2/15/2016

$$\Sigma = \begin{bmatrix} E[(x_1-\mu_1)(x_1-\mu_1)] & \dots & E[(x_1-\mu_1)(x_n-\mu_n)] \\ E[(x_2-\mu_2)(x_1-\mu_1)] & \dots & E[(x_2-\mu_2)(x_n-\mu_n)] \\ \vdots & & \\ E[(x_n-\mu_n)(x_1-\mu_1)] & \dots & E[(x_n-\mu_n)(x_n-\mu_n)] \end{bmatrix}$$

$$\Sigma_{ij} = E[(x_i-\mu_i)(x_j-\mu_j)]$$

$$J(w) = \frac{w^T S_B w}{w^T S_W w} \qquad J(w) = \frac{(m_2-m_1)^2}{s_1^2+s_2^2}$$

$$S_B = (m_2-m_1)(m_2-m_1)^T \qquad m_i = \frac{1}{N_i}\sum_{n\in C_i} x_n$$

$$S_W = \sum_{n\in C_i}(x_n-m_i)(x_n-m_i)^T$$

$$\frac{d}{dw}J(w) = \frac{d}{dw}\left[\frac{w^T S_B w}{w^T S_W w}\right]$$

$$= w^T S_W w \frac{d[w^T S_B w]}{dw} - w^T S_B w \frac{d[w^T S_W w]}{dw}$$

$$= w^T S_W w \, \cancel{2} S_B w - w^T S_B w \, \cancel{2} S_W w = 0$$

$\boxed{\text{BISHOP 4.29}}$ $$\Rightarrow \underbrace{(w^T S_B w)}_{\text{scalar}} S_W w = \underbrace{(w^T S_W w)}_{\text{scalar}} S_B w$$

(2)

$$S_W w \propto S_B w$$

$$S_W^{-1} S_W w \propto S_W^{-1} S_B w$$

$$w \propto S_W^{-1} (m_2 - m_1) \underbrace{(m_2 - m_1)^T w}_{\text{scalar}}$$

$\boxed{\text{BISHOP 4.30}}$ $$w \propto S_W^{-1} (m_2 - m_1)$$

- project points on w
- to find separation, find threshold y_0 such that if $x_i \in C_1$, $y(x_i) \geq y_0$, e.g. by modeling points as a Gaussian distribution (see Bishop 1.2.4)

Perceptron

If $t_n = 1$, want $w^T \phi_n > 0$

If $t_n = -1$, want $w^T \phi_n < 0$

i.e. want $t_n w^T \phi_n > 0$ always

If this is going to be violated, want it to be as little violated (as close to 0) as possible

$$\frac{d[-w^T \phi_n t_n]}{dw} = -\phi_n t_n$$

③

Bayes theorem

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} = \frac{P(B|A)\,P(A)}{\sum_i P(B|A_i)\,P(A_i)}$$

$$P(y=1|x) \overset{?}{>} P(y=0|x)$$

$$\frac{P(y=1|x)}{P(y=0|x)} \overset{?}{>} 1$$

$$\log\left[\frac{P(y=1|x)}{P(y=0|x)}\right] \overset{?}{>} 0$$

$$P(y=1|x) = \frac{1}{1+e^{-w^Tx}}$$

$$P(y=0|x) = 1 - \frac{1}{1+e^{-w^Tx}} = \frac{\cancel{1}+e^{-w^Tx}-\cancel{1}}{1+e^{-w^Tx}}$$

$$\log\left[\frac{P(y=1|x)}{P(y=0|x)}\right] \overset{?}{>} 0 \iff \log\left(\frac{1}{e^{-w^Tx}}\right) \overset{?}{>} 0$$

$$\iff -\log e^{-w^Tx} \overset{?}{>} 0$$

$$\iff w^Tx \overset{?}{>} 0 \quad \text{(linear classifier)}$$