Linear Discrimination

Multiple Classes $(K>2)$

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

$$x_i \in \mathbb{R}^D \qquad y_i \in \{1, 2, \ldots, K\}$$

reference class $\leftarrow K$

$$\log\left[\frac{p(x|y=c)}{p(x|y=K)}\right] = \underline{w_c^T . x + \tilde{w}_{c0}}$$

$$\exp\left[\underbrace{\log\left[\frac{P(y=c|x)}{P(y=K|x)}\right]}\right] = \log\left[\frac{p(x|y=c)\,P(y=c)}{p(x|y=K).P(y=K)}\right]$$

$$= \underbrace{\log\left[\frac{p(x|y=c)}{p(x|y=K)}\right]}_{w_c^T.x+\tilde{w}_{c0}} + \underbrace{\log\left[\frac{P(y=c)}{P(y=K)}\right]}_{\text{this is not a function of } x.}$$

$$\Downarrow$$

$$\exp\left[w_c^T.x + w_{c0}\right]$$

$$w_{c0} = \tilde{w}_{c0} + \log\left[\frac{P(y=c)}{P(y=K)}\right]$$

$$= w_c^T.x + w_{c0}$$

i) $P(y=c|x) \geq 0$ ✓

ii) $\sum_{c=1}^{K} P(y=c|x) = 1$ ✓

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

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

$$\frac{1}{P(y=K|x)} \quad P(y=1|x) + P(y=2|x) + \cdots + P(y=K-1|x) = 1 - P(y=K|x)$$

$$\underbrace{\sum_{c=1}^{K-1} \frac{P(y=c|x)}{P(y=K|x)}} = \boxed{\frac{1-P(y=K|x)}{P(y=K|x)} = \sum_{c=1}^{K-1}\left[\exp(w_c^T.x+w_{c0})\right]}$$

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

$$\theta = \{\underset{D\times 1}{w_1}, \underset{1\times 1}{w_{10}}, \underset{D\times 1}{w_2}, \underset{1\times 1}{w_{20}}, \ldots, \underset{D\times 1}{w_{(K-1)}}, \underset{1\times 1}{w_{(K-1)0}}\} \rightarrow (K-1)(D+1)$$

Result:

$$\underset{1,2,\ldots,K-1}{P(y=c|x)} = \frac{\exp(w_c^T.x+w_{c0})}{1+\sum_{d=1}^{K-1}\exp(w_d^T.x+w_{d0})}$$

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

$$\underset{\geq 0}{P(y=1\,|\,x)} = \frac{\exp\left[w_1^T . x + w_{10}\right] \quad \geq 0}{1 + \underset{\geq 0}{\exp\left[w_1^T . x + w_{10}\right]} + \cdots + \underset{\geq 0}{\exp\left[w_{(k-1)}^T . x + w_{(k-1)0}\right]}}$$

$$\vdots \qquad\qquad\qquad\qquad \vdots$$

$$\underset{\geq 0}{P(y=k-1\,|\,x)} = \frac{\exp\left[w_{(k-1)}^T\, x + w_{(k-1)0}\right]}{1 + \exp\left[w_1^T . x + w_{10}\right] + \cdots + \exp\left[w_{(k-1)}^T . x + w_{(k-1)0}\right]}$$

$$\underset{\geq 0}{P(y=k\,|\,x)} = \frac{1}{1 + \exp\left[w_1^T . x + w_{10}\right] + \cdots + \exp\left[w_{(k-1)}^T . x + w_{(k-1)0}\right]}$$

$$P(y=c \mid x) = \frac{\exp[w_c^T . x + w_{c0}]}{\sum_{d=1}^{K} \exp[w_d^T . x + w_{d0}]} \quad \Big\} \text{ SOFTMAX function}$$

$c \to 1, 2, \ldots, K$

x^* is a datapoint

$$\left.\begin{aligned} w_1^T . x^* + w_{10} &= +2 \\ w_2^T . x^* + w_{20} &= -2 \\ w_3^T . x^* + w_{30} &= +1 \end{aligned}\right\} \text{ pick the maximum one}$$

$$P(y=1 \mid x) = \frac{\exp(2)}{\exp(2) + \exp(-2) + \exp(1)} \cong 0.7214 \qquad +20 \Rightarrow 0.9999$$

$$P(y=2 \mid x) = \frac{\exp(-2)}{\exp(2) + \exp(-2) + \exp(1)} \cong 0.0132 \qquad -20 \cong 0.0000$$

$$P(y=3 \mid x) = \frac{\exp(1)}{\exp(2) + \exp(-2) + \exp(1)} \cong 0.2654 \qquad +10 \cong 0.0000$$

$$w_1^T \cdot x^* + w_{10} = 1000$$

$$w_2^T \cdot x^* + w_{20} = 2000$$

$$w_3^T \cdot x^* + w_{30} = 3000$$

$$p(y=1|x) = \frac{\overset{Inf}{\exp(1000)}}{\underset{Inf}{\exp(1000)} + \underset{Inf}{\exp(2000)} + \underset{Inf}{\exp(3000)}}$$

$$\max(1000, 2000, 3000) = +3000$$

$$= \frac{Inf}{Inf}$$

$$\boxed{\exp(a)/\exp(b) = \exp(a-b)}$$

$$\frac{\exp(1000)\,/\exp(3000)}{\left[\exp(1000) + \exp(2000) + \exp(3000)\right]/\exp(3000)}$$

$$w_1^T \cdot x^* + w_{10} \Rightarrow 1000 - 3000 = -2000$$

$$w_2^T \cdot x^* + w_{20} \Rightarrow 2000 - 3000 = -1000$$

$$w_3^T \cdot x^* + w_{30} \Rightarrow 3000 - 3000 = 0$$

$$p(y=1|x) = \frac{\exp(\overset{\cong 0}{-2000})}{\underset{\cong 0}{\exp(-2000)} + \underset{\cong 0}{\exp(-1000)} + \underset{= 1}{\exp(0)}} = 0$$

$$\left.\begin{aligned} p(y=1|x) &\cong 0 \\ p(y=2|x) &\cong 0 \\ p(y=3|x) &\cong 1 \end{aligned}\right\} \begin{aligned} &\hat{y}_{i1} \\ &\hat{y}_{i2} \\ &\hat{y}_{i3} \end{aligned}$$

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

⚀	⚁	⚂	⚃	⚄	⚅
5/6	0	0	0	1/6	0

$$\Rightarrow P(⚀) = \left(\tfrac{5}{6}\right)^1\left(\tfrac{0}{6}\right)^0\left(\tfrac{0}{6}\right)^0\left(\tfrac{0}{6}\right)^0\left(\tfrac{1}{6}\right)^0\left(\tfrac{0}{6}\right)^0 = 5/6$$

$$P(⚃) = \left(\tfrac{5}{6}\right)^0\left(\tfrac{0}{6}\right)^0\left(\tfrac{0}{6}\right)^0\left(\tfrac{0}{6}\right)^1\left(\tfrac{1}{6}\right)^0\left(\tfrac{0}{6}\right)^0 = 0$$

$1 \Rightarrow 100000$

$2 \Rightarrow 010000$

$\vdots$

$6 \Rightarrow 000001$

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

$$= \prod_{i=1}^{N}\prod_{c=1}^{K} P(y_i=c\mid x_i)^{y_{ic}}$$

$$y = \begin{bmatrix}1\\2\\2\\1\\3\end{bmatrix} \Rightarrow Y = \begin{array}{c|c} & c\text{th column} \\ \hline i\text{th} & \begin{bmatrix}1&0&0\\0&1&0\\0&1&0\\1&0&0\\0&0&1\end{bmatrix}\end{array}$$

one-hot encoding

y_{ic}

$$\text{log-likelihood} = \sum_{i=1}^{N}\sum_{c=1}^{K} y_{ic}\log\Big[\underbrace{P(y_i=c\mid x_i)}_{\hat{y}_{ic}}\Big]$$

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

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

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

Exercise

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

$$w_c^{(t+1)} = w_c^{(t)} \underbrace{- \eta \cdot \frac{\partial \text{Error}}{\partial w_c}}_{\Delta w_c}$$

$$w_{c0}^{(t+1)} = w_{c0}^{(t)} \underbrace{- \eta \cdot \frac{\partial \text{Error}}{\partial w_{c0}}}_{\Delta w_{c0}}$$

Hint

$$\frac{\exp(w_c)}{\sum_{d=1}^{K} \exp(w_d)}$$

w_c also appears in the denominator!!

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

Hint in the next page

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

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

Hint: $\eta \cdot \sum_{i=1}^{N} \sum_{c=1}^{K} y_{ic} (\delta_{cd} - \hat{y}_{id}) . x_i = \eta \sum_{i=1}^{N} (y_{id} - \hat{y}_{id}) . x_i$

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

1 only once

1 only once

ALGORITHM:

STEP #1: Initialize $\{w_1, w_{10}, w_2, w_{20}, \ldots, w_K, w_{K0}\}$ randomly
↳ Uniform $(-0.001, +0.001)$

STEP #2: Calculate gradients.

STEP #3: Update $\{w_1, w_{10}, w_2, w_{20}, \ldots, w_K, w_{K0}\}$ using gradients

STEP #4: Go to STEP#2 if there is enough change in the parameters.

Hint:

$$g_c(w) = \frac{\exp(w_c)}{\sum_{d=1}^{D} \exp(w_d)}$$

$$\frac{\partial g_c(w)}{\partial w_e} = \frac{\frac{\partial \exp(w_c)}{\partial w_e} \cdot \left[\sum_{d=1}^{D} \exp(w_d)\right] - \exp(w_c) \cdot \frac{\partial \left[\sum_{d=1}^{D} \exp(w_d)\right]}{\partial w_e}}{\left[\sum_{d=1}^{D} \exp(w_d)\right]^2}$$

$$= \frac{\exp(w_c) \cdot 1(c=e) \cdot \left[\sum_{d=1}^{D} \exp(w_d)\right] - \exp(w_c) \cdot \exp(w_e)}{\left[\sum_{d=1}^{D} \exp(w_d)\right]^2}$$

$$= \frac{\exp(w_c) \cdot 1(c=e) \cancel{\left[\sum_{d=1}^{D} \exp(w_d)\right]}}{\left[\sum_{d=1}^{D} \exp(w_d)\right] \cancel{\left[\sum_{d=1}^{D} \exp(w_d)\right]}} - \frac{\exp(w_c)}{\left[\sum_{d=1}^{D} \exp(w_d)\right]} \frac{\exp(w_e)}{\left[\sum_{d=1}^{D} \exp(w_d)\right]}$$

$$= g_c(w) \cdot 1(c=e) - g_c(w) \cdot g_e(w)$$