linear regression

$$h_\theta(x) = x \cdot \theta$$

$$L = \frac{1}{2}(h_\theta(x) - y)^2$$

$$\frac{\partial L}{\partial \theta} = x^T(h_\theta(x) - y)$$

$x\ (n, m)$

$\cancel{\theta\ (1, m)}$ $\theta\ (m, 1)$

$y\ (n, 1)$

logistic regression (mse)

$$h_\theta(x) = \text{sigmoid}(x\theta)$$

$$L = \frac{1}{2}(h_\theta(x) - y)^2$$

$$\frac{\partial L}{\partial \theta} = x^T \cdot \left[(h_\theta(x) - y) \cdot h_\theta(x) \cdot (1 - h_\theta(x))\right]$$

logistic regression (cross entropy)

$$h_\theta(x) = \text{sigmoid}(x\theta)$$

$$L = \prod_{i=1}^{n} h_\theta(x^i)^{y^i} \cdot (1 - h_\theta(x^i))^{(1-y^i)}$$

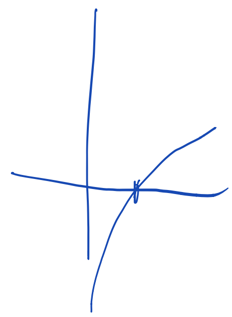
$-\log \rightarrow$ $$\log L = \sum_{i=1}^{n} \left(y^i \log h_\theta(x^i) + (1-y^i)\log(1 - h_\theta(x^i)) \right)$$

$\rightarrow$ $$\frac{\partial \log L}{\partial \theta} = \sum_{i=1}^{n} \left[\frac{y^i}{h_\theta(x^i)} h_\theta'(x^i) - \frac{1-y^i}{1-h_\theta(x^i)} h_\theta'(x^i) \right]$$

$$= \sum_{i=1}^{n} \left[\frac{y^i}{h_\theta(x^i)} - \frac{1-y^i}{1-h_\theta(x^i)} \right] h_\theta(x^i)(1 - h_\theta(x^i)) \cdot x^i$$

$$= \sum_{i=1}^{n} \left(y^i(1 - h_\theta(x^i)) - (1-y^i) h_\theta(x^i) \right) \cdot x^i$$

$$= x^{iT} \cdot \sum_{i=1}^{n} \left(y^i - h_\theta(x^i) \right)$$



perceptron

$$h_\theta(x) = \begin{cases} 1 & x\theta \geq 0 \\ 0 & x\theta < 0 \end{cases}$$

y^i	$h_\theta(x^i)$
0	1
1	0

$$L = \sum_{x \in M_0} x\theta - \sum_{x \in M_1} x\theta$$

$$\cancel{= \sum_{i=1}^{n} \left\{ (1-y^i)x\theta - y^i x\theta \right\}}$$

$$\cancel{= \sum_{i=1}^{n} (1-2y^i)x\theta}$$

$$\cancel{\frac{\partial L}{\partial \theta} = \sum_{i=1}^{1} (1-2y^i)x}$$

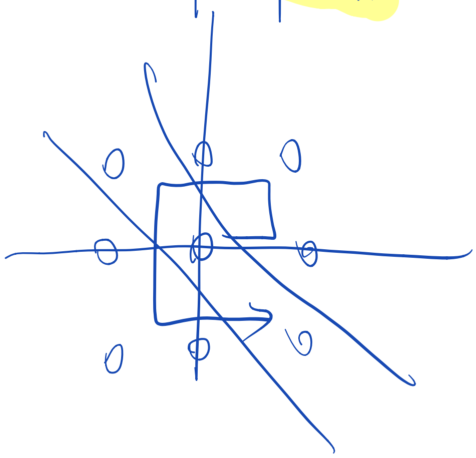
$$= \sum_{i=1}^{n} \left[(1-y^i)h_\theta(x^i) \cdot x^i\theta - y^i(1-h_\theta(x^i))x^i\theta \right]$$

$$= \sum_{i=1}^{n} \left[h_\theta(x^i) - y^i \right] x^i\theta$$

$(n,1)$ (n,m) (t,m)

$$\frac{\partial L}{\partial \theta} = x^{iT} \cdot \sum_{i=1}^{n} (h_\theta(x^i) - y^i)$$

multi-class perceptron



$$h_\theta(x^i) = \max_{j=1}^{c} x^i\theta_j$$

$$L = \sum_{i=1}^{n} \left(\max_{j=1}^{c} x^i\theta_j - x^i\theta_y \right)$$

$$\frac{\partial L}{\partial \theta_v} = \begin{cases} v \neq j,\ v \neq y & \Rightarrow 0 \\ v = j,\ v = y & \Rightarrow 0 \\ v = j,\ v \neq y & \Rightarrow x^i \\ v \neq j,\ v = y & \Rightarrow -x^i \end{cases}$$

Softmax

$$S(x^i\theta_j^T) = \frac{e^{x^i\theta_j}}{\sum_{k=1}^{c} e^{x^i\theta_k}}$$

hypothesis: $h_\theta(x^i) = \prod_{j=1}^{c} S(x^i\theta_j)^{1\{y^i=j\}}$

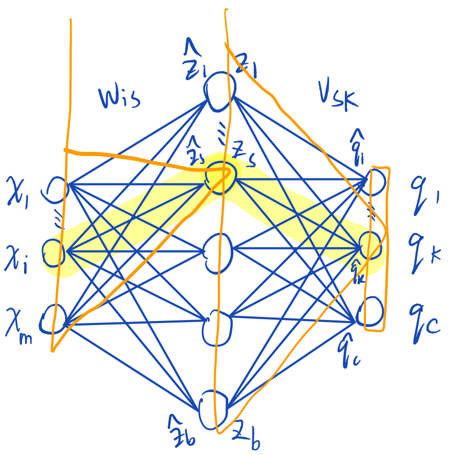
Likelihood $L = \prod_{i=1}^{n}\prod_{j=1}^{c} S(x^i\theta)^{1\{y^i=j\}}$

$$\log L = \sum_{i=1}^{n}\sum_{j=1}^{c} 1\{y^i=j\}\log S(x^i\theta_j)$$

$$\frac{\partial \log L}{\partial \theta_v} = \begin{cases} v=j := \sum_{i=1}^{n} \frac{\left(\sum_{k=1}^{c} e^{x^i\theta_k}\right)}{e^{x^i\theta_v}} \cdot \frac{e^{x^i\theta_v}\cdot x^i \cdot \left(\sum_{k=1}^{c} e^{x^i\theta_k}\right) - e^{x^i\theta_v}\cdot e^{x^i\theta_v} x^i}{\left(\sum_{k=1}^{c} e^{x^i\theta_k}\right)^2} \\ \Downarrow \text{相等} \\ y^i \quad = \sum_{i=1}^{n} x^{iT}\cdot\left(1 - S(x^i\theta_v^T)\right) \\ v \neq j := \sum_{i=1}^{n} \frac{\left(\sum_{k=1}^{c} e^{x^i\theta_k}\right)}{e^{x^i\theta_j}} \cdot \frac{0 - e^{x^i\theta_j}\cdot e^{x^i\theta_v} x^i}{\left(\sum_{k=1}^{c} e^{x^i\theta_k}\right)^2} \\ = \sum_{i=1}^{n} x_i^T \cdot \left(- S(x^i\theta_v^T)\right) \end{cases}$$

$\rightarrow$ 加负号，最小化

ANN



$$\hat{z}_s = \sum_{i=1}^{m} x_i W_{is}$$

$$z_s = \text{Sigmoid}(\hat{z}_s) = \frac{1}{1 + e^{-\left(\sum_{i=1}^{m} x_i W_{is}\right)}}$$

$$\hat{q}_k = \sum_{s=1}^{b} z_s V_{sk}$$

$$q_k = \text{Softmax}(\hat{q}_k) = \frac{e^{\hat{q}_k}}{\sum_{h=1}^{c} e^{\hat{q}_h}} = \frac{e^{\sum_{s=1}^{b} z_s V_{sk}}}{\sum_{h=1}^{c} e^{\sum_{s=1}^{b} z_s V_{sh}}}$$

$$L = -\sum_{k=1}^{c} 1\{k=y\}\log q_k = -\log q_y$$

$$\frac{\partial L}{\partial V_{sk}} = \frac{\partial L}{\partial q_y}\cdot\frac{\partial q_y}{\partial \hat{q}_k}\cdot\frac{\partial \hat{q}_k}{\partial V_{sk}}$$

$$\frac{\partial L}{\partial V_{sk}} = -\frac{1}{q_y} \cdot \begin{cases} k=y \cdot \frac{e^{\hat{q}_y}\cdot\left(\sum_{h=1}^{c} e^{\hat{q}_h}\right) - e^{\hat{q}_y}\cdot e^{\hat{q}_y}}{\left(\sum_{h=1}^{c} e^{\hat{q}_h}\right)^2} \cdot z_s = (q_y - 1) z_s \\ k \neq y \cdot \frac{-e^{\hat{q}_y}\cdot e^{\hat{q}_k}\cdot z_s}{\left(\sum_{h=1}^{c} e^{\hat{q}_h}\right)^2} \cdot z_s = (q_k) z_s \end{cases}$$

$$-\frac{1}{q_y} = -\frac{\left(\sum_{h=1}^{c} e^{\hat{q}_h}\right)}{e^{\hat{q}_y}}$$

$y \rightarrow k=y$

$$\Rightarrow (q_k - y_k) z_s$$

$$\Downarrow$$

$$z^T(q - y)$$

$(5,3)$

$$\frac{\partial L}{\partial w_{is}} = \left(\sum_{k=1}^{C} \frac{\partial L}{\partial q_y} \cdot \frac{\partial q_y}{\partial \hat{q}_k} \right) \cdot \frac{\partial \hat{q}_k}{\partial z_s} \cdot \frac{\partial z_s}{\partial \hat{z}_s} \cdot \frac{\partial \hat{z}_s}{\partial w_{is}}$$

$$= \sum_{k=1}^{C} (q_k - y) \cdot V_{sk} \cdot z_s(1 - z_s) \cdot X_i$$

$$\Downarrow$$

$$\begin{array}{cccc}
C & C & & m \\
\uparrow & \uparrow & & \uparrow \\
(n,3) & (5,3) & (n,5) & (n,3) \\
q-y & V & z(1-z) & x
\end{array}$$

$$\Downarrow$$

$$\frac{x^T \cdot \left((q-y) \cdot V^T \times z(1-z) \right)}{(3,n) \cdot \left((n,3) \cdot (3,5) \times (n,5) \right)}$$