Linear Discrimination

Multiple classes $(K>2)$

$$\mathcal{X} = \{(x_i, y_i)\}_{i=1}^{N}$$

$y_i \in 1, 2, \ldots, K$ reference class

$$\log\left[\frac{p(x \mid y=c)}{p(x \mid y=K)}\right] = \underline{w_c^T \cdot x + w_{c0}^{\bullet}}$$

$$\log\left[\frac{p(y=c \mid x)}{p(y=K \mid x)}\right] = \log\left[\frac{p(x \mid y=c)\, p(y=c)}{p(x \mid y=K)\, p(y=K)}\right]$$

$$= \underbrace{\log\left[\frac{p(x \mid y=c)}{p(x \mid y=K)}\right]}_{w_c^T \cdot x + w_{c0}^{\bullet}} + \underbrace{\log\left[\frac{p(y=c)}{p(y=K)}\right]}_{\text{not a function of } \underline{\underline{x}}}$$

$$\exp\left[\log\left[\frac{p(y=c \mid x)}{p(y=K \mid x)}\right]\right] = \exp\left[w_c^T x + w_{c0}\right]$$

where $w_{c0} = w_{c0}^{\bullet} + \log\left[\frac{p(y=c)}{p(y=K)}\right]$

$$\frac{P(y=c|x)}{P(y=K|x)} = \exp(w_c^T x + w_{c0})$$

$$P(y=1|x) + P(y=2|x) + \dots + \boxed{P(y=K|x)} = 1.$$

$$\underline{P(y=1|x) + P(y=2|x) + \dots P(y=K-1|x)} = 1 - P(y=K|x)$$

$$\sum_{c=1}^{K-1} \frac{P(y=c|x)}{P(y=K|x)} = \frac{1 - P(y=K|x)}{P(y=K|x)} = \sum_{c=1}^{K-1} \exp(w_c^T x + w_{c0})$$

$$P(y=K|x) = ? = \boxed{\frac{1}{1 + \sum_{c=1}^{K-1} \exp(w_c^T x + w_{c0})}}$$

$$\frac{P(y=c|x)}{P(y=K|x)} = \exp(w_c^T . x + w_{c0}) \rightarrow$$

$$P(y=c|x) = \frac{\exp(w_c^T . x + w_{c0})}{1 + \sum_{d=1}^{K-1} \exp(w_d^T . x + w_{d0})}$$

```
sum <- function(a, a) {
}
```

②

$$\theta = \{ \underline{\underline{w_1}}, w_{10}, \underline{w_2}, w_{20}, \ldots\ldots, \underline{\underline{w_{k-1}}}, w_{k-1,0} \}$$

$$p(y=1|x) = \frac{\exp(w_1^T x + w_{10})}{1 + \exp(w_1^T x + w_{10}) + \exp(w_2^T x + w_{20}) + \ldots + \exp(w_{k-1}^T x + w_{k-10})}$$

$$= \cancel{\frac{\exp(w_1^T x + w_{10})}{1 + (k-1) \cdot \exp(w_1^T \cdot x + w_{10})}}$$

```
function (a)
{
    a ← 3
    return a*a
}
```

function (5)
↳ 9

function (7)
↳ 9

$$p(y=c \mid x) = \boxed{\frac{\exp(w_c^T \cdot x + w_{c0})}{\sum_{d=1}^{K} \exp(\underline{w_d}^T \cdot x + \underline{w_{d0}})}} \quad \begin{cases} \geq 0 \\ \underline{\text{sum}} \text{ up to } 1 \end{cases}$$

SOFTMAX

$$w_1^T \cdot x + w_{10} \Rightarrow 2$$
$$w_2^T \cdot x + w_{20} \Rightarrow -2$$
$$\vdots$$
$$w_K^T \cdot x + w_{K0} \Rightarrow 1.$$

$$\underline{p(y=1 \mid x)} = \frac{\exp(2)}{\exp(+2) + \exp(-2) + \ldots + \exp(1)}$$

$$\underline{p(y=2 \mid x)} = \frac{\exp(-2)}{}$$

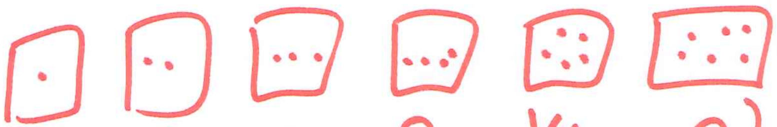
$$\left.\begin{array}{l} p(y=1 \mid x^*) \\ p(y=2 \mid x^*) \\ \vdots \\ p(y=K \mid x^*) \end{array}\right\} \text{maximum of these.}$$

w_1	w_{10} ✓
w_2	$\boxed{w_{20}}$ ✓
$\vdots$	$\vdots$ ✓
w_K	w_{K0}

YES if $c=2$
NO if $c \neq 2$

$$y_i \mid x_i \sim \text{Multinomial}\left(1, \{P(y=c \mid x)\}_{c=1}^{K}\right)$$

$$\ell\left(\{w_c, w_{c0}\}_{c=1}^{K} \mid \mathcal{X}\right) = \prod_{i=1}^{N} \prod_{c=1}^{K} p(y_i = c \mid x_i)^{(y_i = c)}$$



$$\left(5/6,\ 0,\ 0,\ 0,\ 1/6,\ 0\right)$$

$$\underset{(x=1)}{p_1} \cdot \underset{(x=2)}{p_2} \cdot \underset{(x=3)}{p_3} \cdots\cdots$$

0 1

$$p^{x}(1-p)^{1-x}$$

$$p = 0.8$$

$$x = 1$$

$$(0.8)^{1} \cdot (0.2)^{1-1} = 0.8$$

y:

$$p(y=4) = p_1^{(y=1)} \cdot p_2^{(y=2)} \cdots p_6^{(y=6)}$$

$$p(y=1) = 5/6^{\,1} \cdot \boxed{(0)^{0}} \cdot 0^{0} \cdot 0^{0} \cdot (1/6)^{0} \cdot 0^{0} = 5/6$$

$$y = \begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \\ 3 \end{bmatrix} \Rightarrow Y = \begin{array}{c} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ \begin{matrix} \uparrow & \uparrow & \uparrow \\ 1 & 2 & 3 \end{matrix} \end{array}$$

$$\text{Error}\left(\{w_c, w_{c0}\}_{c=1}^{K} \mid \mathcal{X}\right) = -\sum_{i=1}^{N} \sum_{c=1}^{K} y_{ic} \log \left[\underbrace{P(y_i = c \mid x_i)}_{\hat{y}_{ic}}\right]$$

$$\hat{Y} = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.0 & 0.9 & 0.1 \\ 0.2 & 0.8 & 0.0 \\ 0.9 & 0.1 & 0.0 \\ 0 & 0.3 & 0.7 \end{bmatrix}$$

$$\text{Error}(\{w_c, w_{c0}\}_{c=1}^{K} \mid \mathcal{X}) = -\sum_{i=1}^{N}\sum_{c=1}^{K} y_{ic}.\log(\hat{y}_{ic}) \quad \text{where}$$

Exercise #6

$$\boxed{\frac{\partial \text{Error}}{\partial w_c} = ? \quad \frac{\partial \text{Error}}{\partial w_{c0}} = ?}$$

$$\hat{y}_{ic} = \frac{\exp(w_c^T.x_i + w_{c0})}{\sum_{d=1}^{K}\exp(w_d^T x_i + w_{d0})}$$

$$-\eta . \frac{\partial}{\partial}$$

$$\Delta w_d = \eta \sum_{i=1}^{N}\sum_{c=1}^{K} \cancel{\frac{y_{ic}}{\hat{y}_{ic}} . \hat{y}_{ic}} . (\overbrace{\delta_{cd}}^{1(c=d)} - \hat{y}_{id}) x_i$$

$$= \boxed{\eta \sum_{i=1}^{N} (y_{id} - \hat{y}_{id}).x_i}$$

$$\Delta w_{d0} = \eta \sum_{i=1}^{N} (y_{id} - \hat{y}_{id})$$

$$\sum_{c=1}^{K} y_{ic} (\delta_{cd} - \hat{y}_{id}).x_i$$

$$\underline{\sum_{c=1}^{K} y_{ic}.\delta_{cd}.x_i} - \sum_{c=1}^{K} y_{ic}.[\hat{y}_{id}.x_i]$$

$$y_{id}.x_i - \underline{\hat{y}_{id}.x_i}$$

$$\cancel{y_{id}(x_i}$$

$$= \underline{(y_{id} - \hat{y}_{id}).x_i}$$

$$f(x,y) = 2x^2y^3 + 5xy^2 + 5x^2 + 3y^2$$

$$\frac{\partial f(x,y)}{\partial y} = \frac{\partial 2x^2y^3}{\partial y} + \frac{\partial 5xy^2}{\partial y} + \frac{\partial 5x^2}{\partial y} + \frac{3y^2}{\partial y}$$

$$\partial \underbrace{\text{(there is a "y" term)}}_{1} \cdot \frac{\partial \_\_}{\partial y}$$

0

---

1. initrelize $\{w11, w10, \ldots, wK, wK0\}$ randomly.
   ↳ uniform(-0.001, 0.001)
2. calculate gradients
3. Update $\{w1, w10, \ldots, wK, wK0\}$
4. go to step 2 if there is a change in the parameters.