Tutorial 5: Linear Classification

5.) prior: $p(y=c) = \pi_c$

Class cond.: $p(x|y=c, \theta_c)$

$\rightarrow$ data: i.i.d

Likelihood function:

$$p(D|\{\pi_c, \theta_c\}_c) = \prod_n \prod_c [p(x^{(n)}|\theta_c) \cdot \pi_c]^{y_c^{(n)}}$$

$$\log p(D|\{\pi_c, \theta_c\}_c) = \sum_n \sum_c [\underbrace{y_c^{(n)} \log p(x^{(n)}|\theta_c)}_{\text{const in } \pi_c} + y_c^{(n)} \cdot \log \pi_c$$

$$\sum_c \pi_c = 1 \implies \sum_c \pi_c - 1 = 0$$

$$\mathcal{L} = \sum_n \sum_c [y_c^{(n)} \cdot p(x^{(n)}|\theta_c) + y_c^{(n)} \cdot \log \pi_c] - \lambda (\sum_c \pi_c - 1)$$

$$\frac{dL}{d\pi_c} = \underbrace{\sum_n y_c^{(n)}} \cdot \frac{1}{\pi_c} - \lambda$$

$$= N_c \cdot \frac{1}{\pi_c} - \lambda \overset{!}{=} 0$$

$$\pi_c = \frac{N_c}{\lambda} = \frac{N_c}{N}$$

$$\frac{dL}{d\lambda} = -\sum_c \pi_c + 1 \overset{!}{=} 0$$

$$\sum_c \pi_c = 1$$

$$\xrightarrow{\text{use } \pi_c = \frac{N_c}{\lambda}} \sum_c \frac{N_c}{\lambda} = 1 \implies \lambda = \sum_c N_c = N$$

6.) $p(x \mid y=c, \theta_c) = \mathcal{N}(x \mid \mu_c, \Sigma)$

$$= \frac{1}{(2\pi)^{D/2} \, |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu_c)^T \Sigma^{-1} (x-\mu_c)\right]$$

$$\log p(D \mid \{\pi_c, \theta_c\}) = \sum_n \sum_c \left[y_c^{(n)} \log p(x^{(n)} \mid \theta_c) + y_c^{(n)} \log \pi_c \right]$$

$$= \sum_n \sum_c y_c^{(n)} \left[-\frac{D}{2} \log(2\pi) - \frac{1}{2} \log|\Sigma| - \frac{1}{2}(x^{(n)}-\mu_c)^T \Sigma^{-1} (x^{(n)}-\mu_c) + \log \pi_c \right]$$

$$= \sum_n \sum_c -\frac{1}{2} y_c^{(n)} \left[D \log(2\pi) + \log|\Sigma| - 2 \log \pi + x^{(n)T} \Sigma^{-1} x^{(n)} - 2 \mu_c^T \Sigma^{-1} x^{(n)} + \mu_c^T \Sigma^{-1} \mu_c^T \right]$$

Derivative with respect to μ_c:

$$\sum_n -\frac{1}{2} y_c^{(n)} \left[-2 \Sigma^{-1} x^{(n)} - 2 \Sigma^{-1} \mu_c \right]$$

$$= \Sigma^{-1} \left[\sum_n y_c^{(n)} x^{(n)} - N_c \mu_c \right] \stackrel{!}{=} 0$$

$$\mu_c = \frac{1}{N_c} \sum_n y_c^{(n)} x^{(n)}$$

$$= \frac{1}{N_c} \sum_{n \mid x^{(n)} \in C_c} x^{(n)}$$

6.) Rewrite

$$\log p(D|\{\pi_c, \theta_c\}_c) = \sum_n \sum_c -\frac{1}{2} y_c^{(n)} \Big[D\log(2\pi) - 2\log\pi_c + \log|\Sigma| + \underbrace{(x^{(n)}-\mu_c)^T \Sigma^{-1} (x^{(n)}-\mu_c)}_{\text{scalar}} \Big]$$

$tr(a) = a$

$tr(ABC) = tr(BCA)$

$$= \sum_n \sum_c -\frac{1}{2} y_c^{(n)} \Big[D\log(2\pi) - 2\log\pi_c + \log|\Sigma| + tr\left(\Sigma^{-1} (x^{(n)}-\mu_c)(x^{(n)}-\mu_c)^T \right) \Big]$$

Derivative w.r.t. Σ^{-1}

$$\frac{d}{dA} tr(AB) = B^T$$

$$\left.\begin{array}{l} \frac{d}{dA} \log|A| = (A^{-1})^T \\ \log|A^{-1}| = -\log|A| \end{array}\right\} \log|\Sigma| = -\log|\Sigma^{-1}|$$

$$\rightarrow \sum_n \sum_c -\frac{1}{2} y_c^{(n)} \Big[-\Sigma + \sum_n \sum_c y_c^{(n)} (x^{(n)}-\mu_c)(x^{(n)}-\mu_c)^T \overset{!}{=} 0$$

$$\rightarrow \Sigma = \frac{1}{N} \sum_n \sum_c y_c^{(n)} (x^{(n)}-\mu_c)(x^{(n)}-\mu_c)^T \quad // \quad S_c = \frac{1}{N_c} \sum_n y_c^{(n)} (x-\mu_c)^T (x-\mu_c)$$

$$= \sum_c \frac{N_c}{N} S_c$$

1.) prior: $p(y=0) = \frac{1}{2}$

$p(y=1) = \frac{1}{2}$

$p(x|y=0) = \lambda_0 e^{-\lambda_0 x}$

$p(x|y=1) = \lambda_1 e^{-\lambda_1 x}$

$\longrightarrow p(y|x)$

$y \in \{0, 1\}$

$p(y|x = \begin{cases} \cdots & \text{if } y = 0, 1 \\ 0 & \text{else} \end{cases}$

$\longrightarrow$ Bernoulli: $q^{y}(1-q)^{1-y}$

$y_{pred} = \underset{k}{\operatorname{argmax}}\; p(y=k|x)$

$p(y=1|x) > p(y=0|x)$

$\Leftrightarrow \quad \frac{p(y=1|x)}{p(y=0|x)} > 1$

$\Leftrightarrow \ln \frac{p(y=1|x)}{p(y=0|x)} > 0$

$\ln \frac{p(y=1|x)}{p(y=0|x)}$

$= \ln \frac{p(x|y=1) \cdot p(y=1) \cdot \cancel{p(x)}}{\underline{p(x)} \cdot p(x|y=0)\, p(y=0)}$

$= \ln \frac{\lambda_1 \exp[-\lambda_1 x] \cdot \frac{1}{2}}{\lambda_0 \exp[-\lambda_0 x] \cdot \frac{1}{2}}$

$= \ln \frac{\lambda_1}{\lambda_0} + (\lambda_0 - \lambda_1)x \overset{!}{>} 0$

$(\lambda_0 - \lambda_1)x > \ln \lambda_0 - \ln \lambda_1$

Case 1: $(\lambda_0 - \lambda_1) > 0 \Leftrightarrow \lambda_0 > \lambda_1$

$x > \frac{\ln \lambda_0 - \ln \lambda_1}{\lambda_0 - \lambda_2}$

Case 2: $(\lambda_0 - \lambda_1) < 0 \Leftrightarrow \lambda_0 < 1$

$x < \frac{\ln \lambda_0 - \ln \lambda_1}{\lambda_0 - \lambda_1}$

$x \in \begin{cases} \left(\frac{\ln \lambda_0 - \ln \lambda_1}{\lambda_0 - \lambda_1}, \infty\right) & \text{if } \lambda_0 > \lambda_1 \\ \left[0, \frac{\ln \lambda_0 - \ln \lambda_1}{\lambda_0 - \lambda_1}\right) & \text{if } \lambda_0 < \lambda_1 \end{cases}$

2.) • seperating hyperplane

→ all points on correct side

• magnitude of $w \to \infty$

• prevent: regularization

3.) $$p(y=k|x) = \frac{e^{w_k^T x}}{\sum_c e^{w_c^T x}}$$ (softmax)

$$= \frac{\exp(w_1^T x)}{\exp(w_1^T x) + \exp(w_0^T x)}$$

$$= \frac{1}{1 + \frac{\exp(w_0^T x)}{\exp(w_1^T x)}}$$

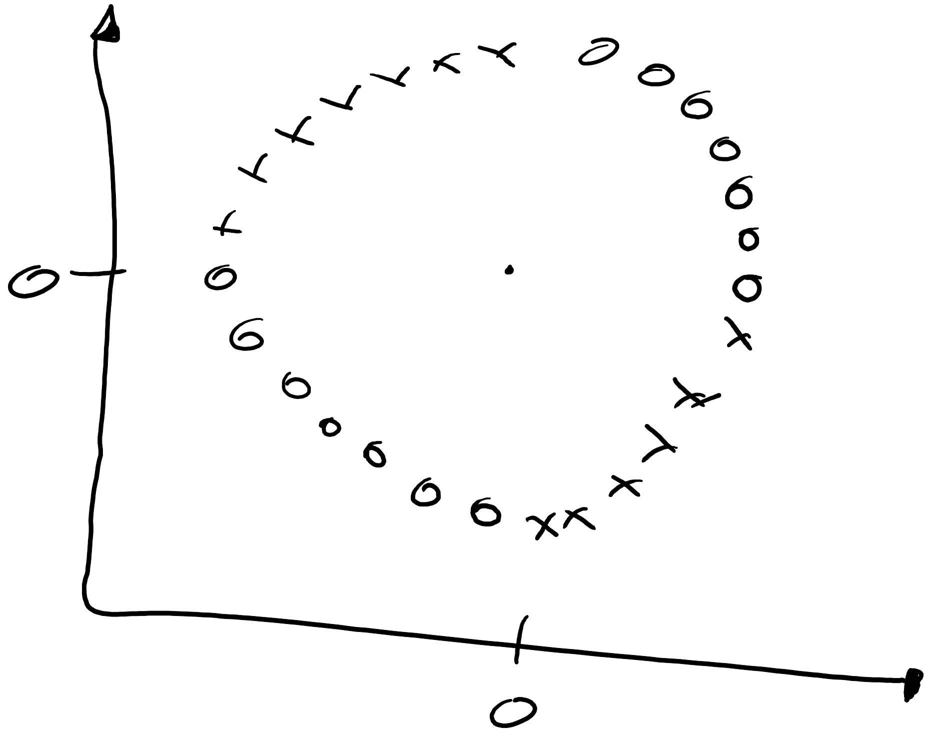
$$= \frac{1}{1 + \exp[-(w_1 - w_0)^T x]} \quad // \hat{w} = w_1 - w_0$$

$$= \frac{1}{1 + \exp(-\hat{w}^T x)}$$

sigmoid

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

4.)



$$\phi(x_1, x_2) = x_1 x_2$$