Lecture 8: Bayes Classification

Richa Singh

Google classroom code: wgzuohn

Slides are prepared from several information sources including Duda, Hart, Stork

Recap: Analyzing Covariance Matrix

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

- Case $\Sigma_i = \sigma^2.I$ (I stands for the identity matrix)
- Case $\Sigma_i = \Sigma$ (covariance of all classes are identical but arbitrary!)
- Case Σi = actual covariance

Case $\Sigma_i = \sigma^2.I$ (I stands for the identity matrix)

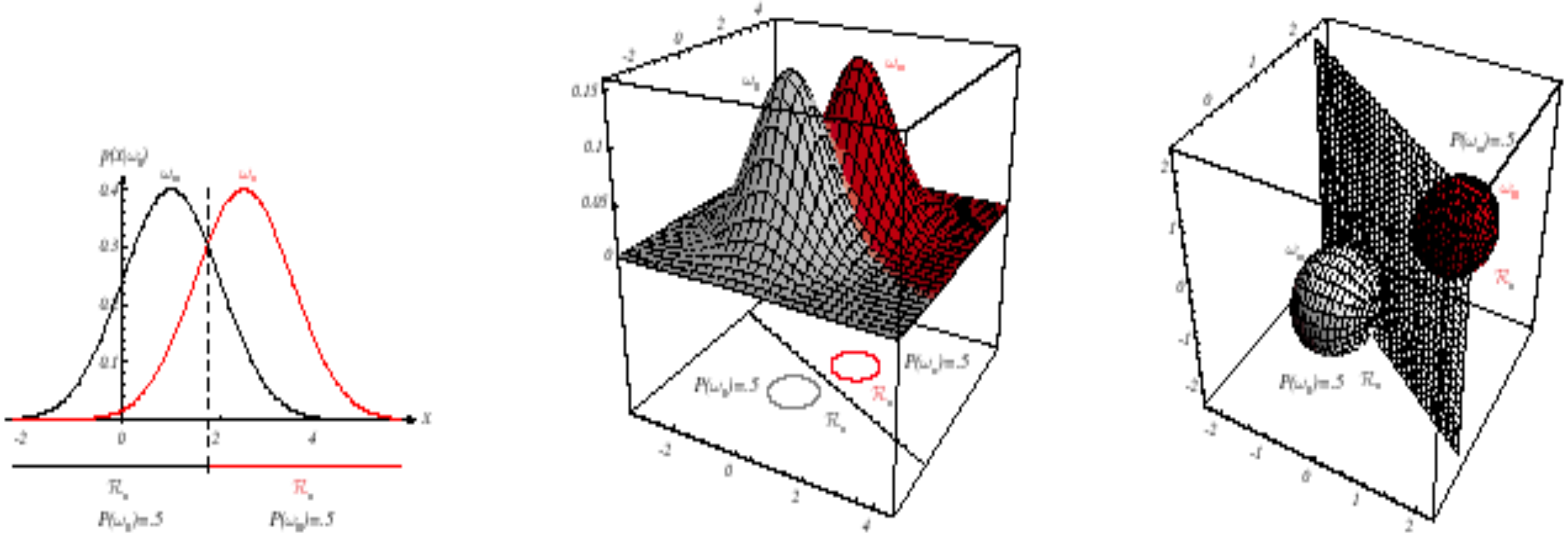


FIGURE 2.10. If the covariance matrices for two distributions are equal and proportional to the identity matrix, then the distributions are spherical in d dimensions, and the boundary is a generalized hyperplane of $d-1$ dimensions, perpendicular to the line separating the means. In these one-, two-, 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.

Solve the questions in the form.

It will be used for taking the attendance.

Discriminant Functions for the Normal Density…

- Case $\Sigma_i = \Sigma$ (covariance of all classes are identical but arbitrary!)

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

  - Expand the term and disregard the quadratic expression

where :

$$g_i(x) = w_i^t x + w_{i0} \qquad w_i = \sum^{-1}\mu; \;\; w_{i0} = -\frac{1}{2}\mu_i^t \sum^{-1}\mu_i + \ln P(\omega_i)$$

Discriminant Functions for the Normal Density…

$$x_0 = \frac{1}{2}(\mu_i + \mu_j) - \frac{\ln\left[P(\omega_i)/P(\omega_j)\right]}{(\mu_i - \mu_j)^t \Sigma^{-1}(\mu_i - \mu_j)} .(\mu_i - \mu_j)$$

- Comments about this hyperplane:
  - It passes through $\mathbf{x_0}$
  - It is NOT orthogonal to the line linking the means.
  - What happens when *$P(\omega_i)= P(\omega_j)$* ?
  - If *$P(\omega_i)$!= $P(\omega_j)$*, then $\mathbf{x_0}$ shifts away from the more likely mean.

Discriminant Functions for the Normal Density…

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

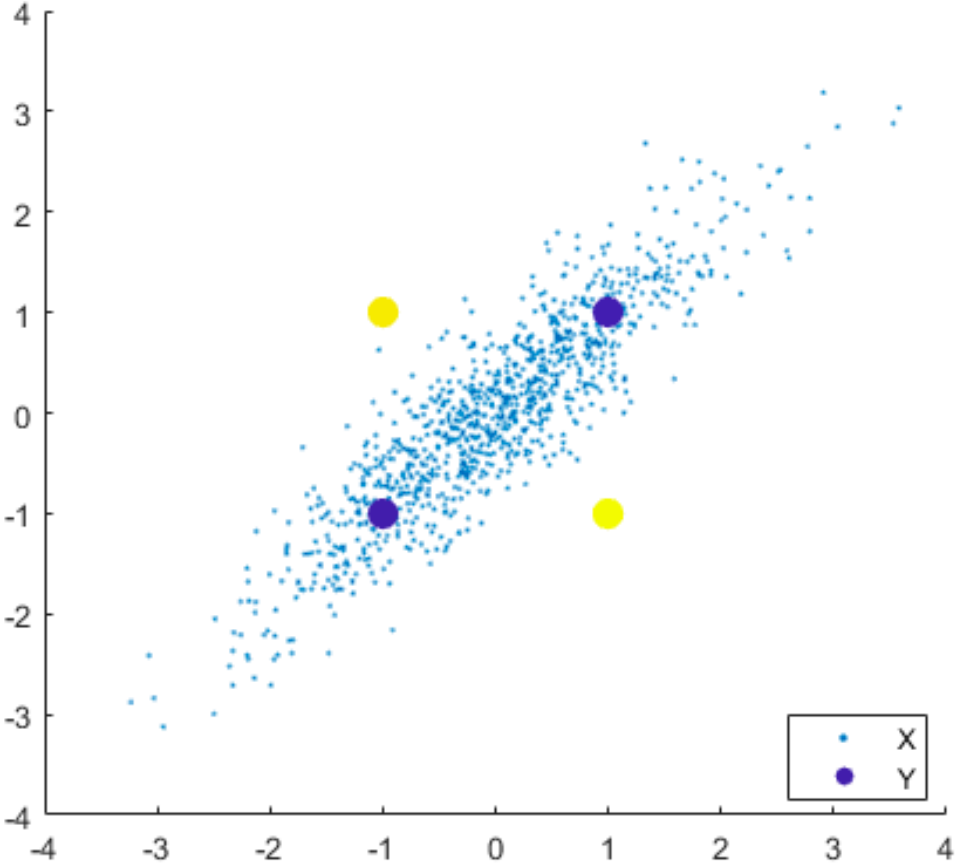
- When $P(\omega_i)$ is the same for each of the *c* classes
- Case I: Euclidean distance classifier

$$g_i(x) = -\|x-\mu_i\|^2$$

- Case II: Mahalanobis distance classifier

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^t \Sigma_i^{-1}(x-\mu_i)$$

Mahalanobis Distance



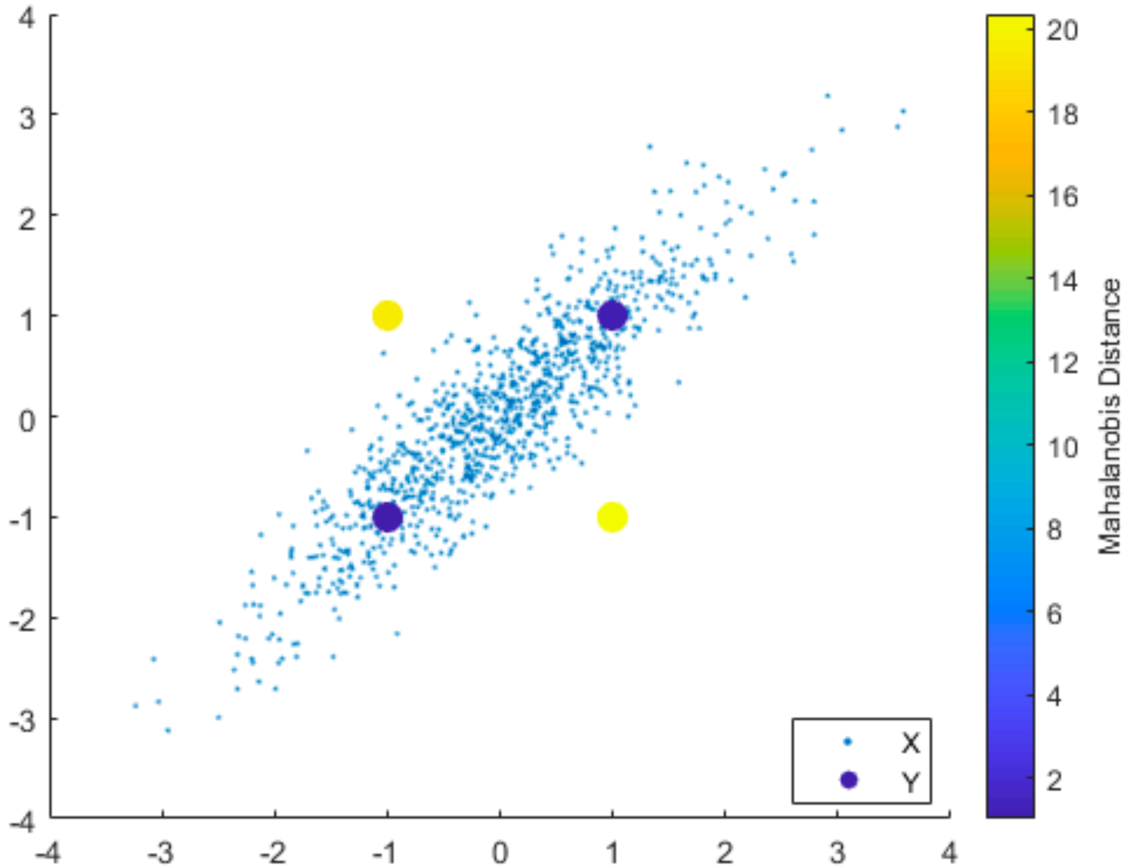
Euclidean Distance:

$$d = \sqrt{\sum_{i=1}^{n} (X_i - Y_i)^2}$$

Mahalanobis Distance:

$$d\left(\vec{x}, \vec{y}\right) = \sqrt{\left(\vec{x} - \vec{y}\right)^{\top} S^{-1} \left(\vec{x} - \vec{y}\right)}$$

Mahalanobis Distance



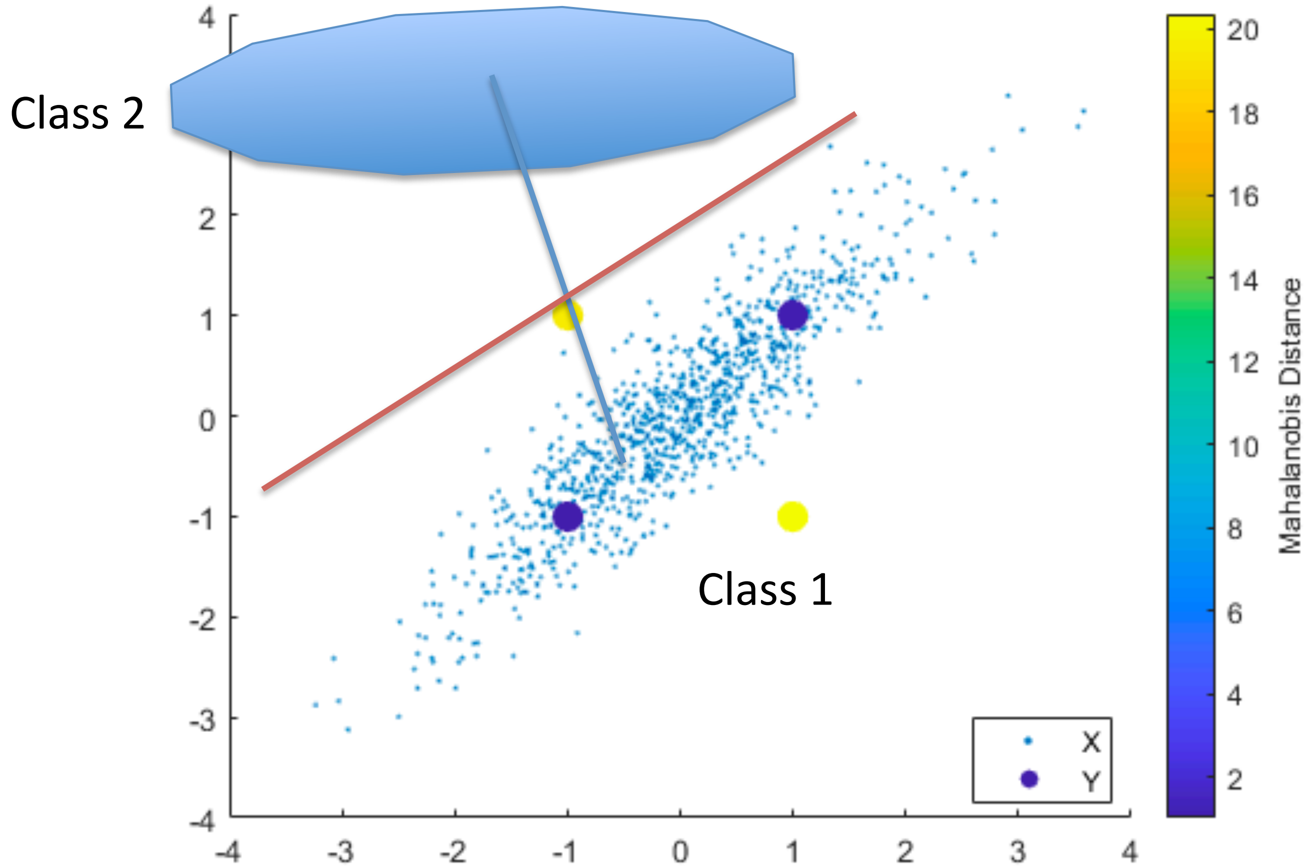
$$d = \sqrt{\sum_{i=1}^{n} (X_i - Y_i)^2}$$

Euclidean Distance:

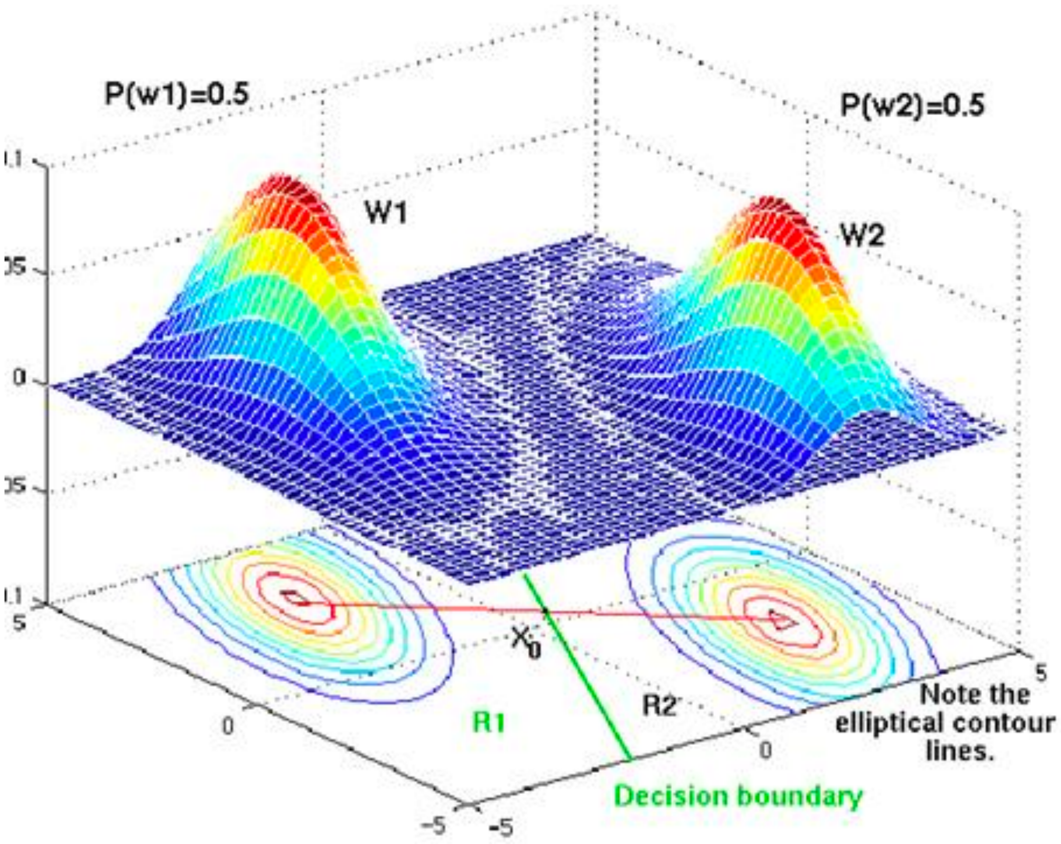
All points are equidistant

$$d\left(\vec{x}, \vec{y}\right) = \sqrt{\left(\vec{x} - \vec{y}\right)^{\top} S^{-1} \left(\vec{x} - \vec{y}\right)}$$

Mahalanobis Distance

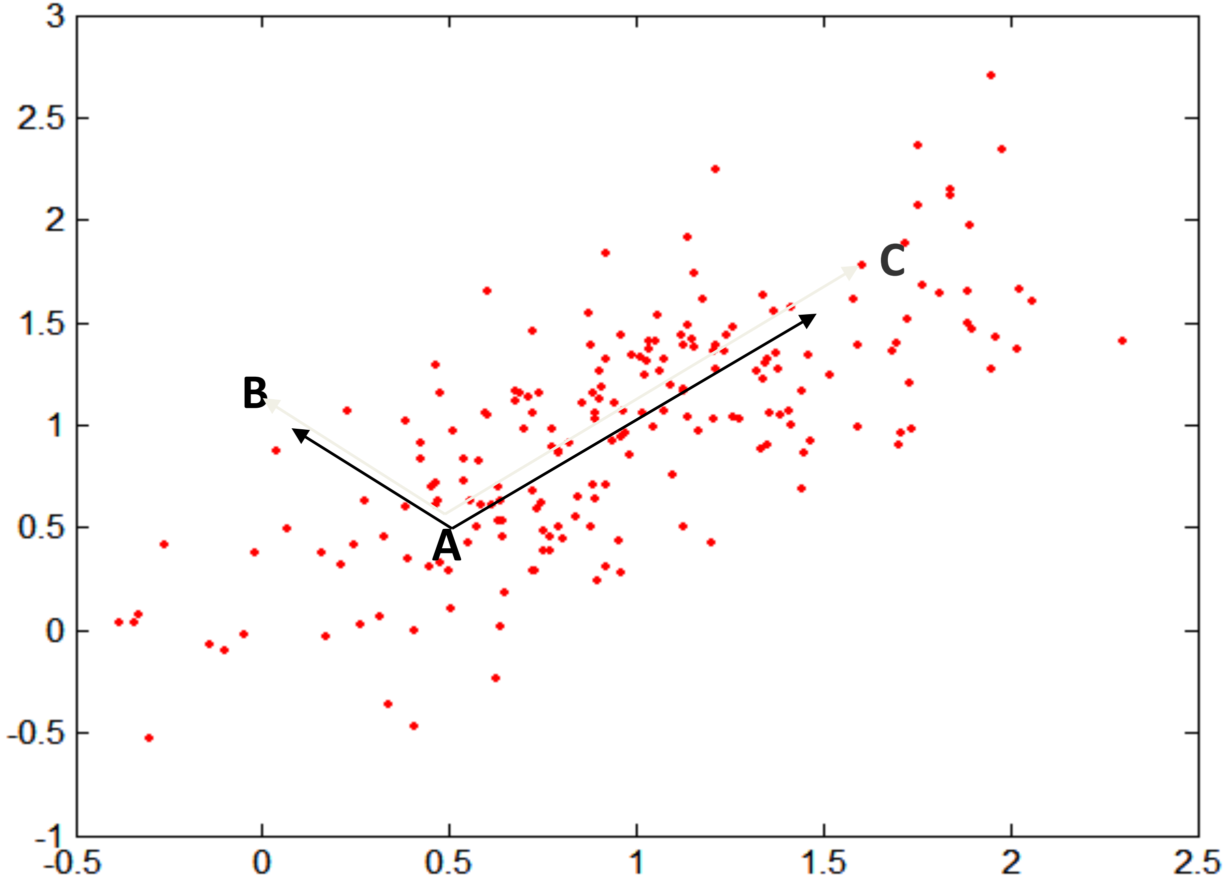


Discriminant Functions for the Normal Density…



The contour lines are elliptical in shape because the covariance matrix is not diagonal. However, both densities show the same elliptical shape. The prior probabilities are the same, and so the point x0 lies halfway between the 2 means. The decision boundary is not orthogonal to the red line. Instead, it is is tilted so that its points are of equal distance to the contour lines in w1 and those in w2.

Mahalanobis Distance



Covariance Matrix:

$$\Sigma = \begin{bmatrix} 0.3 & 0.2 \\ 0.2 & 0.3 \end{bmatrix}$$

A: (0.5, 0.5)

B: (0, 1)

C: (1.5, 1.5)

Euclid(A,B)

Euclid(A,C)

Mahal(A,B)

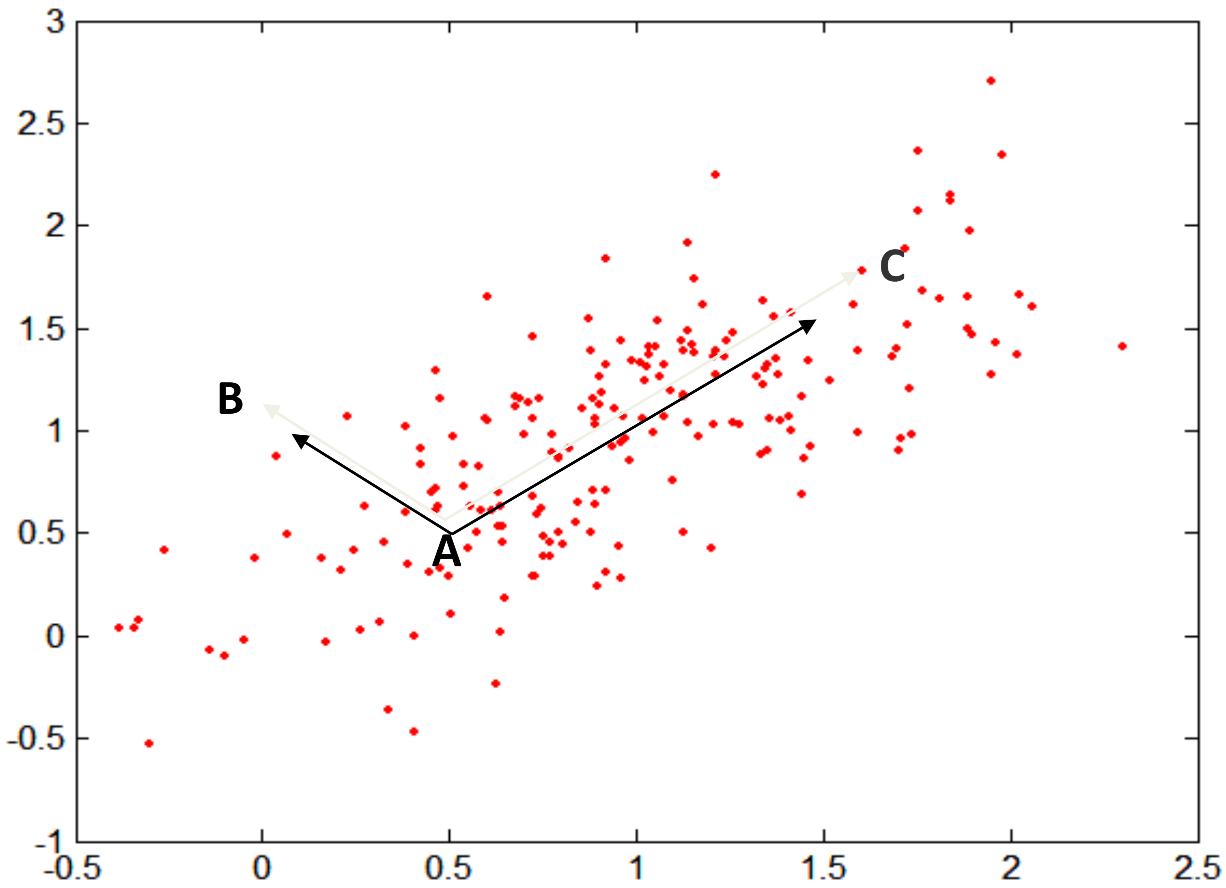
Mahal(A,C)

Compute squared versions

$$d = \sqrt{\sum_{i=1}^{n} (X_i - Y_i)^2}$$

$$d\left(\vec{x}, \vec{y}\right) = \sqrt{\left(\vec{x} - \vec{y}\right)^{\top} S^{-1} \left(\vec{x} - \vec{y}\right)}$$

Mahalanobis Distance



Covariance Matrix:

$$\Sigma = \begin{bmatrix} 0.3 & 0.2 \\ 0.2 & 0.3 \end{bmatrix}$$

A: (0.5, 0.5)

B: (0, 1)

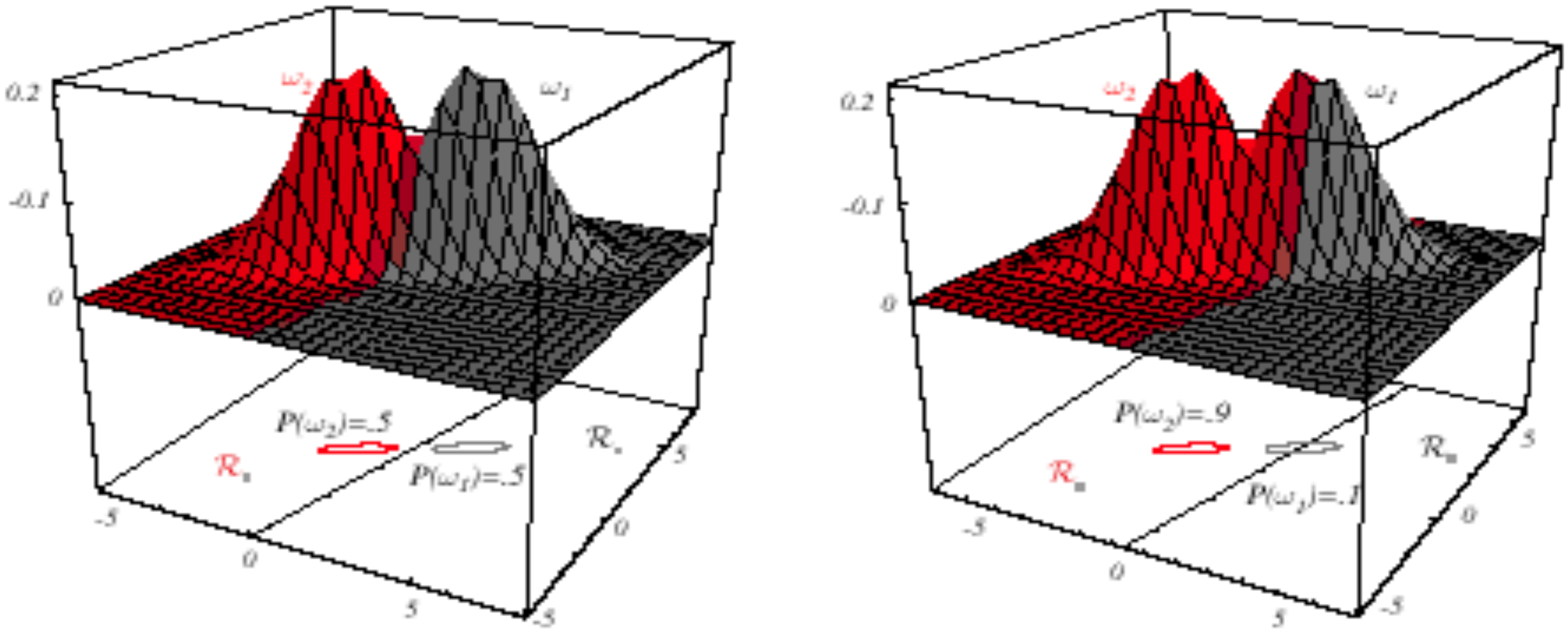
C: (1.5, 1.5)

Euclid(A,B) = 0.5

Euclid(A,C) = 2

Mahal(A,B) = 5

Mahal(A,C) = 4



0.2
-0.1
0
ω_2
ω_1
$P(\omega_2)=.5$
$P(\omega_1)=.5$
$\mathcal{R}_2$
$\mathcal{R}_1$
-5
0
5
$P(\omega_2)=.9$
$P(\omega_1)=.1$

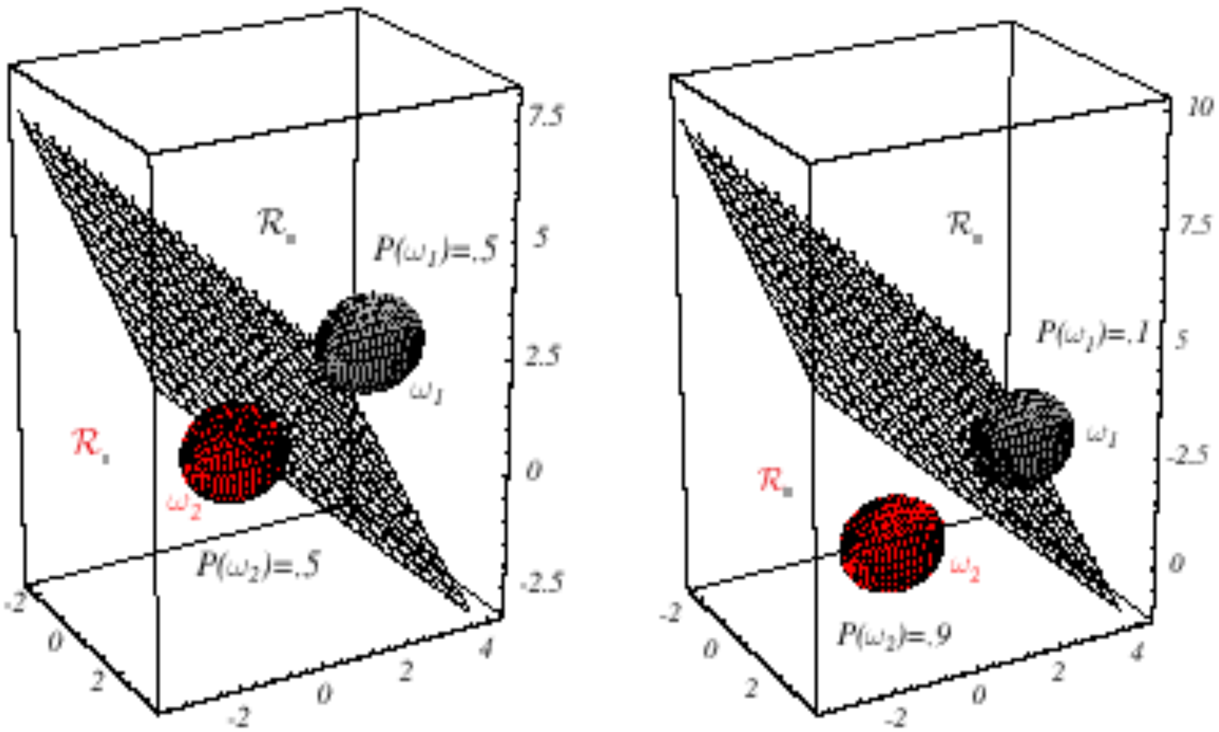


FIGURE 2.12. Probability densities (indicated by the surfaces in two dimensions and ellipsoidal surfaces in three dimensions) and decision regions for equal but asymmetric Gaussian distributions. The decision hyperplanes need not be perpendicular to the line connecting the means. From: Richard O. Duda,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

Solve the questions in the form.

Discriminant Functions for the Normal Density…

- Case Σi = arbitrary
  - The covariance matrices are different for each category

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

where :

$$W_i = -\frac{1}{2}\Sigma_i^{-1}$$

$$w_i = \Sigma_i^{-1}\mu_i$$

$$w_{i0} = -\frac{1}{2}\mu_i^t\Sigma_i^{-1}\mu_i - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i)$$

Discriminant Functions for the Normal Density…

Disconnected decision regions

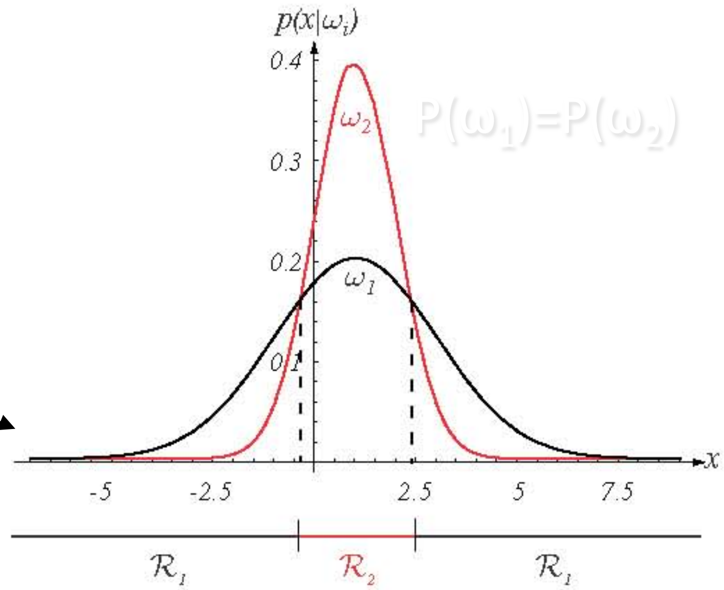
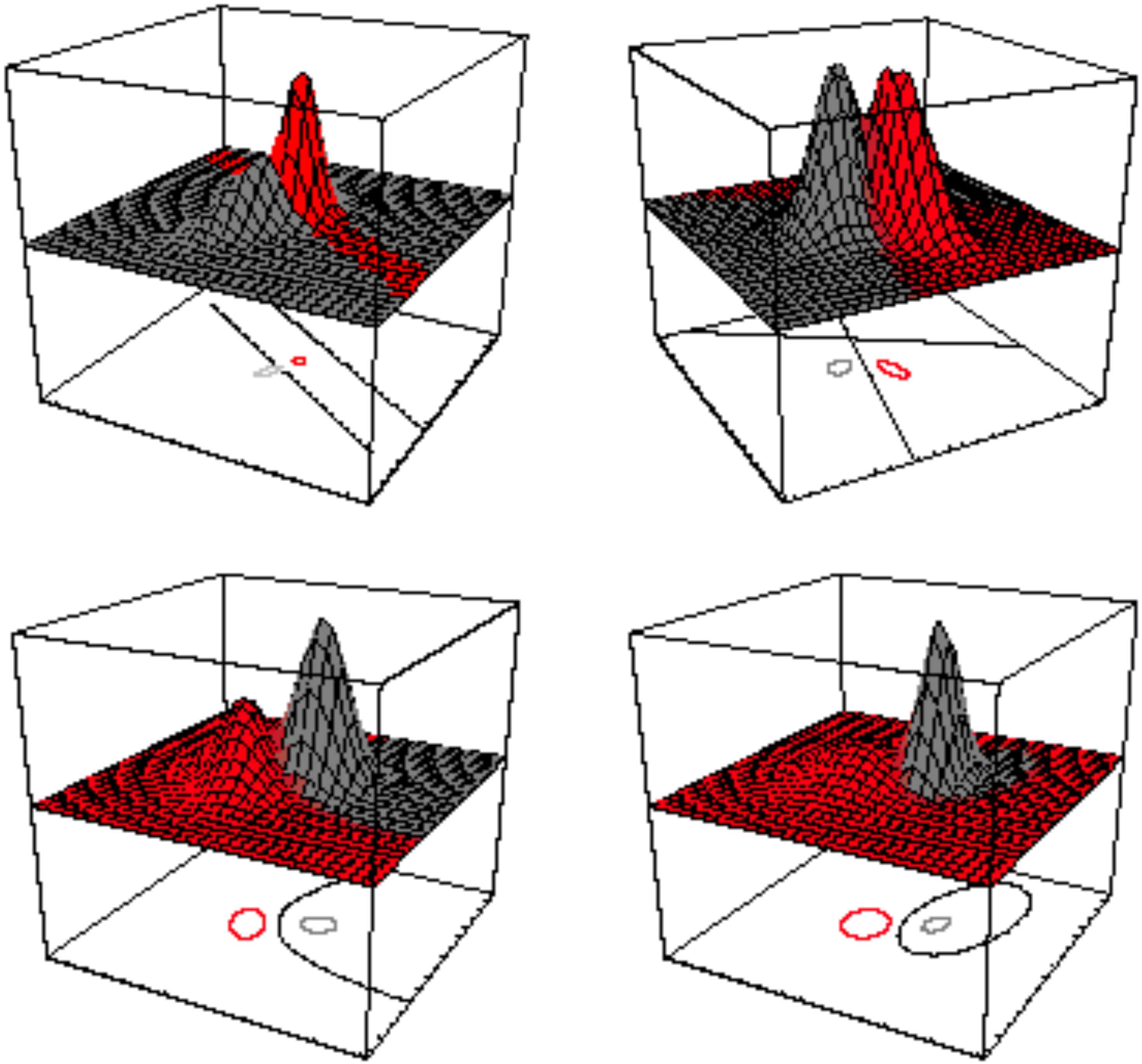


FIGURE 2.13. Non-simply connected decision regions can arise in one dimensions for Gaussians having unequal variance. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.



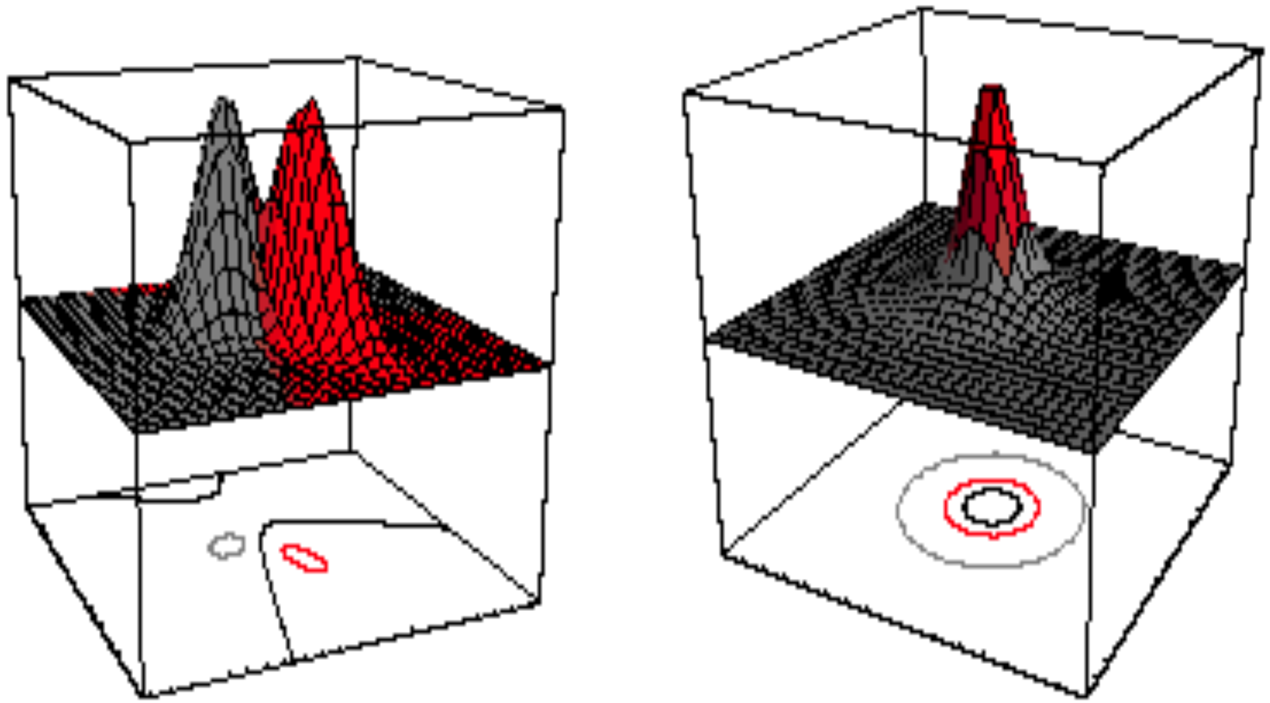


FIGURE 2.14. Arbitrary Gaussian distributions lead to Bayes decision boundaries that are general hyperquadrics. Conversely, given any hyperquadric, one can find two Gaussian distributions whose Bayes decision boundary is that hyperquadric. These variances are indicated by the contours of constant probability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

Discriminant Functions for the Normal Density

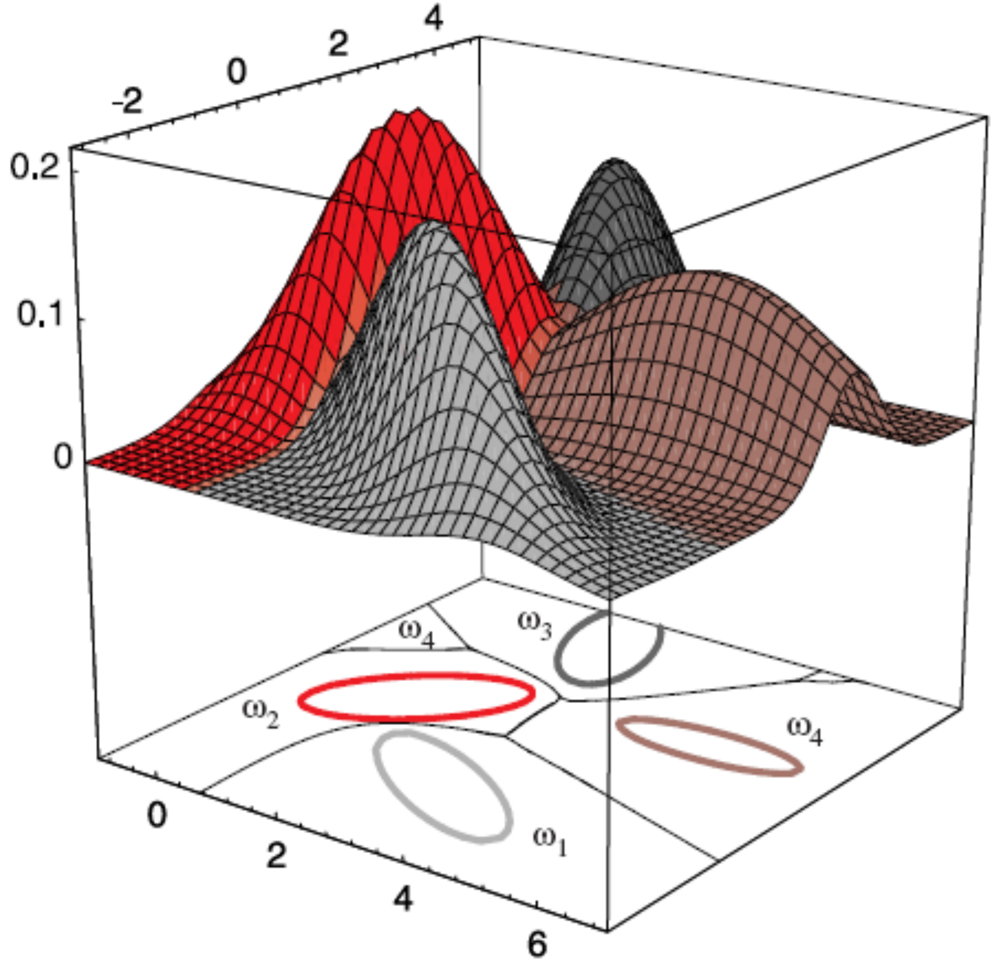
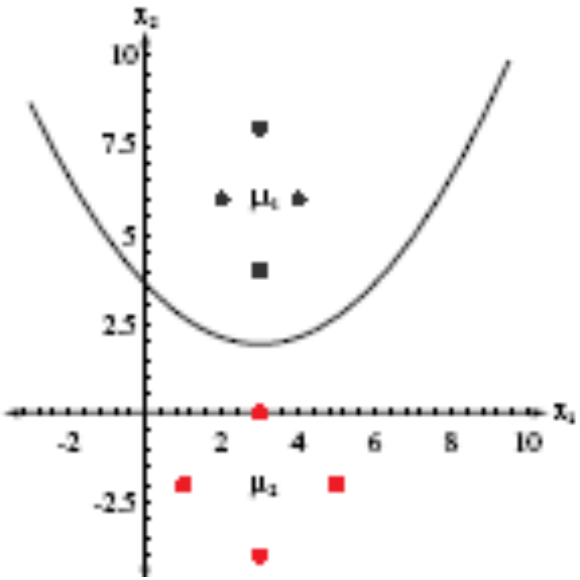


Figure 2.16: The decision regions for four normal distributions. Even with such a low number of categories, the shapes of the boundary regions can be rather complex.

Decision Regions for Two-Dimensional Gaussian Data

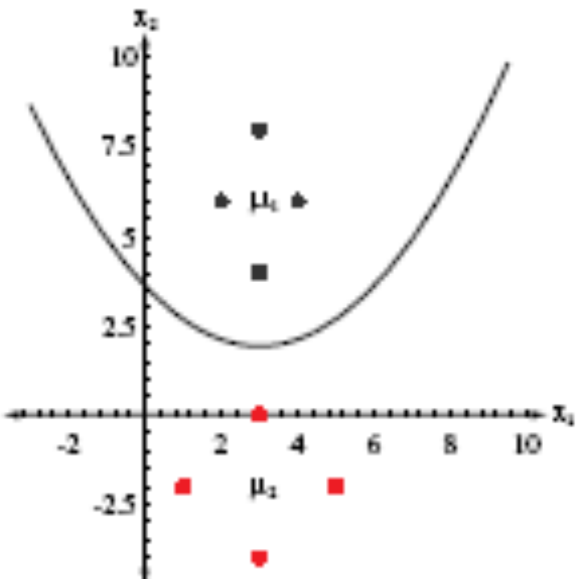


$$\mu_1 = \begin{bmatrix} 3 \\ 6 \end{bmatrix}; \quad \Sigma_1 = \begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix} \text{ and } \mu_2 = \begin{bmatrix} 3 \\ -2 \end{bmatrix}; \quad \Sigma_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

$$\Sigma_1^{-1} = \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix} \quad \text{and} \quad \Sigma_2^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}.$$

$$P(\omega_1) = P(\omega_2) = 0.5,$$

Decision Regions for Two-Dimensional Gaussian Data



$$\mu_1 = \begin{bmatrix} 3 \\ 6 \end{bmatrix}; \quad \Sigma_1 = \begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix} \text{ and } \mu_2 = \begin{bmatrix} 3 \\ -2 \end{bmatrix}; \quad \Sigma_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

$$\Sigma_1^{-1} = \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix} \quad \text{and} \quad \Sigma_2^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}.$$

$$P(\omega_1) = P(\omega_2) = 0.5,$$

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

Error Probabilities and Integrals

- 2-class problem: There are two types of errors

$$
\begin{aligned}
P(error) &= P(\mathbf{x} \in \mathcal{R}_2, \omega_1) + P(\mathbf{x} \in \mathcal{R}_1, \omega_2) \\
&= P(\mathbf{x} \in \mathcal{R}_2 | \omega_1) P(\omega_1) + P(\mathbf{x} \in \mathcal{R}_1 | \omega_2) P(\omega_2) \\
&= \int_{\mathcal{R}_2} p(\mathbf{x}|\omega_1) P(\omega_1)\, d\mathbf{x} + \int_{\mathcal{R}_1} p(\mathbf{x}|\omega_2) P(\omega_2)\, d\mathbf{x}.
\end{aligned}
$$

- Multi-class problem
  - Simpler to computer the prob. of being correct (more ways to be wrong than to be right)

$$
P(correct) = \sum_{i=1}^{c} P(\mathbf{x} \in \mathcal{R}_i, \omega_i) \quad
\begin{aligned}
&= \sum_{i=1}^{c} P(\mathbf{x} \in \mathcal{R}_i | \omega_i) P(\omega_i) \\
&= \sum_{i=1}^{c} \int_{\mathcal{R}_i} p(\mathbf{x}|\omega_i) P(\omega_i)\, d\mathbf{x}.
\end{aligned}
$$

Fractal 1 - Last class.

Topic left: Support Vector Machine

Best of luck.