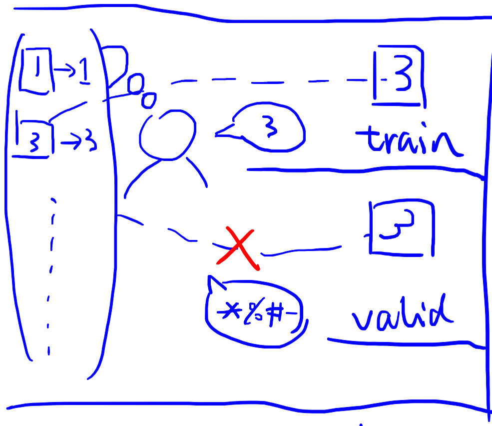
$$\underset{10}{output} = A'(\underset{784}{l_0} \cdot \underset{784\times10}{\boxed{W_1}} + \underset{10}{\boxed{b_1}})$$

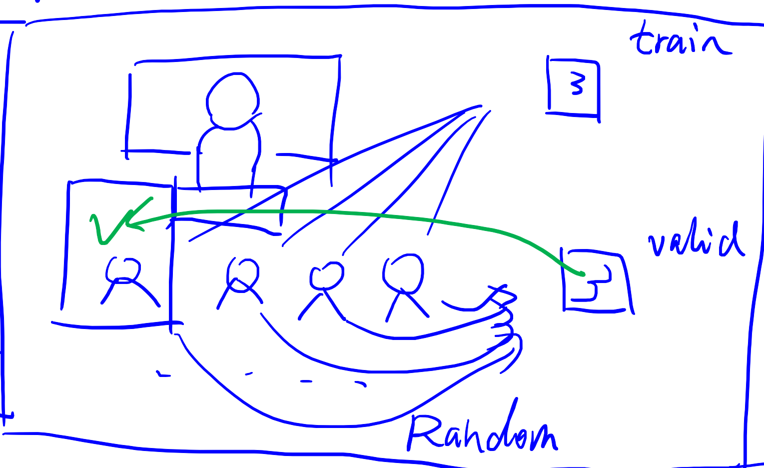
$$\begin{cases} A = \tanh \\ A' = \text{softmax} \end{cases}$$

b_0 : 784
b_1 : 10
W_1 : [784, 10]

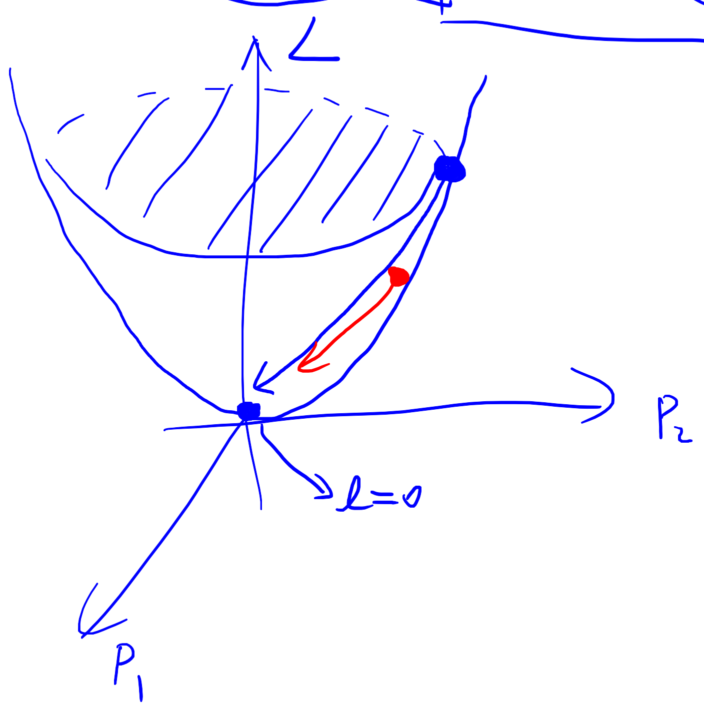
$b_0 \quad b_1 \rightarrow 0$

$$W_1 \in \left[-\sqrt{\frac{6}{784+10}}, \sqrt{\frac{6}{784+10}}\right)$$

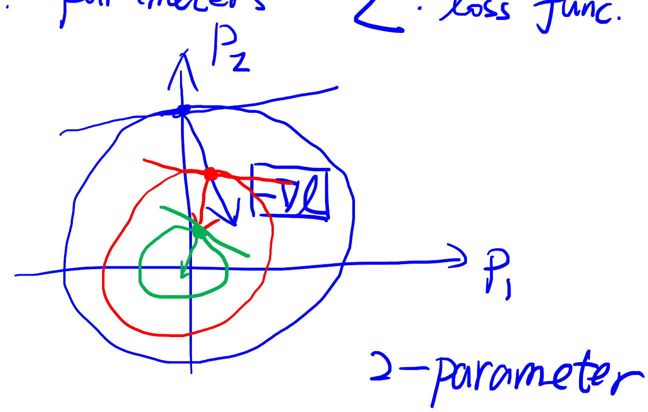


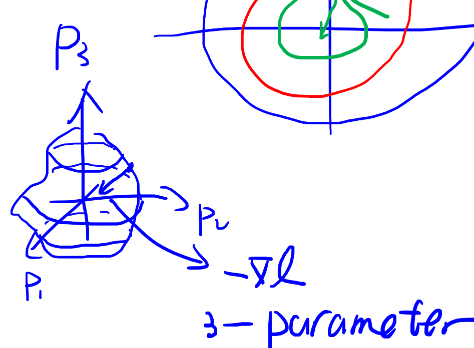


train
validation
test
real world

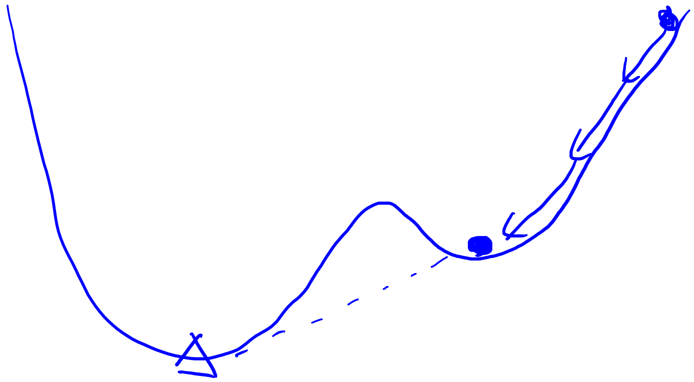


P_i : parameters L : loss func.



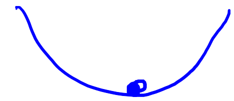


$$\vec{p} \to \vec{p} - k\nabla \vec{\ell}$$



$\left\{\begin{array}{l}\text{会出现} \\ \text{不用太担心，}\end{array}\right.$

已优化.



① $\longrightarrow$ ② $\frac{d}{dx}(af) = a\frac{d}{dx}f$

$$\frac{d}{dx}\left[f(g(x))\right] = f'(g(x))\,\underline{g'(x)}$$

③ ④

$$\frac{d}{dx}(f+g) = \frac{d}{dx}f + \frac{d}{dx}g$$

$$\frac{\partial}{\partial x}f(x,y) \longrightarrow \frac{d}{dx}f(x;y=y_0)$$

$$\frac{d}{dx}f(g_1(x), g_2(x)) = \sum_{i=1}^{2}(\partial_i f)(g_1,g_2)\,g_i'(x)$$

$$L = (y - y_{pred})^2 \qquad \vec{p} = (\vec{b}_1, \overleftrightarrow{w}_1, \vec{b}_0)$$

$$= (y - A_2(\vec{\ell}_0 \cdot w_1 + b_1))^2$$

$$= (y - A_2(A_1(x + b_0)\cdot w_1 + b_1))^2$$

$$\left(L = L_0 + \cdots + L_9 \qquad \frac{\partial L}{\partial \cdots} = \sum_{k=0}^{9}\frac{\partial L_k}{\partial \cdots}\right)$$

$$\frac{\partial L_0}{\partial b_{1,i}} = \frac{\partial}{\partial b_{1,i}}\left(y_0 - y_{pred,0}(b_{1,i})\right)^2$$

$$= 2(y_0 - y_{pred,0}(b_{1,i}))\,\frac{\partial}{\partial b_{1,i}}(\cancel{y_0} - y_{pred,0}(b_{1,i}))$$

$$= -2(y_0 - y_{pred,0}(b_{1,i}))\,\frac{\partial}{\partial b_{1,i}}y_{pred,0}(b_{1,i})$$

$$= \underbrace{\boxed{-2(y_0 - y_{pred,0}(b_{1,i}))}}_{1}\,\frac{\partial}{\partial b_{1,i}}A_{2,0}(\vec{\ell}_0 \cdot \overleftrightarrow{w}_1 + \vec{b}_1)$$

$$= \boxed{1} \sum_m (\partial_m A_{l_0})(\vec{l}_0 \cdot \vec{w}_1 + \vec{b}_1) \frac{\partial}{\partial b_{1i}} \left[\cancel{(\vec{l}_0 \cdot \vec{w}_1)_m} + \vec{b}_{1m} \right]$$

$$\begin{cases} 1 & i = m \\ 0 & i \neq m \end{cases}$$

$$= \boxed{1} (\partial_i A_{2d})(\vec{l}_0 \cdot \vec{w}_1 + b_1)$$

$$\boxed{\frac{\partial L}{\partial b_{1,i}} = -2(y - y_{pred})(\partial_i A_2)(\vec{l}_0 \cdot \vec{w}_1 + \vec{b}_1)}$$

$$\frac{\partial L_0}{\partial w_{1ij}} = \boxed{1} \sum_m (\partial_m A_{l_0})(\vec{l}_0 \cdot \vec{w}_1 + \vec{b}_1) \frac{\partial}{\partial w_{1ij}} \left[(\vec{l}_0 \cdot \vec{w}_1)_m + \cancel{\vec{b}_{1m}} \right]$$

$$= \boxed{1} \sum_m (\partial_m A_{l_0})(\vec{l}_0 \cdot \vec{w}_1 + \vec{b}_1) \frac{\partial}{\partial w_{1ij}} \left(\sum l_{0n} w_{nm} \right)$$

$$\begin{cases} 1 & i = n \ \& \ j = m \\ 0 & \text{other} \end{cases}$$

$$= \boxed{1} (\partial_j A_{2d})(\cdots) \, l_{0i}$$

$$\boxed{\frac{\partial L}{\partial w_{1ij}} = -2 \, l_{0i} \, (\partial_j A_2)(\vec{l}_0 \cdot \vec{w}_1 + \vec{b}_1) \cdot (y - y_{pred})}$$

$$\frac{\partial L_0}{\partial b_{0,i}} = \boxed{\boxed{1} \sum_m (\partial_m A_{l_0})(\vec{l}_0 \cdot \vec{w}_1 + \vec{b}_1)} \frac{\partial}{\partial b_{0,i}} \left[(\vec{l}_0 \cdot \vec{w}_1)_m + \cancel{\vec{b}_{1m}} \right]$$

$$= \boxed{2} \left[\left(\frac{\partial}{\partial b_{0i}} \vec{l}_0 \right) \cdot \vec{w}_1 \right]_m$$

$$\frac{\partial l_{0j}}{\partial b_{0i}} = \frac{\partial}{\partial b_{0i}} A_{1,j}(x + \vec{b}_0)$$

$$= \sum_n (\partial_n A_{1,j})(x + \vec{b}_0) \frac{\partial}{\partial b_{0i}} (\cancel{x_n} + b_{0n})$$

$$\begin{cases} 1 & i = n \\ 0 & i \neq n \end{cases}$$

$$= (\partial_i A_{1,j})(x + \vec{b}_0)$$

$$\frac{\partial L_0}{\partial b_{0i}} = |1| \sum_{n} (\partial_m A_{2\,0})(\vec{l_0} \cdot \vec{w} + b_1)\left[\sum_j (\partial_i A_{1,j})(x+\vec{b_0})\, W_{jm}\right]$$

$$= -2(y_0 - y_{pred_0})(\partial_i A_1)\,\underline{(\vec{x}+\vec{b_0})} \cdot \vec{W_1} \cdot \partial A_{2\,0}\,\underline{(\vec{l_0} \cdot \vec{w_1} + b_1)}$$

$$= -2\,(\partial_i A_1) \cdot W_1 \cdot \partial A_2 \cdot (y - y_{pred})$$

$$\frac{d}{dx}\tan(x) = \operatorname{sech}^2(x) = \frac{1}{\cosh^2(x)}$$

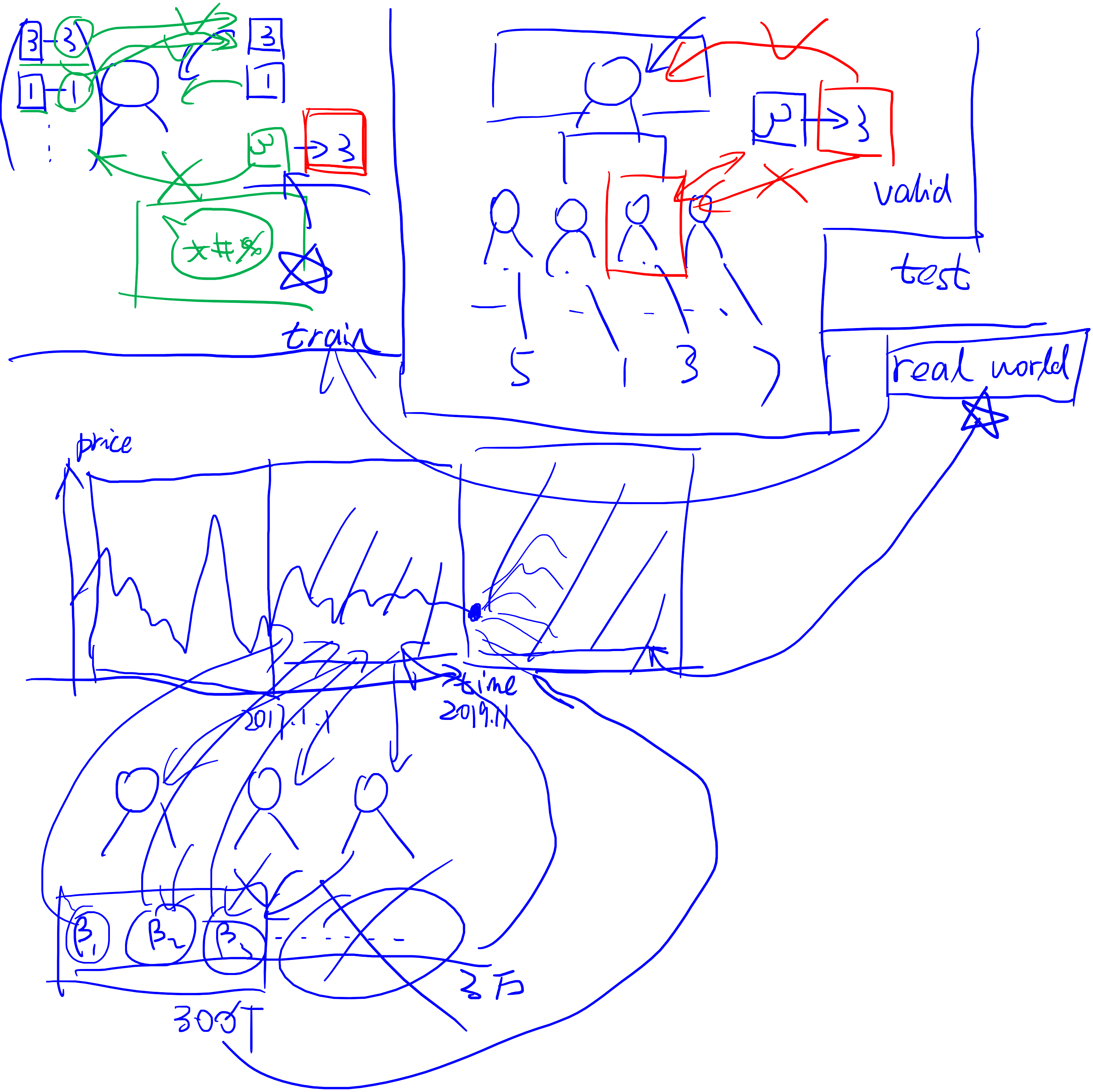
$$\tanh(x) \longrightarrow \tanh(\vec{x}) \to (\tanh(x_1), \tanh(x_2), \cdots \tanh(x_n))$$

$$\frac{d}{dx_i}\tanh_j(x) \longrightarrow \begin{cases} \frac{1}{\cosh^2(x_i)} & i=j \\ 0 & i \neq j \end{cases} = \begin{pmatrix} \frac{1}{\cosh^2(x_1)} & & \\ & \frac{1}{\cosh^2(x_2)} & \\ & & \ddots & \frac{1}{\cosh(x_n)} \end{pmatrix}$$

$$\operatorname{Softmax}_j(\vec{x}) = \frac{e^{x_j}}{\sum_m e^{x_m}}$$

$$\frac{d}{dx_i}\operatorname{Softmax}_j(\vec{x}) = \frac{\delta_{ij} e^{x_j}\left(\sum_m e^{x_m}\right) - e^{x_j} e^{x_i}}{\left(\sum_m e^{x_m}\right)^2}$$

$$= -\operatorname{Softmax}_i \cdot \operatorname{Softmax}_j + \delta_{ij} \cdot \operatorname{Softmax}_i$$



3 → 3
1 → 1
3 → 3
x#%
train
3 → 3
valid
test
5
1
3
7
real world
price
time
2017.1.1
2019.11
β1
β2
β3
300T
3万

$O_1 = (0.1, 0.1, \dots\dots 0.1)$

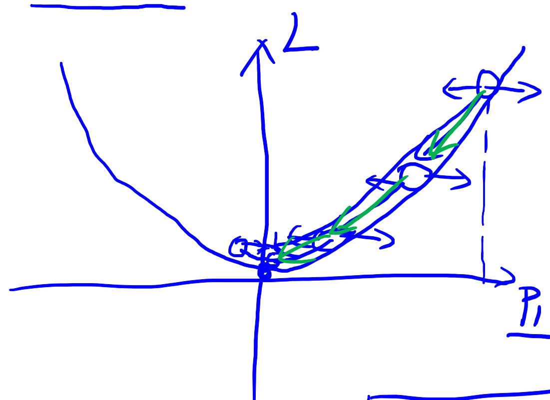
$\underline{O_2} = (0.01, 0.01, \dots\dots, \underline{0.91})$

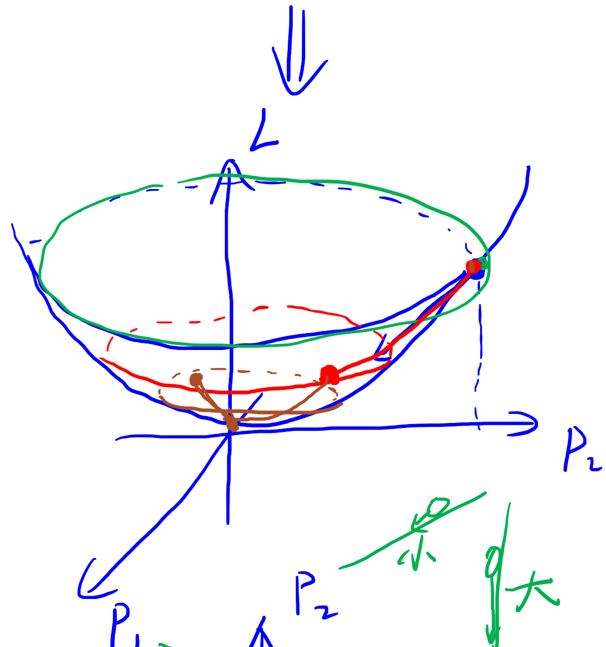
$O_3 = (0, 0, \dots\dots, \underline{1})$

$$L = \underline{(O_3 - O_2)^2} = \underline{(0.01)^2 + \dots\dots (1-0.91)^2}$$

loss function

$\underline{L(P_i)}$ P_i : parameters



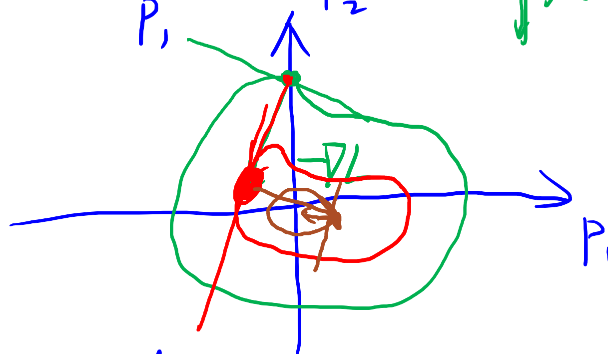


$$\boxed{P_i \to P_i - k \nabla L}$$

$$\nabla f(\vec{x}) = \left(\frac{\partial}{\partial x_i} f(\vec{x})\right)$$

$$\boxed{\frac{d}{dx} f(x)}$$

$$\partial_i f(\vec{x}) = \frac{\partial}{\partial x_i} f$$



$$\frac{d}{dx}(f+g) = \frac{d}{dx} f + \frac{d}{dx} g$$

$$\frac{d}{dx}(af) = a \frac{d}{dx} f$$

$$\frac{d}{dx}(fg) = \left(\frac{d}{dx} f\right) g + f\left(\frac{d}{dx} g\right)$$

$$\frac{d}{dx}\left(\frac{f}{g}\right) = \frac{\left(\frac{d}{dx} f\right) g - \left(\frac{d}{dx} g\right) f}{g^2}$$

$$\frac{d}{dx} f(g(x)) = \left(\frac{d}{dx} f\right)(g(x))\left(\frac{d}{dx} g\right)$$

$$\frac{\partial}{\partial x} f(x,y) = \frac{d}{dx} f(x; y = y_0)$$

$$\frac{d}{dx} f(g_1(x), g_2(x)) = (\partial_1 f)\frac{d}{dx} g_1(x) + (\partial_2 f)\frac{d}{dx} g_2(x)$$

$$L = (y_{(10)} - y_{pred\,(10)})^2$$

$$= \underbrace{\boxed{(y_0 - y_{pred,0})^2}}_{L_0} + \underbrace{(y_1 - y_{pred,1})^2}_{L_1} + \cdots \quad \cdots \quad L_9$$

$$= (y - A_2(\vec{l_0} \cdot \overleftrightarrow{w_1} + \vec{b_1}))^2$$

$$= (y - A_2(A_1(x + \vec{b_0}) \cdot \overleftrightarrow{w_1} + b_1))^2$$

$$\left(\frac{\partial L}{\partial b_{1i}}, \frac{\partial L}{\partial b_{0i}}, \frac{\partial L}{\partial w_{ij}} \right) = ?$$

$$\frac{\partial L}{\partial b_{1i}} = \frac{\partial L_0}{\partial b_{1i}} + \cdots$$

$$\frac{\partial L_0}{\partial b_{1i}} = (y_0 - y_{pred\,0})^2$$

$$= 2(y_0 - y_{pred\,0}) \frac{\partial}{\partial b_{1,i}} (\cancel{y_0} - y_{pred\,0})$$

$$= \underbrace{\boxed{-2(y_0 - y_{pred\,0})}}_{1} \frac{\partial}{\partial b_{1i}} y_{pred\,0}$$

$$= \boxed{1} \frac{\partial}{\partial b_{1,i}} A_{2,0} (\vec{l_0} \cdot \overleftrightarrow{w_1} + \vec{b_1})$$

$$= \boxed{1} \sum_m (\partial_m A_{2,0}) \boxed{\frac{\partial}{\partial b_i} (\cancel{l_0 \cdot w_{1m}} + \vec{b}_{1,m})} = \begin{cases} 1 & i = m \\ 0 & i \neq m \end{cases}$$

$$= \boxed{1} (\partial_i A_{2,0})(\vec{l_0} \cdot \overleftrightarrow{w_1} + \vec{b_1})$$

$$\frac{\partial L_0}{\partial b_{1,i}} = -2(y_0 - y_{pred_0})(\partial_i A_{2,0})(\vec{l_0}\cdot\hat{w} + \vec{b}_1)$$

$$\frac{\partial L_1}{\partial b_{1,i}} = -2(y_1 - y_{pred_1})(\partial_i A_{2,1})(\vec{l_0}\cdot\hat{w} + \vec{b}_1)$$

$$- \; - \; -$$

$$\boxed{\frac{\partial L}{\partial b_{1,i}} = -2(\vec{y} - \vec{y}_{pred})\cdot \partial_i A_2(\vec{l_0}\cdot\vec{w} + \vec{b}_1)}$$

$$\frac{\partial L_0}{\partial w_{1,ij}} = \boxed{1}\sum_m (\partial_m A_{2,0}) \frac{\partial}{\partial w_{1,ij}}(l_0\cdot w_{1m} + \cancel{\vec{b}_{1,m}})$$

$$\frac{\partial}{\partial w_{1,ij}}(\vec{l_0}\cdot\vec{w}_1)_m$$

$$\frac{\partial}{\partial w_{1,ij}}\sum_n l_{0,n}\boxed{w_{1nm}} \rightarrow \begin{cases} 1 & i=n \;\&\; j=m \\ 0 & \text{other} \end{cases}$$

$$= \boxed{1}\,\partial_j A_{2,0}\, l_{0,i}$$

$$\frac{\partial L_0}{\partial w_{1,ij}} = -2(y_0 - y_{pred_0})\;\overset{\text{outer}}{\underline{l_{0i}\;\partial_j A_{2,0}}}$$

$$\boxed{\frac{\partial L}{\partial w_{1,ij}} = -2\, l_{0i}\, \partial_j A_2\cdot(y - y_{pred})}$$

$$\frac{\partial L_0}{\partial b_{0,i}} = \underbrace{\boxed{1}\sum_m (\partial_m A_{2,0})}_{2}\frac{\partial}{\partial b_{0,i}}((l_0\cdot w_1)_m + \cancel{\vec{b}_{1,m}})$$

$$= \boxed{2}\left(\left(\frac{\partial}{\partial b_{0i}} l_0\right) \cdot w_1\right)_m$$

$$\frac{\partial}{\partial b_{0i}} A_1(x + b_0)$$

$$= \sum_n (\partial_n A_1) \frac{\partial}{\partial b_{0i}} (\cancel{x_n} + b_{0,n}) \rightarrow \begin{cases} 1 & i = n \\ 0 & i \neq n \end{cases}$$

$$= \partial_i A_1 (x + b_0)$$

$$\frac{\partial L_0}{\partial b_{0i}} = \boxed{1}\left[\sum_m \partial_m A_{20} (\partial_i A_1 \cdot w)_m\right]$$

$$= \boxed{1} \; (\partial_i A_1 \cdot \overleftrightarrow{w_1} \cdot \vec{\partial} A_{2,0})$$

$$\frac{\partial L_0}{\partial b_{0,i}} = -2(y_0 - y_{pred_0})(\partial_i A_1 \cdot \overleftrightarrow{w_1} \cdot \vec{\partial} A_{20})$$

$$\boxed{\frac{\partial L}{\partial b_{0,i}} = -2 \, \boxed{\partial_i A_1} \cdot \overleftrightarrow{w_1} \cdot \vec{\partial} A_2 \cdot (\vec{y} - \vec{y}_{pred})}$$

$$\boxed{\frac{\partial L}{\partial b_{1i}} = -2 \, \boxed{\partial_i A_2} (\vec{l_0} \cdot \overleftrightarrow{w} + \vec{b_1}) \cdot (\vec{y} - \vec{y}_{pred})}$$

$$\boxed{\frac{\partial L}{\partial w_{1ij}} = -2 \, l_{0i} \, \partial_j A_2 \cdot (y - y_{pred})}$$

$$A_1 = \tanh = \frac{e^{x} - e^{-x}}{e^{x} + \sigma^{x}}$$

$$\frac{d}{da} A_1 = \frac{1}{\cosh^2 x}$$

$$A_1(\vec{x}) = (\underline{\tanh x_0}, \tanh x_1, \cdots)$$

$$\boxed{\underline{\partial_i A_1}} = \begin{pmatrix} \frac{1}{\cosh^2 x_0} & 0 & \cdots & & \\ 0 & \frac{1}{\cosh^2 x_1} & & & \\ 0 & & \ddots & 0 & \\ 0 & & & \ddots & \\ 0 & 0 & & & \ddots \end{pmatrix}$$

$$\mathrm{Softmax}_i(x) = \frac{e^{x_i}}{\sum_m e^{x_m}}$$

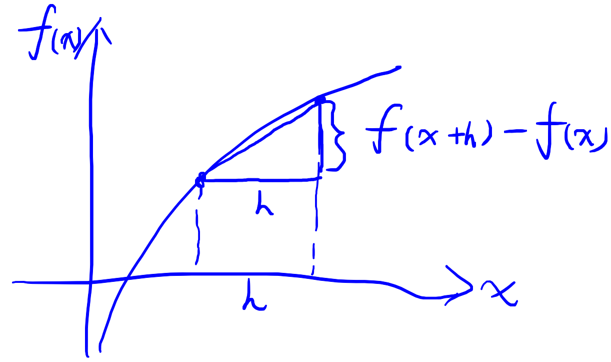
$$\frac{\partial}{\partial x_i} \mathrm{Softmax}_j(x) = \frac{\delta_{ij} e^{x_i} (\sum_m e^{x_m}) - e^{x_i} \cdot e^{x_j}}{(\sum_m e^{x_m})^2}$$

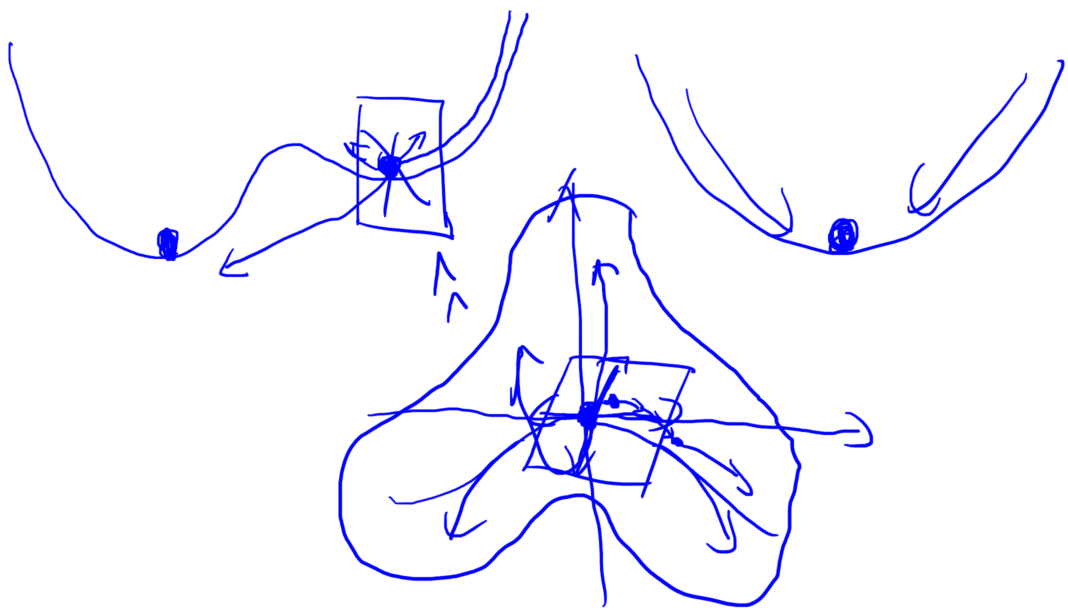
$$\boxed{\underline{\delta_{ij}} \, e^{x_i} = \begin{cases} e^{x_i} & i = j \\ 0 & i \neq j \end{cases}}$$

$$\partial_i A_2(x) = \delta_{ij} \, \mathrm{Softmax}_i(x) - \mathrm{Softmax}_i(x) \, \mathrm{Softmax}_j(x)$$

$$\frac{d}{dx}f(x) = \lim_{\substack{h \to 0 \\ 0.0001}} \frac{f(x+h) - f(x)}{h}$$

$$\boxed{\frac{\Delta f(x)}{\Delta x}} \approx \boxed{\frac{d}{dx} f(x)}$$





$$\vec{l}_{n,0} = \vec{A}_n(\vec{l}_{n-1,0} \cdot \vec{w}_n + \vec{b}_n)$$

$$\frac{d}{dx} f(g(x)) = f' \times g'(x)$$

$$\frac{d}{dx} f_1(f_2(f_3 \cdots f_n(x)) \cdots) = \underbrace{f_1'}_{1} \times \underbrace{f_2'}_{2} \times \underbrace{f_3'}_{3} \cdots f_n'(x)$$

$$\frac{d}{db_{0,i}} \vec{l}_{n,0} = \sum_k (\partial_k \vec{A}_n) \frac{\partial}{\partial b_{0i}} \left[(l_{n-1\,0} \cdot w_n)_k + b_{nk} \right]$$

$$= \sum_k (\partial_k \vec{A}_n) \left[\left(\frac{\partial}{\partial b_{0i}} l_{n-1\,0} \right) \cdot w_n \right]_k$$

$$\frac{d}{db_{0i}} \vec{l}_{n,0} = \left(\frac{\partial}{\partial b_{0i}} l_{n-1,0} \right) \cdot w_n \cdot \vec{\partial} \vec{A}_n$$