JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Applied and Computational Mathematics

Data Mining

625.740

Normal Density Discriminant Functions

Mike Weisman

email: `data.mining.625.740@gmail.com`

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

Applied and Computational Mathematics

Data Mining

625.740

Normal Density Discriminant Functions

Normal Density Discriminant Function

Recall that for minimum-error-rate classification, we can use the discriminant function

$$g_j(\mathbf{x}) = \log p(\mathbf{x}|\omega_j) + \log P(\omega_j);$$

If $p(\mathbf{x}|\omega_j) \sim \mathscr{N}(\mu_j, \Sigma_j)$, then

$$p(\mathbf{x}|\omega_j) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma_j|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu_j)^T\Sigma_j^{-1}(\mathbf{x}-\mu_j)\right] \quad \text{and}$$

$$g_j(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\mu_j)^T\Sigma_j^{-1}(\mathbf{x}-\mu_j) - \frac{n}{2}\log 2\pi - \frac{1}{2}\log|\Sigma_j| + \log P(\omega_j)$$

Case I: $\Sigma_j = \sigma^2 I$

$$g_j(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\mu_j)^T(\mathbf{x}-\mu_j) + \log P(\omega_j)$$

$$g_j(\mathbf{x}) = -\frac{1}{2\sigma^2}[\mathbf{x}^T\mathbf{x} - 2\mu_j^T\mathbf{x} + \mu_j^T\mu_j] + \log P(\omega_j)$$

or equivalently

$$g_j(\mathbf{x}) = \mathbf{w}_j^T\mathbf{x} + w_{j0}$$

where

$$\mathbf{w}_j = \frac{\mu_j}{\sigma^2}$$

and

$$w_{j0} = -\frac{1}{2\sigma^2}\mu_j^T\mu_j + \log P(\omega_j)$$

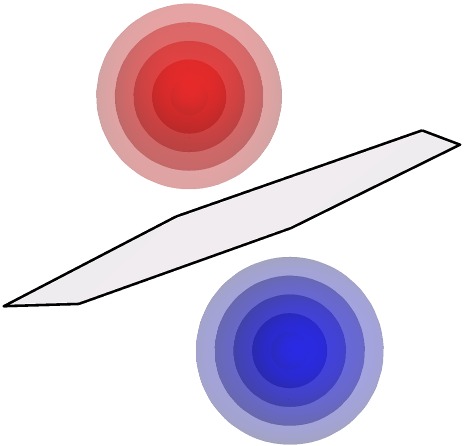

Figure: Case I

Case II: $\Sigma_j = \Sigma$

$$g_j(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \mu_j)^T \Sigma^{-1} (\mathbf{x} - \mu_j) + \log P(\omega_j)$$

or equivalently

$$g_j(\mathbf{x}) = \mathbf{w}_j^T \mathbf{x} + w_{j0}$$

$$\text{where } \mathbf{w}_j = \Sigma^{-1} \mu_j$$

$$\text{and } w_{j0} = -\frac{1}{2}\mu_j^T \Sigma^{-1} \mu_j + \log P(\omega_j)$$

The separating hyperplane between Region i and Region j has the equation

$$\mathbf{w}^T(\mathbf{x} - \mathbf{x}_0) = 0$$

$$\text{where } \mathbf{w} = \Sigma^{-1}(\mu_i - \mu_j)$$

$$\text{and } \mathbf{x}_0 = -\frac{1}{2}(\mu_i + \mu_j) - \frac{\log P(\omega_i) - \log P(\omega_j)}{(\mu_i - \mu_j)^T \Sigma^{-1} (\mu_i - \mu_j)}(\mu_i - \mu_j)$$

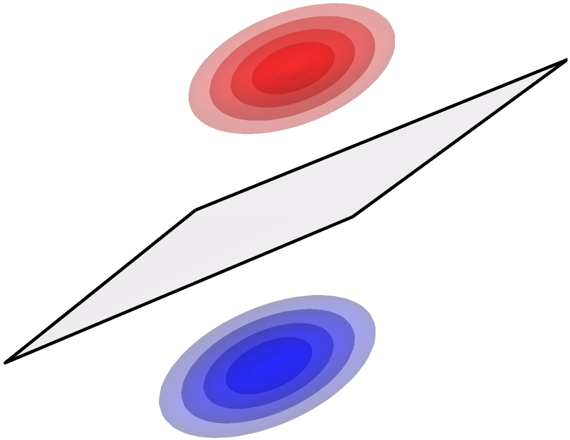

Figure: Case II

Case III: Σ_j Arbitrary

$$g_j(\mathbf{x}) = \mathbf{x}^T W_j \mathbf{x} + \mathbf{w}_j^T \mathbf{x} + w_{j0}$$

where

$$W_j = -\frac{1}{2}\Sigma_j^{-1},$$

$$\mathbf{w}_j = \Sigma^{-1}\mu_j,$$

and

$$w_{j0} = -\frac{1}{2}\mu_j^T \Sigma^{-1}\mu_j - \frac{1}{2}\log|\Sigma_j| + \log P(\omega_j)$$

The decision boundaries are hyperquadrics: pairs of hyperplanes, hyperspheres, hyperellipsoids, hyperparaboloids, and hyperhyperboloids.

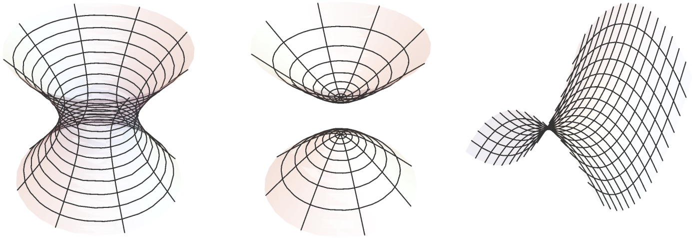

Figure: Case III, Examples of boundaries that may occur for arbitrary Σ_j: hyperboloids of one and two sheets, hyperbolic paraboloid

JOHNS HOPKINS
WHITING SCHOOL of ENGINEERING

Applied and Computational Mathematics

Data Mining

625.740

Normal Density Discriminant Functions