Linear Discrimination:

Multiple Classes $(K > 2)$

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

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

reference class

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

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

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

is not a function of x

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

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

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

$$P(y=1\mid x) + P(y=2\mid x) + \ldots + P(y=K\mid x) = 1$$

$$P(y=1\mid x) + P(y=2\mid x) + \ldots + P(y=K-1\mid x) = 1 - P(y=K\mid x)$$

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

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

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

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

$$\underbrace{P(y=c\mid x)}_{\text{classes } 1,2,\ldots,K-1} = \frac{\exp(w_c^T . x + w_{c0})}{1+\sum_{d=1}^{K-1}\exp\left[w_d^T . x + w_{d0}\right]}$$

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

total # of parameters $= (K-1)(D+1)$

all classes except the reference class

$$P(y=1 \mid x) = \frac{\exp[w_1^T . x + w_{10}]}{1 + \exp[w_1^T x + w_{10}] + \ldots + \exp[w_{(K-1)}^T x + w_{(K-1)0}]}$$

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

$$P(y=K-1 \mid x) = \frac{\exp[w_{(K-1)}^T . x + w_{(K-1)0}]}{1 + \exp[w_1^T . x + w_{10}] + \ldots + \exp[w_{(K-1)}^T x + w_{(K-1)0}]}$$

reference class

$$P(y=K \mid x) = \frac{1}{1 + \exp[w_1^T . x + w_{10}] + \ldots + \exp[w_{(K-1)}^T x + w_{(K-1)0}]}$$

$$P(y=c \mid x) = \frac{\exp[w_c^T . x + w_{c0}]}{\sum_{d=1}^{K} \exp[w_d^T . x + w_{d0}]} \quad \left.\right\} \begin{array}{l} \geq 0 \\ \leq 1 \\ \sum_{c=1}^{K} P(y=c \mid x) = 1 \end{array}$$

SOFTMAX

x^* is a new data point

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

$$P(y=1 \mid x) = \frac{\exp(2)}{\exp(2) + \exp(-2) + \exp(1)} \quad \text{maximum}$$

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

$$P(y=3 \mid x) = \frac{\exp(1)}{\exp(2) + \exp(-2) + \exp(1)}$$

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

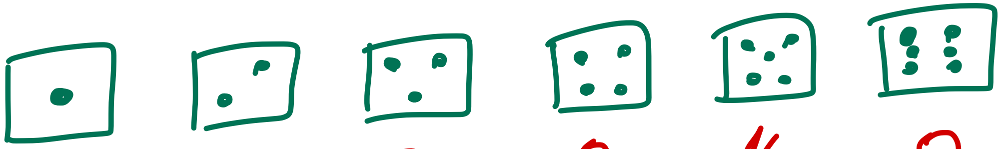
$$w_2^T x^* + w_{20} = 2000 - 3000 = -1000$$

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

$$\frac{Inf}{Inf} = \frac{\exp(1000)\,[\exp(-3000)]}{(\exp(1000) + \exp(2000) + \exp(3000))\exp(-3000)} = P(y=1 \mid x)$$

$$0 \cong \frac{\overbrace{\exp(-2000)}^{0}}{\underbrace{\exp(-2000)}_{0} + \underbrace{\exp(-1000)}_{0} + \underbrace{\exp(0)}_{1}} = P(y=1 \mid x)$$

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



5/6 0 0 0 1/6 0

$\boxed{1}\ (5/6)^1 (0)^0 (0)^0 (0)^0 (1/6)^0 (0)^0$

$\boxed{2}\ (5/6)^0 (0)^1 (0)^0 (0)^0 (1/6)^0 (0)^0$

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

$y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)$

$y_{i1} \log(\hat{y}_{i1}) + y_{i2}\log(\hat{y}_{i2})$

$$y = \begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \\ 3 \end{bmatrix} \Rightarrow Y = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad = \prod_{i=1}^{N} \prod_{c=1}^{K} P(y_i=c \mid x_i)^{y_{ic}}$$

one-hot encoding

ith row, cth column of Y

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

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

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

Exercise #6

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

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

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

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

gradient descent updates

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

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

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

binary class.

$$\eta \sum_{i=1}^{N} (y_i - \hat{y}_i) . x_i$$

$$\eta \sum_{i=1}^{N} (y - \hat{y}_i)$$

Hint:

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

1 only once   1 only once

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

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

ALGORITHM:

STEP#1: initialize $\{w_1, w_{10}, w_2, w_{20}, \ldots, w_K, w_{K0}\}$ randomly
$\hookrightarrow$ 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.