#### **Bayes Decision Theory**

Discriminant Functions for the Normal Density

# Discriminant Functions for the Normal Density

 We saw that the minimum error-rate classification can be achieved by the discriminant function

$$g_i(x) = \ln P(x \mid \omega_i) + \ln P(\omega_i)$$

Case of multivariate normal

$$g_{i}(x) = -\frac{1}{2}(x - \mu_{i})^{t} \sum_{i}^{-1} (x - \mu_{i}) - \frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\Sigma_{i}| + \ln P(\omega_{i})$$

## Case of Independent Features with same variance

• Case 1:  $\Sigma_i = \sigma^2 I$  (I stands for the identity matrix)

 $g_i(x) = w_i^t x + w_{i0}$  (linear discriminant function) where:

$$w_{i} = \frac{\mu_{i}}{\sigma^{2}}; \ w_{i0} = -\frac{1}{2\sigma^{2}}\mu_{i}^{t}\mu_{i} + \ln P(\omega_{i})$$

( $\omega_{i0}$  is called the threshold for the ith category!)

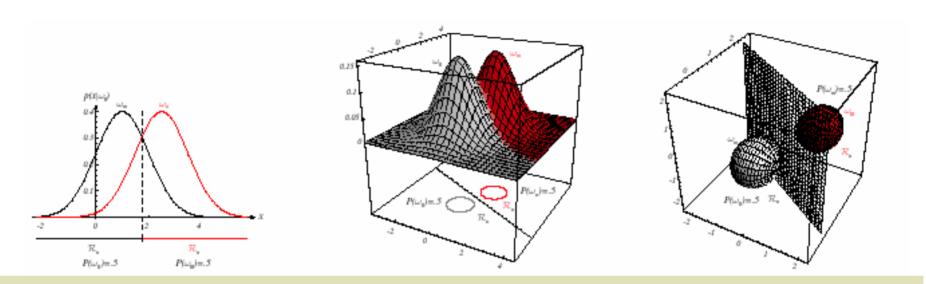
#### Linear Machine

 A classifier that uses linear discriminant functions is called "a linear machine"

 The decision surfaces for a linear machine are pieces of hyperplanes defined by:

$$g_i(x) = g_i(x)$$

## Hyperplane Decision Boundaries



If 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 in d-1 dimensions, perpendicular to the line joining the means

#### Hyperplane Equation

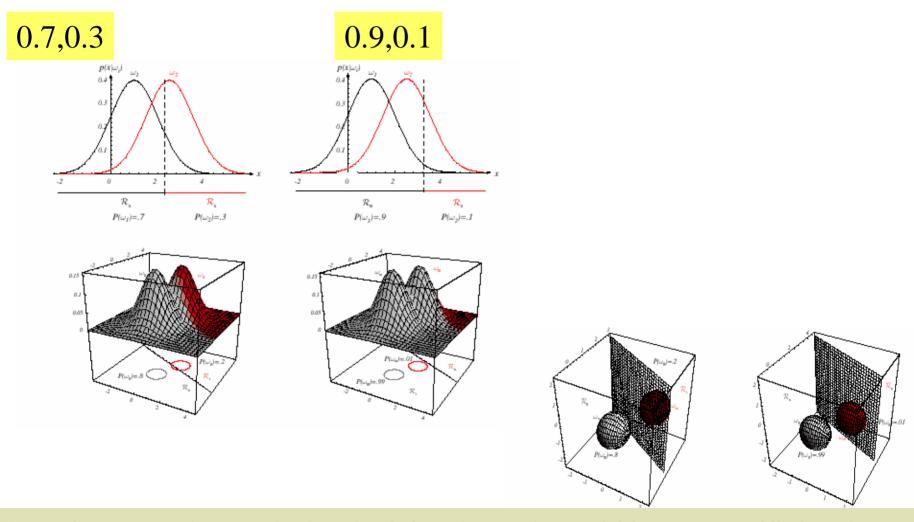
ullet The hyperplane separating  $R_i$  and  $R_i$ 

$$x_{0} = \frac{1}{2}(\mu_{i} + \mu_{j}) - \frac{\sigma^{2}}{\|\mu_{i} - \mu_{j}\|^{2}} \ln \frac{P(\omega_{i})}{P(\omega_{j})} (\mu_{i} - \mu_{j})$$

always orthogonal to the line linking the means

if 
$$P(\omega_i) = P(\omega_j)$$
 then  $x_0 = \frac{1}{2}(\mu_i + \mu_j)$ 

#### Influence of Priors on Decision Boundaries



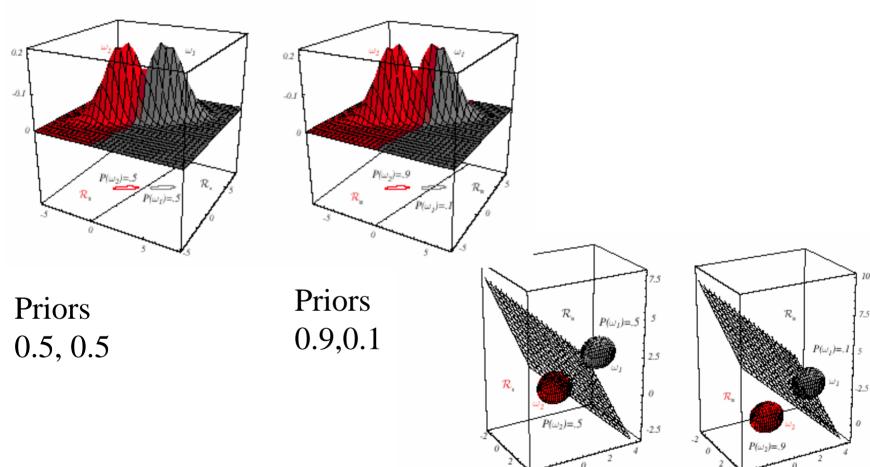
As priors are changed, the decision boundary shifts. For sufficiently disparate priors the boundary will not lie between the means.

- Case 2:  $\Sigma_i = \Sigma$  (covariance of all classes are identical but arbitrary!)
  - ullet Hyperplane separating  $R_i$  and  $R_j$

$$x_{0} = \frac{1}{2}(\mu_{i} + \mu_{j}) - \frac{\ln[P(\omega_{i})/P(\omega_{j})]}{(\mu_{i} - \mu_{j})^{t} \Sigma^{-1}(\mu_{i} - \mu_{j})}.(\mu_{i} - \mu_{j})$$

(the hyperplane separating  $R_i$  and  $R_j$  is generally not orthogonal to the line between the means!)

## Decision Regions for Equal but Asymmetric Gaussians



Decision hyperplanes need not be perpendicular to line joining the means

#### • Case 3: $\Sigma_i$ = arbitrary

The covariance matrices are different for each category

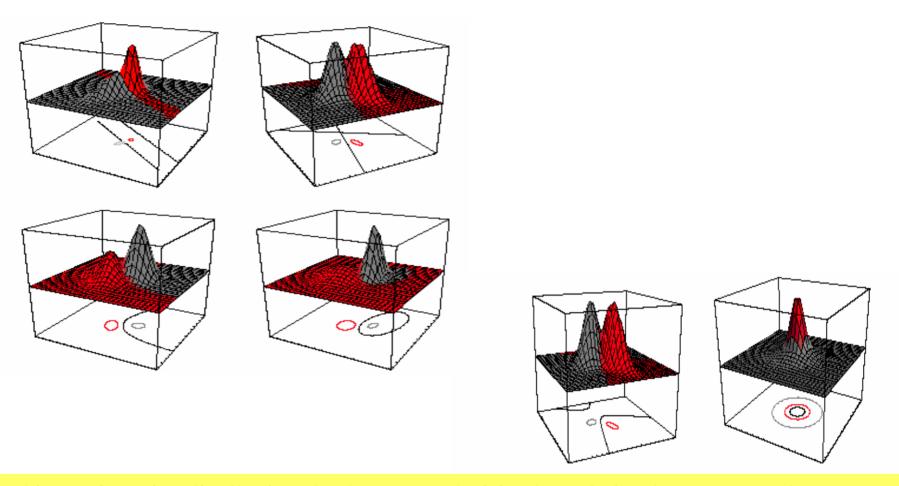
$$g_{i}(x) = x^{t}W_{i}x + w_{i}^{t}x = w_{i0}$$
where:
$$W_{i} = -\frac{1}{2}\Sigma_{i}^{-1}$$

$$w_{i} = \Sigma_{i}^{-1}\mu_{i}$$

$$w_{i0} = -\frac{1}{2}\mu_{i}^{t}\Sigma_{i}^{-1}\mu_{i} - \frac{1}{2}\ln|\Sigma_{i}| + \ln P(\omega_{i})$$

(Hyperquadrics which are: hyperplanes, pairs of hyperplanes, hyperspheres, hyperellipsoids, hyperparaboloids, hyperhyperboloids)

### Hyperquadric Decision Boundaries



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.

Variances are indicated by contours of constant density.