Practical Pattern Recognition

Thomas M. Breuel

December 7, 2009

Chapter 1

Decision Regions for the Normal Density

1.1 Recall: Multivariate Gaussian Density

$$N(\mu, \Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu)}$$

note the exponent $-\frac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu)$ this is a quadratic form in x Σ is positive definite and symmetric (or at least positive semi definite) does everybody remember what that means? also recall the eigendecomposition $\Sigma = PDP^T$ where P is orthonormal and D is a diagonal matrix Motiviation measurement x is generated by taking some ideal prototype vector μ_i for class i and corrupting it with (measurement) noise ν if the noise ν is the result of the addition of many small noise contributions $\nu = \sum_k \delta_k$, then ν has approximately a normal density (by the central limit theorem) furthermore, measurement devices are usually adjusted so that they have no "bias", so $\tilde{\nu} N(0, \Sigma_i)$ for some Σ_i, and

$$p(x|\omega_i) = N(x; \mu_i, \Sigma_i)$$

1.2 Decision Regions for the Normal Density

assume that we have classes $\omega_i \in \{\omega_1 \ldots \omega_C\}$ assume that the ccds are all normal, with parameters μ_i, Σ_i recall that we can use $g_i(x) = \ln p(x|\omega_i) + \ln P(\omega_i)$ as a discriminant function for Bayes-optimal classification under a zero-one loss function what do the decision regions and decision boundaries look like? we can plug in the normal density and obtain

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

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

it turns out that there are several useful cases to distinguish: – all $\Sigma_i = \sigma^2 I$ – all $\Sigma_i = \Sigma$ (all the same) – arbitrary Σ_i let's look at these in turn

1.3 $\Sigma_i = \sigma^2 I$

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

becomes

$$g_i(x) = -\frac{1}{2}(x - \mu_i) \cdot (\sigma^{-2} I) \cdot (x - \mu_i) - \frac{1}{2} \ln \sigma^2 + \ln P(\omega_i)$$

leave out the common term

$$g_i(x) = -\frac{1}{2}(x - \mu_i) \cdot (\sigma^{-2} I) \cdot (x - \mu_i) + \ln P(\omega_i)$$

simplify the quadratic form

$$g_i(x) = -\frac{1}{2}\sigma^{-2}(x - \mu_i) \cdot (x - \mu_i) + \ln P(\omega_i)$$

rewrite

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

let's assume uniform priors for all classes, then this is equivalent to the discriminant function

$$g_i(x) = -\frac{1}{2}\sigma^{-2}||x - \mu_i||^2$$

or even simpler

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

this means that in the case of uniform priors and equal, spherical covariance matrices, we classify each x according to which of the prototype vectors μ_i it is closest to the decision regions are the cells of the voronoi diagram generated by the μ_i (who has seen the voronoi diagram already?) — let's return now to the more general case

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

expand the quadratic

$$g_i(x) = -\frac{1}{2}\sigma^{-2}(x \cdot x - 2\mu_i \cdot x + \mu_i \cdot \mu_i) + \ln P(\omega_i)$$

now, $x \cdot x$ is the same for all i (it's independent of i), so we get the same decision regions from

$$g_i(x) = -\frac{1}{2}\sigma^{-2}(-2\mu_i \cdot x + \mu_i \cdot \mu_i) + \ln P(\omega_i)$$

$$g_i(x) = \sigma^{-2}\mu_i \cdot x - \frac{1}{2}\sigma^{-2}\mu_i \cdot \mu_i + \ln P(\omega_i)$$

this is of the form

$$g_i(x) = w_i \cdot x + b_i$$

that is, it is a linear function of x we have *linear discriminant functions* linear discriminant functions are very important here, we have derived them in a model-based approach however, later, we will find ways of learning them directly what is the shape of the decision boundaries for linear discriminant functions? recall that the decision boundaries are where $g_i(x) = g_j(x)$

$$w_i \cdot x + b_i = w_j \cdot x + b_j$$

or

$$(w_i - w_j) \cdot x + (b_i - b_j) = 0$$

this is the equation for a plane that is, the decision boundaries are hyperplanes again, this is true even for unequal priors

1.4 $\Sigma_i = \Sigma$

that is, all the Σ_i are the same

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

so, we drop the subscript on the Σ_i

$$g_i(x) = -\frac{1}{2}(x - \mu_i) \cdot \Sigma^{-1} \cdot (x - \mu_i) - \frac{1}{2}\ln|\Sigma| + \ln P(\omega_i)$$

now, we get rid of the terms that are the same for everything

$$g_i(x) = -\frac{1}{2}(x - \mu_i) \cdot \Sigma^{-1} \cdot (x - \mu_i) + \ln P(\omega_i)$$

we multiply out the quadratic form

$$g_i(x) = -\frac{1}{2}(x \cdot \Sigma^{-1} \cdot x - 2\mu_i \cdot \Sigma^{-1} \cdot x + \mu_i \cdot \Sigma_i \cdot \mu_i) + \ln P(\omega_i)$$

as before, the $x\Sigma^{-1}x$ is independent of i, so we can drop it and get

$$g_i(x) = \mu_i \cdot \Sigma^{-1} \cdot x - \frac{1}{2}\mu_i \cdot \Sigma_i \cdot \mu_i) + \ln P(\omega_i)$$

this is again of the form

$$g_i(x) = w_i \cdot x + b_i$$

with $w_i = \mu_i \cdot \Sigma^{-1}$ so we again have linear discriminant functions and linear decision boundaries however, the interpretation as Euclidean distance fails instead, in the case of equal priors, this amounts to classification using a different distance, namely the Mahalanobis distance

$$d(x, y) = \sqrt{(x - y) \cdot \Sigma^{-1} \cdot (x - y)}$$

1.5 $\Sigma_i =$ arbitrary

is there anything we can say about the general case?

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

if we multiply this out, we get discriminant functions of the form

$$g_i(x) = xQx + wx + b$$

these are general quadratic functions their intersections are called *hyperquadrics* they include points, lines, hyperplanes, hyperellipsoids, hyperparaboloids, and hyperhyperboloids (my favorite!)

1.6 Examples

[show slides with example decision surfaces]

this concludes, for now, our treatment of Chapter 2 in DHS

1.7 Equal, Spherical Covariances, Equal Priors

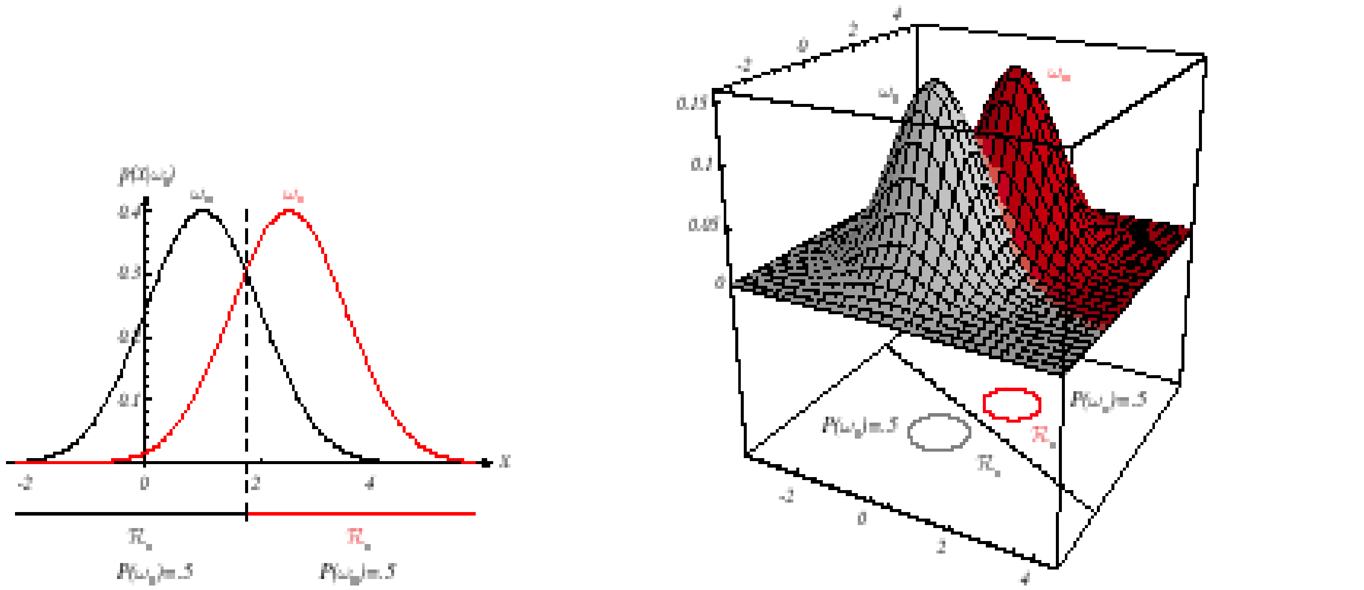
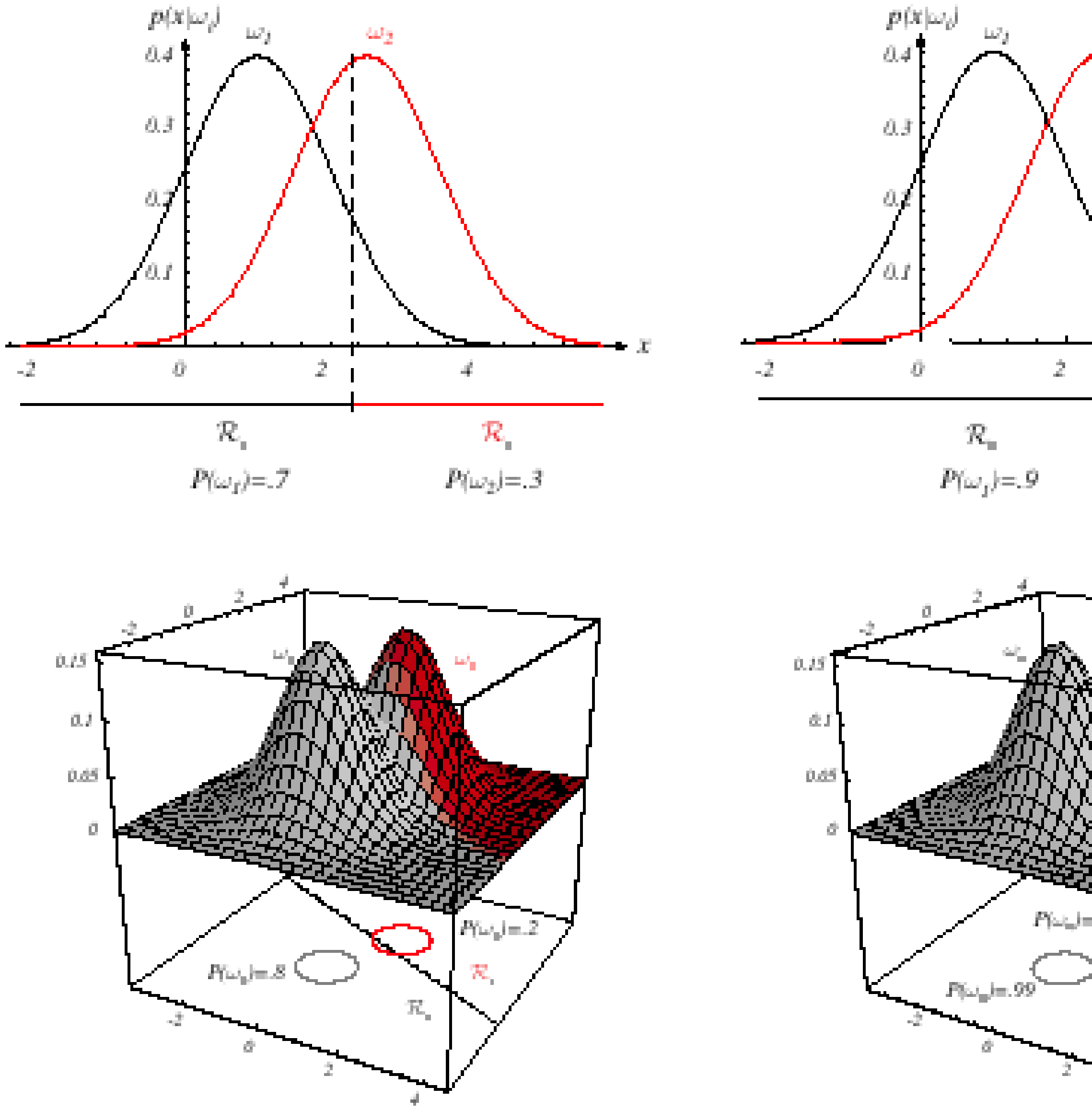


FIGURE 2.10. If the covariance matrices for two distributions are equal and … matrix, then the distributions are spherical in d dimensions, and the boundary is … $d-1$ dimensions, perpendicular to the line separating the means. In these one-, … examples, we indicate $p(\mathbf{x}|\omega_i)$ and the boundaries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

1.8 Equal, Spherical Covariances, Unequal Priors



1.9 Equal, Spherical Covariances, Unequal Priors

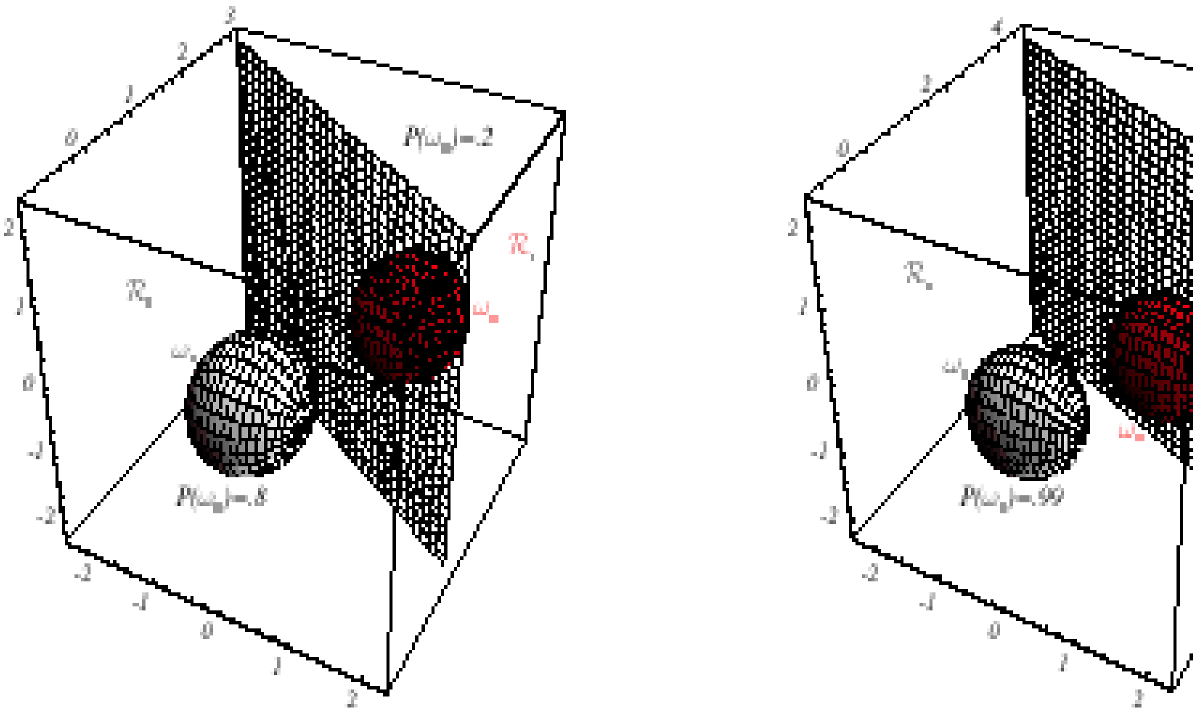
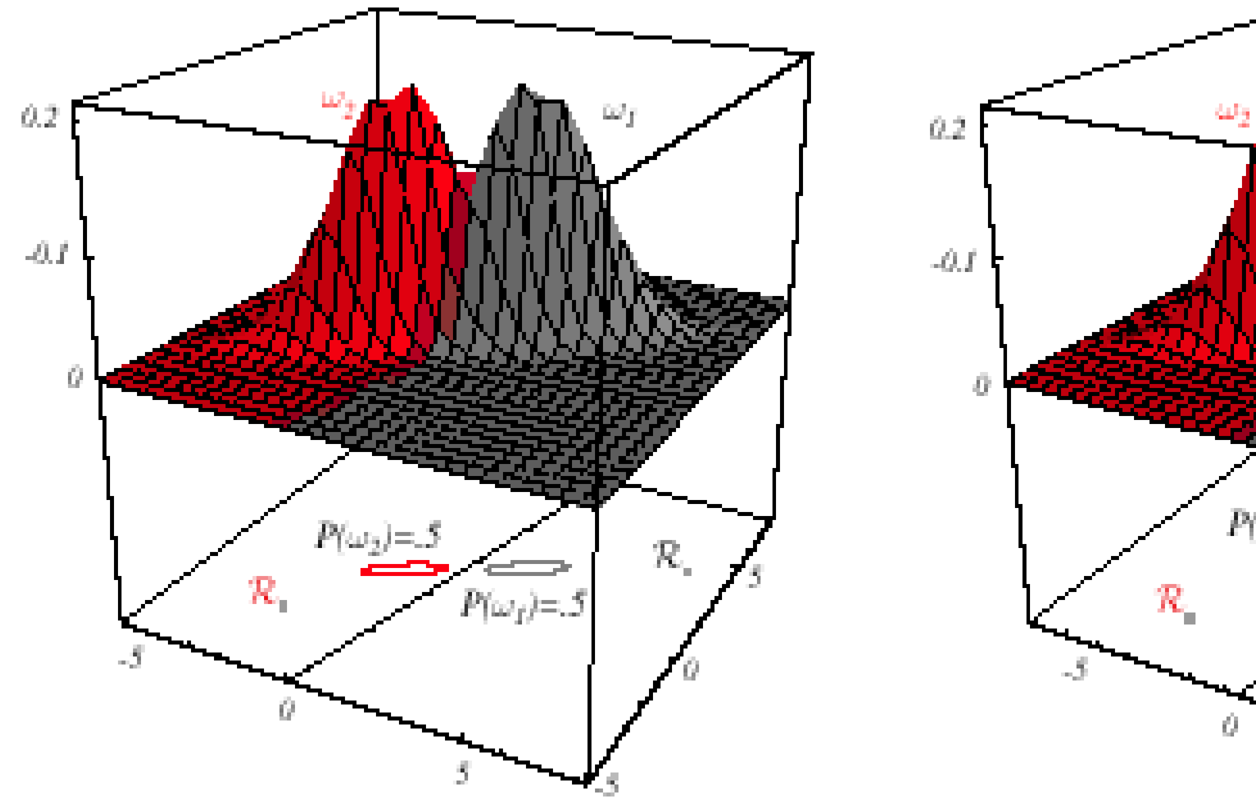


FIGURE 2.11. As the priors are changed, the decision boundary shifts
disparate priors the boundary will not lie between the means of these
three-dimensional spherical Gaussian distributions. From: Richard O.
Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 b
Sons, Inc.

1.10 Equal, Non-Spherical Covariances



1.11 Equal, Non-Spherical Covariances

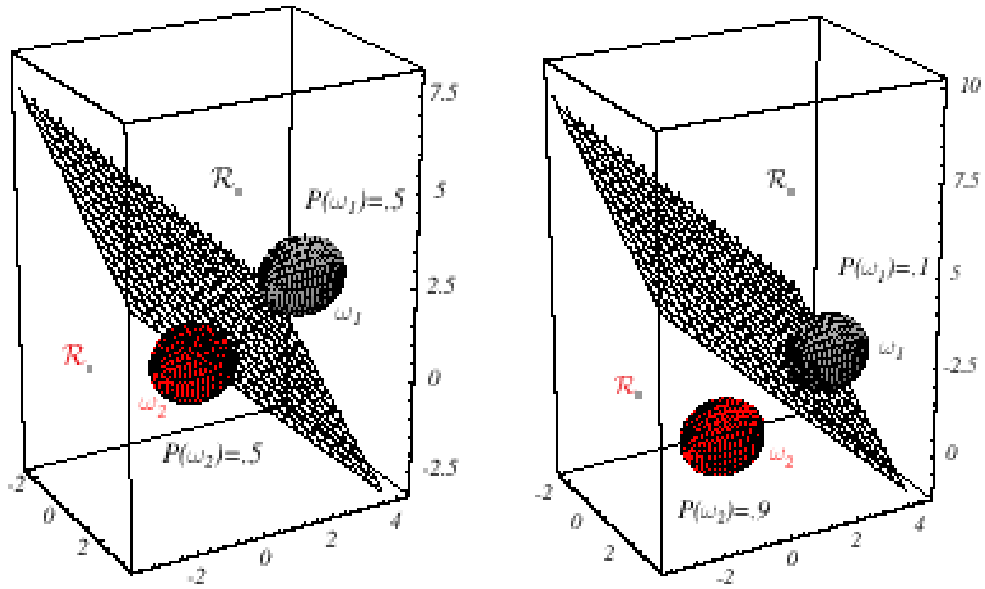
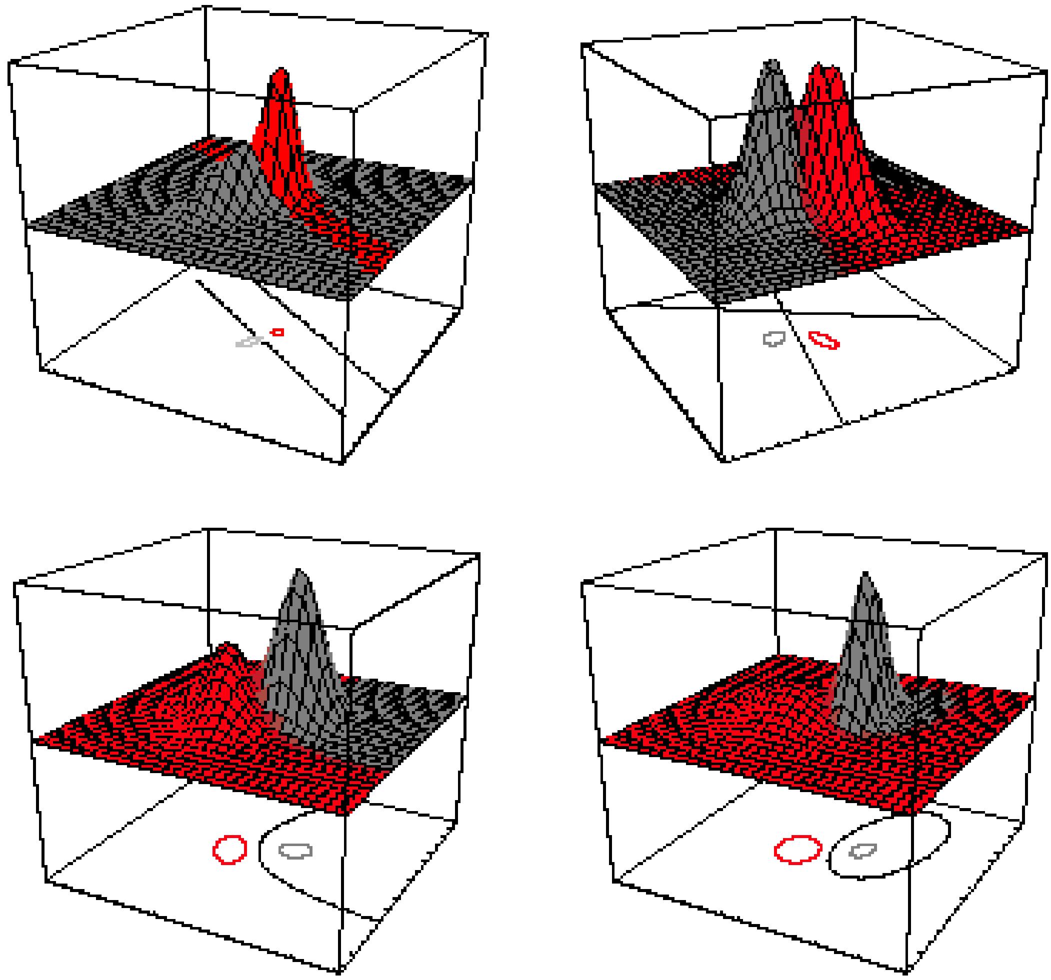


FIGURE 2.12. Probability densities (indicated by the surfaces in two dimensions and ellipsoidal surfaces in three dimensions) and decision regions for equal but asymmetric Gaussian distributions. The decision hyperplanes need not be perpendicular to the line connecting the means. From: Richard O. Duda,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

1.12 Equal, Non-Spherical Covariances



1.13 Arbitrary Covariances

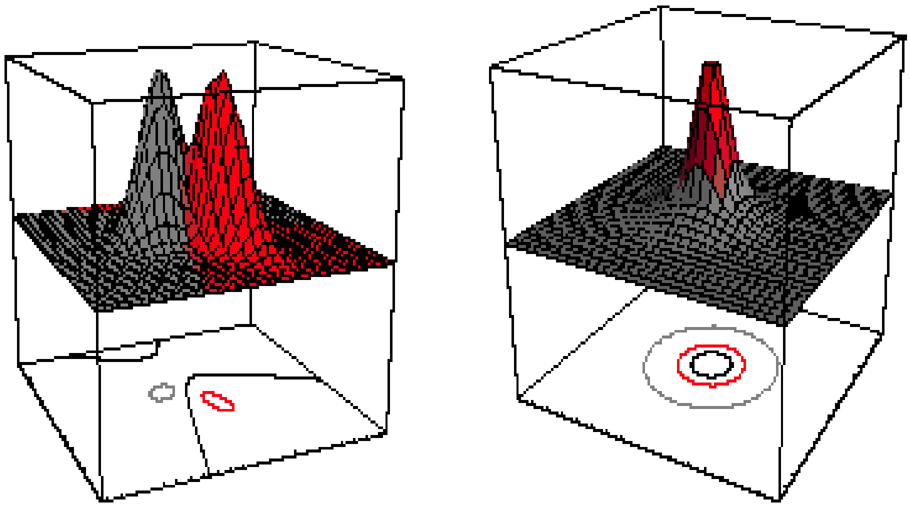


FIGURE 2.14. Arbitrary Gaussian distributions lead to Bayes decision boundaries that are general hyperquadrics. Conversely, given any hyperquadric, one can find two Gaussian distributions whose Bayes decision boundary is that hyperquadric. These variances are indicated by the contours of constant probability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