KHU-BME
Pattern Classification
Lecture 10

## (Revisit DHS 2.6)

### Examples: Bayes Minimum Error for Normal Density

Given: $p(\underline{x}|S_k)=N(\underline{x},\underline{m}_k,\Sigma_k)$

$$=\frac{1}{(2\pi)^{N/2}|\sum_k|^{1/2}}\exp\{-1/2(\underline{x}-\underline{m}_k)^T\Sigma_k^{-1}(\underline{x}-\underline{m}_k)\}$$

That is the likelihood is normal!

Want to maximize $p(x|S_i)P(S_i)$

Choose $g_i(x)=\ln[p(\underline{x}|S_i)P(S_i)]=\ln p(\underline{x}|S_i)+\ln P(S_i)$ <- DHS p. 36 Eq. (48)

$g_i(\underline{x})=-1/2\ln|\Sigma_i|-1/2(\underline{x}-m_i)^T\Sigma_i^{-1}(\underline{x}-m_i)+\ln P(S_i)$ (B1)

Note if $|\Sigma_1|=|\Sigma_2|=...$

And $P(S_1)= P(S_2)=\ldots$ then <- called uniform priors

$g_i(\underline{x})=-(\underline{x}-m_i)^T\Sigma_i^{-1}(\underline{x}-m_i)$ <- assign x to the closest mean of the class

⇨ minimum Mahalanobis distance to class means classifier.

Comments

- Include $\ln P(S_i)$ term => decision surface shifts to favor class with larger $P(S_i)$.
- $-1/2\ln|\Sigma_i|$ term incorporates differing ellipsoids form one class to another.

$g_i(\underline{x})=-1/2\ln|\Sigma_i|-1/2d_M^2(\underline{x},m_i)+\ln P(S_i)$ (B1 again)

**[REVIEW]**

Let's go back to Case 1, 2, and 3 again.

**Case 1: (Revisit DHS 2.6.1)**

- Features are statistically independent
- Each feature has the same variance $\sigma^2$

$\Sigma_i=\sigma^2 I$

$d_M^2(\underline{x},m_i)=1/\sigma^2\ d_E^2(\underline{x},m_i)$

$|\Sigma_i|=|\Sigma|$=independent of i

$g_i(\underline{x})=-1/(2\sigma^2)(\underline{x} - m_i)^T(\underline{x} - m_i) + \ln P(S_i)$

$g_i(\underline{x})= 1/(2\sigma^2)(2\underline{x}^T m_i - m_i^T m_i) + \ln P(S_i)$ we can drop $\underline{x}^T\underline{x}$ since same for all i

Note: it's linear: $g_i(\underline{x})=\underline{w}^{(i)T}\underline{x}+\underline{w}^{(i)}_{N+1}$

$w^{(i)}$=? $\underline{w}^{(i)}_{N+1}$=?

Minimum Euclidean distance to class means except favors classes with higher a priori probability.

Review Fig. 2.10 & Fig. 2.11 again.

**Case 2: (DHS 2.6.2)**

$\Sigma_i = \Sigma$ Same for different classes

$d_M^2$=hyperellipsoid

$g_i(\underline{x}) = -1/2\ (\underline{x} - m_i)^T \Sigma^{-1} (\underline{x} - m_i) + \ln P(S_i)$

$d_M$=constant surfaces are identically hyperellipsoids

➔ classifier is linear

$g_i(\underline{x}) = w^{(i)T} \underline{x} + w^{(i)}_{N+1}$

$w^{(i)}$=?

$\underline{w}^{(i)}_{N+1}$=?

Review Fig. 2.12 again.

## <u>Case 3 (DHS 2.6.3)</u>

**$\Sigma_i$=arbitrary**

$d_M^2(\underline{x},\underline{m}_i)$=different for each class $S_i$

$\underline{x}^T\Sigma_i^{-1}\underline{x}$ does not drop out.

⇨ $g_i$ are not linear.

Hyperquadractic decision surface (polynomials of degree 2)

Can express as: $g_i(x)=\underline{x}^T W^{(i)}\underline{x} + w^{(i)T}\underline{x} + w^{(i)}_{N+1}$ (B2)

Hyperquadractic decision surface

Hyperellipsoid

Hyperhyperboloid

Hyperparaboloid

Hypersphere

⇨ Revisit Fig. 2.14, Fig. 2.15, and Fig. 2.16

**How to calculate $W^{(i)}$ and $w^{(i)}$ for (B2)**

$g_i(x)=-1/2\ln|\Sigma_i|-1/2(\underline{x}-m)^T\Sigma_i^{-1}(\underline{x}-m_i)+\ln P(S_i)$
$= -1/2\underline{x}^T\Sigma_i^{-1}\underline{x} + 1/2[\underline{x}^T\Sigma_i^{-1}m_i + m^T\Sigma_i^{-1}\underline{x}] +$ (constant of $\underline{x}$ terms)

$1/2[\underline{x}^T\Sigma_i^{-1}\underline{m}_i + \underline{m}_i^T\Sigma_i^{-1}\underline{x}]=1/2[(\Sigma_i^{-1}\underline{m}_i)^T\underline{x} + m_i^T\Sigma_i^{-1}\underline{x}]$
$=1/2[(\Sigma_i^{-1}\underline{m}_i)^T\underline{x} + (\Sigma_i^{-1}\underline{m}_i)^T\underline{x}]$
$=(\Sigma_i^{-1}\underline{m}_i)^T\underline{x}$

$g_i(\underline{x})=\underline{x}^T[-1/2\Sigma_i^{-1}]\underline{x} + [\Sigma_i^{-1}m_i]^T\underline{x} +$ (constant of $\underline{x}$ terms)

General Bayes minimum error, normal densitites:

$g_i(\underline{x})=\underline{x}^T W^{(i)}\underline{x} + \underline{w}^{(i)T}\underline{x} + w_{N+1}^{(i)}$

⇨ $W^{(i)}=-1/2\Sigma_i^{-1}$ (DHS p. 41 Eq. (67))

$w^{(i)}= \Sigma_i^{-1}m_i$ (DHS p. 41 Eq. (68))

$w_{N+1}^{(i)}=-1/2\ln|\Sigma_i| -1/2\underline{m}_i^T\Sigma_i^{-1}\underline{m}_i + \ln P(S_i)$ (DHS p. 41 Eq. (69))

## Example 1 (DHS p. 44)

**Example 1: Decision regions for two-dimensional Gaussian data**

To clarify these ideas, we explicitly calculate the decision boundary for the two-category two-dimensional data in the Example figure. Let $\omega_1$ be the set of the four black points, and $\omega_2$ the red points. Although we will spend much of the next chapter understanding how to estimate the parameters of our distributions, for now we simply assume that we need merely calculate the means and covariances by the discrete versions of Eqs. 39 & 40; they are found to be:

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 3 \\ 6 \end{bmatrix}; \quad \boldsymbol{\Sigma}_1 = \begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix} \text{ and } \boldsymbol{\mu}_2 = \begin{bmatrix} 3 \\ -2 \end{bmatrix}; \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

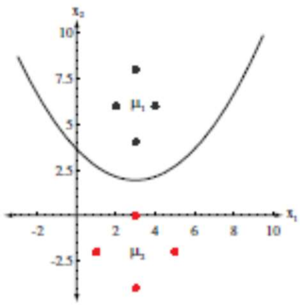
The inverse matrices are then,

$$\boldsymbol{\Sigma}_1^{-1} = \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_2^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}.$$

We assume equal prior probabilities, $P(\omega_1) = P(\omega_2) = 0.5$, and substitute these into the general form for a discriminant, Eqs. 64 – 67, setting $g_1(\mathbf{x}) = g_2(\mathbf{x})$ to obtain the decision boundary:

$$x_2 = 3.514 - 1.125x_1 + 0.1875x_1^2.$$

This equation describes a parabola with vertex at $\binom{3}{1.83}$. Note that despite the fact that the variance in the data along the $x_2$ direction for both distributions is the same, the decision boundary does not pass through the point $\binom{3}{2}$, midway between the means, as we might have naively guessed. This is because for the $\omega_1$ distribution, the probability distribution is "squeezed" in the $x_1$-direction more so than for the $\omega_2$ distribution. Because the overall prior probabilities are the same (i.e., the integral over space of the probability density), the distribution is increased along the $x_2$ direction (relative to that for the $\omega_2$ distribution). Thus the decision boundary lies slightly lower than the point midway between the two means, as can be seen in the decision boundary.




The computed Bayes decision boundary for two Gaussian distributions, each based on four data points.