Discriminant functions for Normal Density

- Lecture 3 -

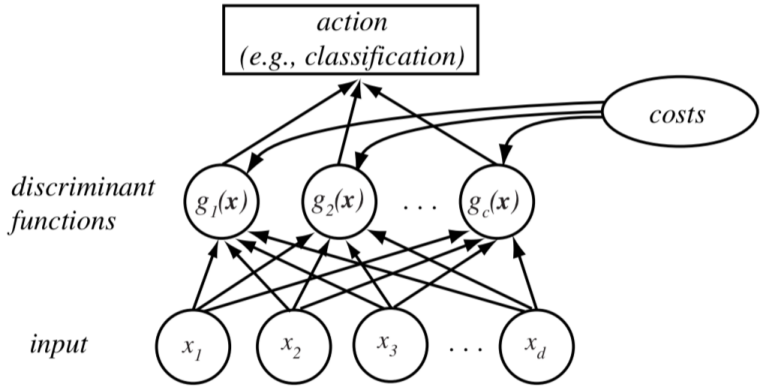


FIGURE 2.5. The functional structure of a general statistical pattern classifier which includes d inputs and c discriminant functions $g_i(\mathbf{x})$. A subsequent step determines which of the discriminant values is the maximum, and categorizes the input pattern accordingly. The arrows show the direction of the flow of information, though frequently the arrows are omitted when the direction of flow is self-evident. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

The minimum error rate classification can be achieved by the use of discriminant functions:

$$g_i(x) = \ln P(\omega_i|x) = \ln p(\omega_i|x) + \ln P(\omega_i)$$

if $p(x|\omega_i) \sim N(\mu_i, \Sigma_i)$

$$g_i(x) = \ln\left(\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu_i)^t \Sigma_i^{-1}(x-\mu_i)\right]\right)$$

$$g_i(x) = -\frac{d}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_i| - \frac{1}{2}(x-\mu_i)^t\Sigma_i^{-1}(x-\mu_i) + \ln P(\omega_i)$$

Examine the special cases:

$$g_i(x) = -\frac{d}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_i| - \frac{1}{2}(x-\mu_i)^t\Sigma_i^{-1}(x-\mu_i) + \ln P(\omega_i)$$

Case 1: $\Sigma_i = \sigma^2 I$

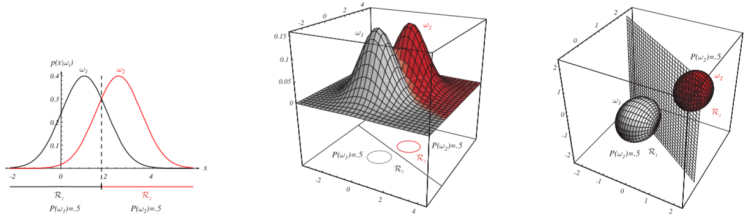


FIGURE 2.10. If the covariance matrices for two distributions are equal and proportional to the identity matrix, then the distributions are spherical in d dimensions, and the boundary is a generalized hyperplane of $d-1$ dimensions, perpendicular to the line separating the means. In these one-, two-, and three-dimensional examples, we indicate $p(\mathbf{x}|\omega_i)$ and the boundaries for the case $P(\omega_1) = P(\omega_2)$. In the three-dimensional case, the grid plane separates $\mathcal{R}_1$ from $\mathcal{R}_2$. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

$$|\Sigma_i| = \sigma^{2d} \qquad \Sigma_i^{-1} = \frac{1}{\sigma^2} I$$

- Σ_i and $\frac{d}{2}\ln(2\pi)$ constants

$$g_i(x) = -\frac{\|x-\mu_i\|^2}{2\sigma^2} + \ln P(\omega_i)$$

$$\|x-\mu_i\| = (x-\mu_i)^t(x-\mu_i)$$

$$g_i(x) = -\frac{1}{2\sigma^2}\left[x^t x - 2\mu_i^t x + \mu_i^t\mu_i\right] + \ln P(\omega_i)$$

same for all i → ignore

Hence, we get

$$g_i(x) = w_i^t x + w_{i0}$$

where, $w_i = \frac{1}{\sigma^2} \mu_i$

$$w_{i0} = -\frac{1}{2\sigma^2} \mu_i^t \mu_i + \ln P(w_i)$$ → threshold or bias for i^{th} category.

Decision surface $g_i(x) = g_j(x)$ → hyperplanes

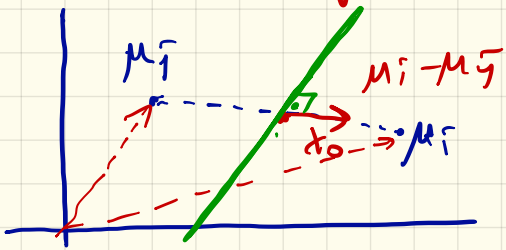
In this case, $g_i(x) = g_j(x)$ can be rewritten as:

$$w^t (x - x_0) = 0 \text{ , where}$$

$$\vec{w} = \mu_i - \mu_j$$

$$x_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{\sigma^2}{\|\mu_i - \mu_j\|} \ln\left(\frac{P(w_i)}{P(w_j)}\right)(\mu_i - \mu_j)$$

This eqn. defines a hyperplane through the point x_0 and orthogonal to the vector $\vec{w}$



if $P(w_i) = P(w_j)$

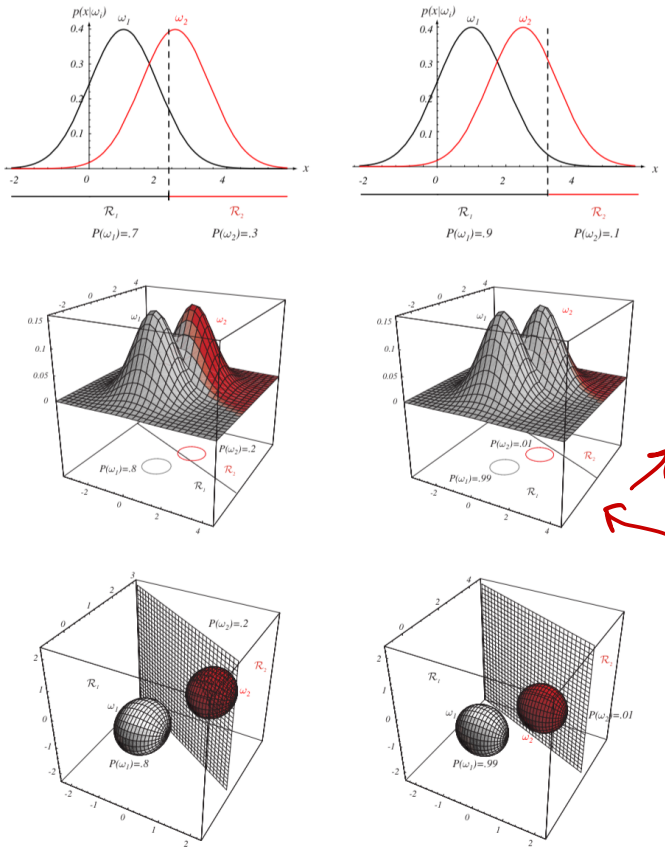


FIGURE 2.11. As the priors are changed, the decision boundary shifts; for sufficiently disparate priors the boundary will not lie between the means of these one-, two- and three-dimensional spherical Gaussian distributions. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

if $P(\omega_i) \neq P(\omega_j)$, the point x_0 shifts away from the more likely mean

Case 2: $\Sigma_i = \Sigma$ (identical for all classes but arbitrary)

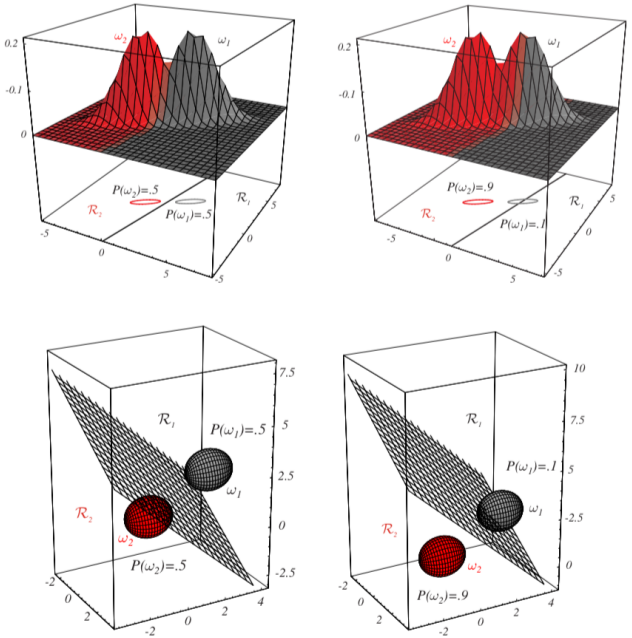


FIGURE 2.12. Probability densities (indicated by the surfaces in two dimensions and ellipsoidal surfaces in three dimensions) and decision regions for equal but asymmetric Gaussian distributions. The decision hyperplanes need not be perpendicular to the line connecting the means. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

$g_i(x) = w_i x + w_{i0}$, where

$w_i = \Sigma^{-1} \mu_i$,

$w_{i0} = -\frac{1}{2} \mu_i^t \Sigma^{-1} \mu_i + \ln P(w_i)$

$\Rightarrow$ decision boundaries are hyperplanes

Case $\Sigma_i = \Sigma$

if priors are equal for all c classes, $\ln P(\omega_i)$ can be ignored. Then the optimal decision rule:

To classify $\vec{x}$, measure the squared Mahalanobis distance $(x-\mu)^t \Sigma^{-1} (x-\mu_i)$ from $\vec{x}$ to each of the c mean vectors (μ_i) and assign $\vec{x}$ to the category of the nearest mean.

Case 3 : Σ_i = arbitrary

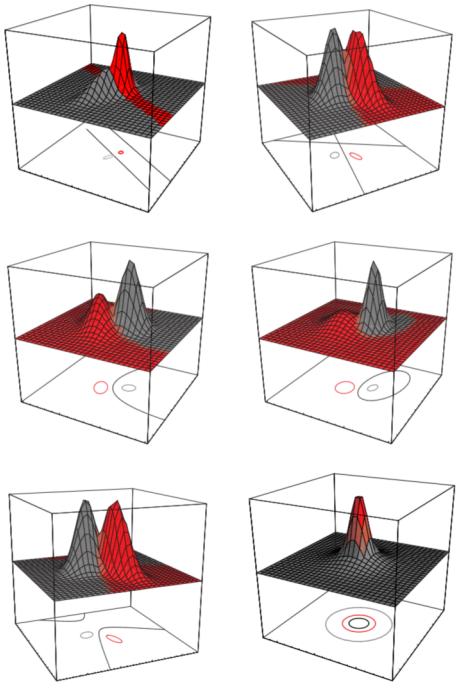


FIGURE 2.14. Arbitrary Gaussian distributions lead to Bayes decision boundaries that are general hyperquadrics. Conversely, given any hyperquadric, one can find two Gaussian distributions whose Bayes decision boundary is that hyperquadric. These variances are indicated by the contours of constant probability density. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

- cov. matrices are different for each category.

$$g_i(x) = x^t W_i x + w_i^t x + w_{i0}$$

Decision surface: (2 category case)

hyperquadrics:

- hyperplanes, pairs of hyperplanes, hyperspheres, hyperellipsoids, hyperparaboloids etc.