TUTORIAL 10

Notiztitel 27.06.2014

Outline:

- Probabilistic classification

Exercise 22

(Eq. 3.78)

Bayes for normal (Gaussian) distributed vectors:

$$d_i = \log P(w_i) - \frac{1}{2}\log|\underset{\sim}{C}_i| - \frac{1}{2}(\underline{y}-\underline{m}_i)^T \cdot \underset{\sim}{C}^{-1} \cdot (\underline{y}-\underline{m}_i)$$

a) $\underset{\sim}{C}_i = \underset{\sim}{C}$

$$d_i = \log P(w_i) - \underbrace{\frac{1}{2}\log|\underset{\sim}{C}|}_{\text{equal for all } d_i \Rightarrow \text{not nec.}} - \frac{1}{2}\Big(\overbrace{\underline{y}^T \cdot \underset{\sim}{C}^{-1}\underline{y}}^{\text{equal for all } d_i \Rightarrow \text{not nec.}} - 2\underline{y}^T\underset{\sim}{C}^{-1}\underline{m}_i + \underline{m}_i^T \cdot \underset{\sim}{C}^{-1}\underline{m}_i\Big)$$

$$d_i = \log P(w_i) + \underline{y}^T\underset{\sim}{C}^{-1}\underline{m}_i - \frac{1}{2}\underline{m}_i^T\underset{\sim}{C}^{-1}\underline{m}_i$$

b) $C = I$ $\qquad P(w_1) = P(w)$

$$d_i = \underbrace{\log P(w)}_{\Rightarrow \text{ not nec.}} - \underbrace{\frac{1}{2}\log|\underset{\sim}{I}|}_{\Rightarrow \text{ not nec.}} - \frac{1}{2}\left[\underline{y}^T \underset{\sim}{I}^{-1} \underline{y} - 2\underline{y}^T \underset{\sim}{I}^{-1} \underline{m}_i + \underline{m}_1^T \underset{\sim}{I}^{-1} \underline{m}_1\right]$$

$$\underline{d_1} = \underline{y}^T \cdot \underline{m}_i - \frac{1}{2}\underline{m}_i^T \underline{m}_i = \underline{m}_1^T \underline{y} - \underline{\frac{1}{2}\underline{m}_i^T \underline{m}_1}$$

$$= \underbrace{\left[m_{1,1}, m_{1,2}, \ldots, m_{1,N}, -\frac{1}{2}\underline{m}_i^T \cdot \underline{m}_1\right]}_{w} \cdot \begin{bmatrix} \underline{y} \\ 1 \end{bmatrix}$$

$\Rightarrow$ decision function $\rightarrow$ distance classifier (Eq. 3.11)

c) Vector $\underline{m}_i$ is the prototype of the i-th class

$$\underline{m}_i = \frac{1}{\underline{K}_i} \sum_{i=1}^{\underline{K}_i} y_i$$

The prototype is equal to the mean of all vectors of the class ω_i

Exercise 23

Decision function of the Bayes - classifier (Eq. 3.75)

$$d_i(\underline{y}) = p(\underline{y} \mid \Omega_i) \cdot p(\Omega_i) \leftarrow$$

a) Probability of a wrong classification $p(e \mid \underline{y})$

" " " correct " $p(c \mid \underline{y})$

$$p(e \mid \underline{y}) = 1 - p(c \mid \underline{y})$$

if $\underline{y}$ belongs to class Ω_i then:

$$p(c \mid \underline{y}) = p(\Omega_i \mid \underline{y})$$

$$p(e|y) = 1 - p(\Omega_i|y)$$

b) $$p(e) = \int_{Y} p(e, y)\, dy = \int_{Y} p(e|y) \cdot p(y)\, dy =$$

$$= \int_{Y} \left[1 - p(\Omega_i|y)\right] p(y)\, dy =$$

$$= \underbrace{\int_{Y} p(y)\, dy}_{1} - \int_{Y} \overbrace{p(\Omega_i|y) \cdot p(y)}^{p(\Omega_i, y)}\, dy$$

$$\Rightarrow p(e) = 1 - \int_{Y} p(\Omega_i, y)\, dy$$

c) $d_i = p(y \mid \Omega_i)\, p(\Omega_i) = p(y, \Omega_i) =$

$= p(\Omega_i \mid y) \cdot p(y)$ (Bayes)

From b) : $p(e) = 1 - \int_Y p(\Omega_i, y)\, dy$

$\Rightarrow p(e) = 1 - \int_Y d_i(y)\, dy$

$\Rightarrow$ error probability is minimised if and only if the integral is maximised.

That is, d_i is maximised.

$\Rightarrow$ d_i has been chosen as decision function

$\Rightarrow$ decision function is maximised by the class selected

$\Rightarrow$ error probability is minimised.