Q&A Linear Classification

Wednesday, 2 December 2020 11:01

$$\frac{\partial}{\partial \pi_c} \sum_{n=1}^{N}\sum_{c=1}^{C} y_c^{(n)} \log \pi_c - \lambda\left(\sum_{c=1}^{C}\pi_c - 1\right). \overset{?}{=}> \pi_c = \frac{1}{\lambda}\sum_{n=1}^{N} y_c^{(n)} = \frac{N_c}{\lambda}.$$

$$\frac{\partial}{\partial \pi_{\bar{c}}} \sum_n \sum_c \log \pi_c$$

$$\sum_{n=1}^{N} y_c^{(n)} \frac{1}{\pi_c} - \lambda \overset{!}{=} 0 \iff \frac{1}{\lambda}\sum_{n=1}^{N} y_c^{(n)} = \pi_c$$

$\pi_{\bar{c}}$, π_c are independent if $\bar{c} \neq c$

$$D = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots\}$$

$$p(D \mid \pi, \theta) = \prod_n p(x^n, y^n \mid \pi, \theta)$$

$$y^{(2)} = \begin{pmatrix}1\\0\\0\end{pmatrix} \qquad y_1^{(2)} = 1 \quad y_2^{(2)} = 0 \quad y_3^{(2)} = 0$$

$$p(D \mid \pi, \theta) = p(\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots\} \mid \pi, \theta)$$

$$= \prod_n p(x^{(n)}, y^{(n)} \mid \pi, \theta)$$

$$\sum_{c=1}^{C} y_c^{(n)} \log \pi_c = \frac{\partial}{\partial \pi_2} \sum_{n=1}^{N}\left(y_1^{(n)} \log \pi_1 + y_2^{(n)} \log \pi_2 + y_3^{(n)} \log \pi_3\right)$$

$$\frac{\partial}{\partial \pi_2} \overbrace{y_1^{(n)} \log \pi_1}^{\text{const wrt. } \pi_2} = 0$$

$$\frac{\partial}{\partial \pi_2} y_2^{(n)} \log \pi_2 = y_2^{(n)} \frac{1}{\pi_2}$$

$p(y \mid x)$ vs. $p(y=0 \mid x)$

general function of y | specific for $y=0$

MVN

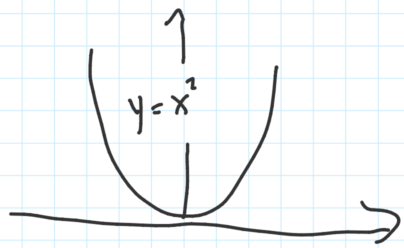
$$(2\pi)^{-\frac{D}{2}} |\Sigma|^{-\frac{1}{2}}$$

general function of y | specific for $y = 0$

$(2\pi)^{-\frac{D}{2}} |\Sigma|^{-\frac{1}{2}}$

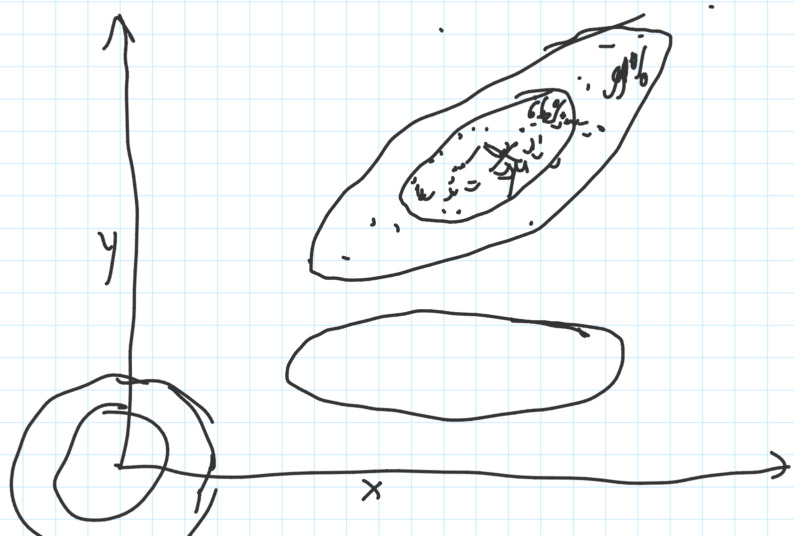
Assume a single observation $(\bar{x}, \bar{y})$ and compute $p(y|x)$

$p(y = \bar{y} \mid x = \bar{x})$



$$= -\frac{1}{2}\sum_{n=1}^{N}\sum_{c=1}^{C} y_c^{(n)} \left(\cancel{D\log 2\pi} + \cancel{\log\det(\Sigma)} + (x^{(n)} - \mu_c)^{\mathrm{T}}\Sigma^{-1}(x^{(n)} - \mu_c) - \cancel{2\log\pi_c}\right)$$

$x \sim \mathcal{N}\left(\begin{pmatrix}10\\10\end{pmatrix} \middle| \begin{pmatrix}5 & 0\\0 & 2\end{pmatrix}\right)$

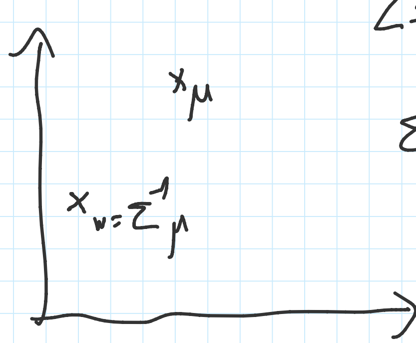


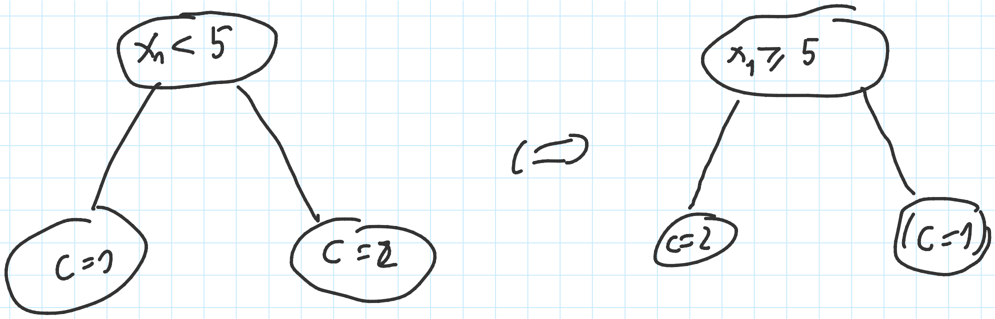
5.7

$\boxed{\phi(x,y) = \text{formula}}$

$\Sigma = \begin{pmatrix}4 & 0\\0 & 2\end{pmatrix}$

$\Sigma^{-1} = \begin{pmatrix}1/4 & 0\\0 & 1/2\end{pmatrix}$





$a = \mathrm{Tr}(a)$ for all $a \in \mathbb{R}$ and $\mathrm{Tr}(\boldsymbol{ABC}) = \mathrm{Tr}(\boldsymbol{BCA})$.

$\mathrm{Tr}(A) = \sum_i A_{ii}$

$ABC \to BCA \to CAB$

$a \in \mathbb{R} \implies \mathrm{Tr}(a) = \mathrm{Tr}((a)) = a$

$\mathrm{Tr}(ABC) = \mathrm{Tr}(BCA) = \mathrm{Tr}(CAB)$

but $\mathrm{Tr}(ABC) \neq \mathrm{Tr}(BAC)$ in general

$A \in \mathbb{R}^{n \times n}$ but $\mathrm{Tr}(A) \in \mathbb{R}$

$$Av \neq \mathrm{Tr}(A) \cdot v$$

$v^T A v \in \mathbb{R} \implies v^T A v = \mathrm{Tr}(v^T A v)$

$$\frac{\partial}{\partial v} v^T A v = (A + A^T) v$$