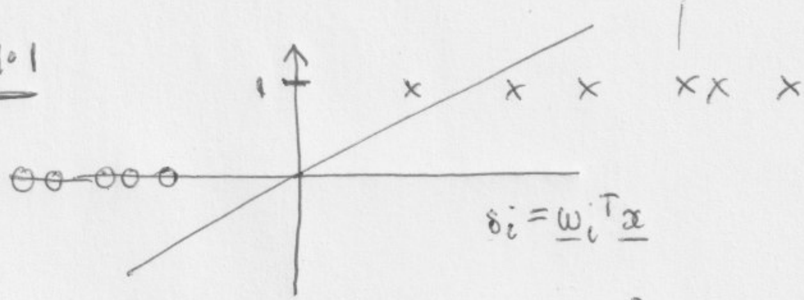
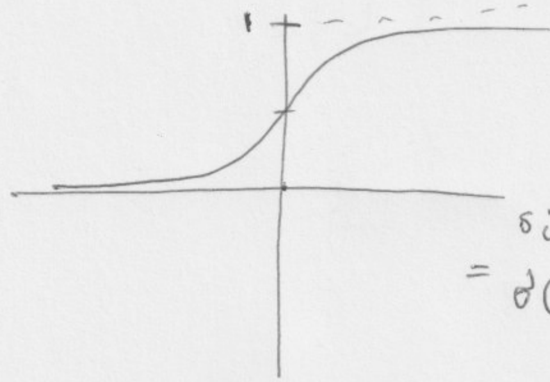
19.1



$s_i = \underline{w}_i^T \underline{x}$

$|w_i^T \underline{x} - \underline{t}|^2$



s_i

$= \sigma(\underline{w}_i^T \underline{x})$

$\sigma(s_i) = \frac{1}{e^{-s_i}+1}$ logistic

$\underline{s} = W^T \underline{x}$ $\qquad$ $\underline{\sigma}(\underline{s}) = \frac{e^{\underline{s}}}{\sum_i e^{s_i}}$ softmax

19.2

softmax predictor $\underline{\sigma}(\underline{s}) = \frac{e^{\underline{s}}}{\sum_i e^{s_i}}$ $\qquad$ $e^{\underline{s}} = \begin{pmatrix} e^{s_0} \\ e^{s_1} \\ e^{s_2} \\ \vdots \end{pmatrix}$

cross entropy loss $\quad e = -\sum_i t_i \log \underline{\sigma}(\underline{s})_i$

it can be shown $\quad \frac{\partial e}{\partial \underline{s}} = \underbrace{\underline{\sigma}(s)}_{= h} - \underline{t}$

linear predictor $\underline{s} = \underline{W}^T x$, $\quad$ L2 loss $|h-t|^2 = e$

$$\frac{\partial e}{\partial \underline{s}} = h - t$$

19.3

2 layer network + relu

$$h = W_2 \underbrace{\max(0, W_1 x)}_{\text{relu}}$$

with no activation $\quad h = \underbrace{W_2 W_1}_{W_3} x \quad \times$ no use.

19.4

$$h = w_2 \underbrace{\max(0, w_1 x)}_{\underline{a}}$$

$$\begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} = \underline{a} = \max(w_1 x, 0) = \max\left(\underbrace{\begin{pmatrix} w_{00} & w_{01} & w_{02} \\ w_{10} & w_{11} & w_{12} \\ w_{20} & w_{21} & w_{22} \\ w_{30} & w_{31} & w_{32} \end{pmatrix}}_{= \, w_1} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix}, 0 \right)$$

19.5

$$y = w_2(\max(0, \underline{w}_1 x + b_1)) + b_2 \qquad L = |y - t|^2$$

$$\frac{\partial L}{\partial w_1} = 2(y-t)\frac{\partial}{\partial w_1}(y-t) = 2(y-t)\frac{\partial y}{\partial w_1}$$

$$= 2(y-t)\, w_2 \frac{\partial}{\partial w_1} \max(0, w_1 x + b_1)$$

$$= 2(y-t)\, w_2 \, \mathbb{I}_{w_1 x + b_1 > 0} \frac{\partial}{\partial w_1} w_1 x + b_1$$

$$\mathbb{I}_x = \begin{cases} 1, & x \text{ is true} \\ 0, & \text{oth.} \end{cases}$$

$$= \underline{2(y-t)\, w_2 \, \mathbb{I}_{w_1 x + b_1 > 0} \, x}$$

19.6 Forward pass: compute all values + activations

input → output

$i_1 = w_1 x + b_1 = 4$

$a = \max(0, i_1) = 4$

$y = w_2 a + b_2 = 3$

$i_2 = y - t = 2$

$L = i_2^2 = 4$

Backward pass: compute $\frac{\partial L}{\partial \cdot}$ applying chain rule

output → input

$L = i_2^2 \quad \frac{\partial L}{\partial i_2} = 2 i_2 = 4$

$i_2 = y - t \quad \frac{\partial L}{\partial y} = \frac{\partial L}{\partial i_2} \frac{\partial i_2}{\partial y} = 4$

$y = a w_2 + b_2 \quad \frac{\partial L}{\partial b_2} = \frac{\partial L}{\partial y} \underbrace{\frac{\partial y}{\partial b_2}}_{=1} = 4 \qquad \frac{\partial L}{\partial w_2} = \underbrace{\frac{\partial L}{\partial y}}_{=4} \underbrace{\frac{\partial y}{\partial w_2}}_{=a=4} = 16$

$\frac{\partial L}{\partial a} = \frac{\partial L}{\partial y} \underbrace{\frac{\partial y}{\partial a}}_{=w_2=2} = 8$

$a = \max(0, i_1) \quad \frac{\partial L}{\partial i_1} = \frac{\partial L}{\partial a} \frac{\partial a}{\partial i_1} \rightarrow$

19.7 $\frac{\partial L}{\partial a_0} = \frac{\partial L}{\partial h_0} \frac{\partial h_0}{\partial a_0} + \frac{\partial L}{\partial h_1} \frac{\partial h_1}{\partial a_0}$

in general $\frac{\partial L}{\partial a_0} = \sum_{i=1}^{n} \frac{\partial L}{\partial h_i} \frac{\partial h_i}{\partial a_0}$