Today's Discussion

Review from last time (Sect. 2.1, 2.4)

- Bayesian decision theory
- Discriminant functions

Normal density: (Sect. 2.5, 2.6)

- Normal density function
- Discriminant functions for normal densities

Maximum likelihood estimation: (Sect. 3.1, 3.2)

Review from last time

- **Assume knowledge of** $P(\omega_i)$ **and** $p(\mathbf{x}|\omega_i)$ **for all classes.**
- ***Bayes Optimal Classifier*:**

Decide ω_i **if** $P(\omega_i|\mathbf{x}) > P(\omega_j|\mathbf{x})$, $\forall j \neq i$ **where,**

$$P(\omega_i|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_i)P(\omega_i)}{p(\mathbf{x})}$$

$$p(\mathbf{x}) = \sum_{j=1}^{k} p(\mathbf{x}|\omega_j)P(\omega_j)$$

Discriminant functions

Three choices:

#1: $g_i(\mathbf{x}) = P(\omega_i|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_i)P(\omega_i)}{p(\mathbf{x})}$

#2: $g_i(\mathbf{x}) = p(\mathbf{x}|\omega_i)P(\omega_i)$

#3: $g_i(\mathbf{x}) = \ln p(\mathbf{x}|\omega_i) + \ln P(\omega_i)$

Bayes Decision Rule:

Decide ω_i **if** $g_i(\mathbf{x}) > g_j(\mathbf{x})$, $\forall j \neq i$

Two-category case

Let:

$$g(\mathbf{x}) \equiv g_1(\mathbf{x}) - g_2(\mathbf{x})$$

Bayes Optimal Classifier (*Dichotomizer*):

Decide ω_1 **if** $g(\mathbf{x}) > 0$ **; else decide** ω_2 **.**

$$g(\mathbf{x}) = P(\omega_1|\mathbf{x}) - P(\omega_2|\mathbf{x})$$

$$g(\mathbf{x}) = \ln\frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} + \ln\frac{P(\omega_1)}{P(\omega_2)}$$

Normal (Gaussian) density

One-dimensional (univariate) density:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

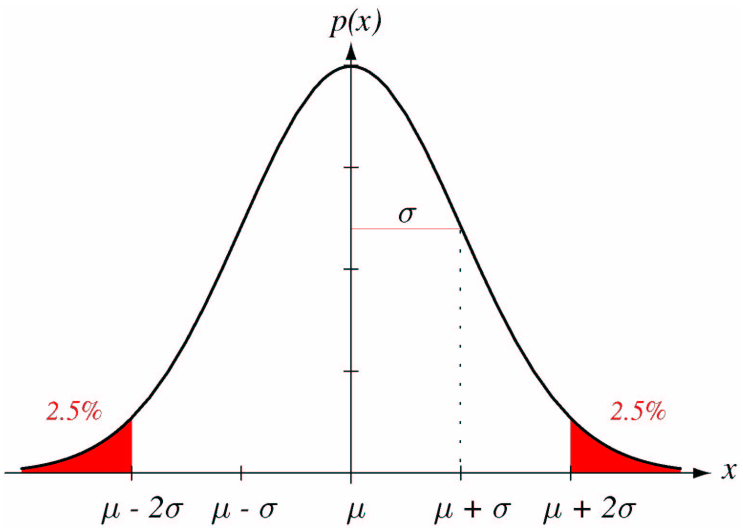
$$\mu \equiv E[x] = \int_{-\infty}^{\infty} x p(x) dx \text{ = mean}$$

$$\sigma^2 = E[(x-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 p(x) dx \text{ = variance}$$

σ **= standard deviation**

Let's confirm integrals in *Mathematica* ...

Univariate density



Multivariate normal density

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}}\exp\left[-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right]$$

$$\mu \equiv E[\mathbf{x}] = \int \mathbf{x} p(\mathbf{x}) d\mathbf{x} \text{ = mean vector}$$

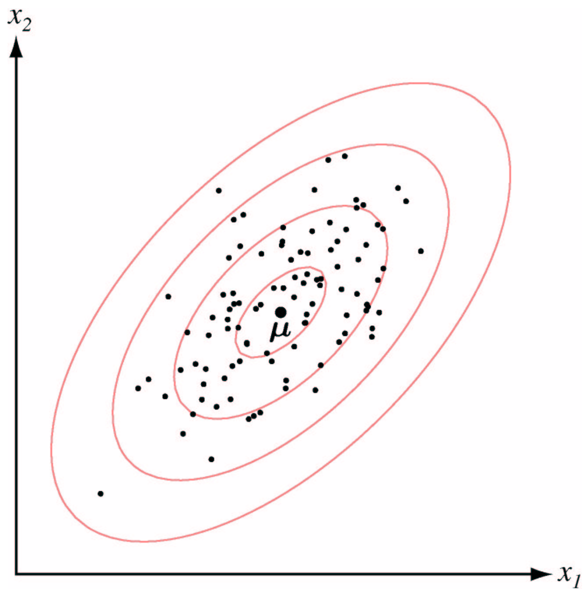
$$\Sigma \equiv E[(\mathbf{x}-\mu)(\mathbf{x}-\mu)^T] = \int (\mathbf{x}-\mu)(\mathbf{x}-\mu)^T p(\mathbf{x}) d\mathbf{x} =$$

covariance matrix (symmetric, $\Sigma \geq 0$)

d **= dimension**

$p(\mathbf{x}) \sim N(\mu, \Sigma)$ **(alternate notation)**

Multivariate density example



Discriminant functions for normal density

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

$$p(\mathbf{x}|\omega_i) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu_i)^T\Sigma_i^{-1}(\mathbf{x}-\mu_i)\right]$$

$$\ln p(\mathbf{x}|\omega_i) = -\frac{1}{2}(\mathbf{x}-\mu_i)^T\Sigma_i^{-1}(\mathbf{x}-\mu_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_i|$$

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

Case 1: $\Sigma_i = \sigma^2\mathbf{I}$

Equal, uniform covariance matrices:

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

$$|\Sigma_i| = \sigma^{2d}, \quad \Sigma_i^{-1} = \frac{1}{\sigma^2}\mathbf{I}$$

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

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

Discriminant function (Case 1)

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

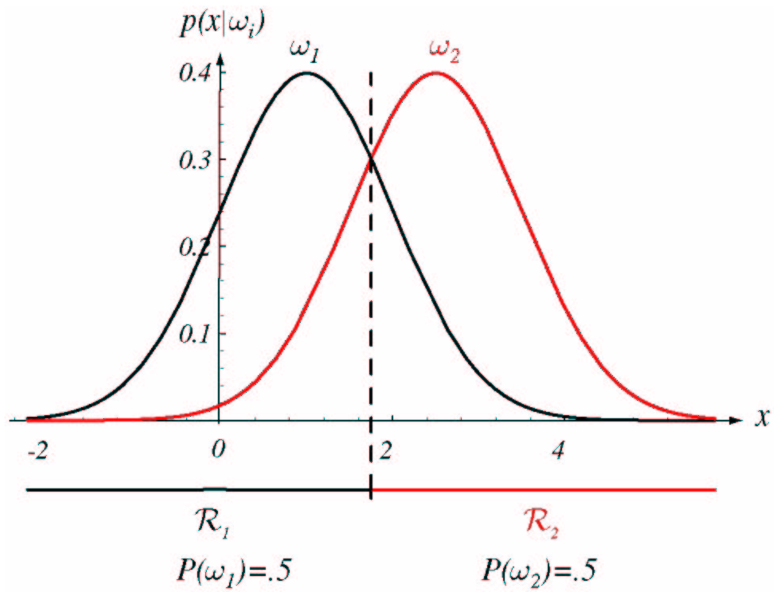
$$g_i(\mathbf{x}) = \frac{-1}{2\sigma^2}[\mathbf{x}^T\mathbf{x} - 2\mu_i^T\mathbf{x} + \mu_i^T\mu_i] + \ln P(\omega_i)$$

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

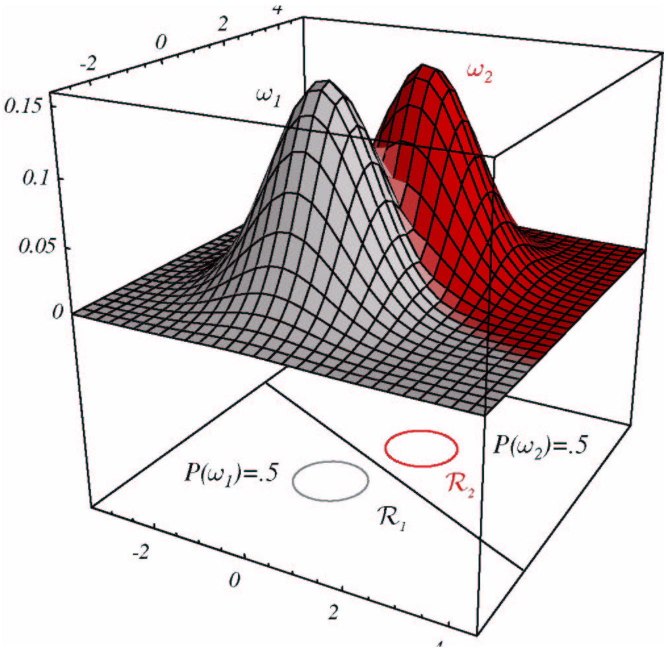
$g_i(\mathbf{x}) = \mathbf{w}_i^T\mathbf{x} + w_{i0}$ **=> linear discriminant function**

$$\mathbf{w}_i = \frac{\mu_i}{\sigma^2}, \quad w_{i0} = -\frac{\mu_i^T\mu_i}{2\sigma^2} + \ln P(\omega_i)$$

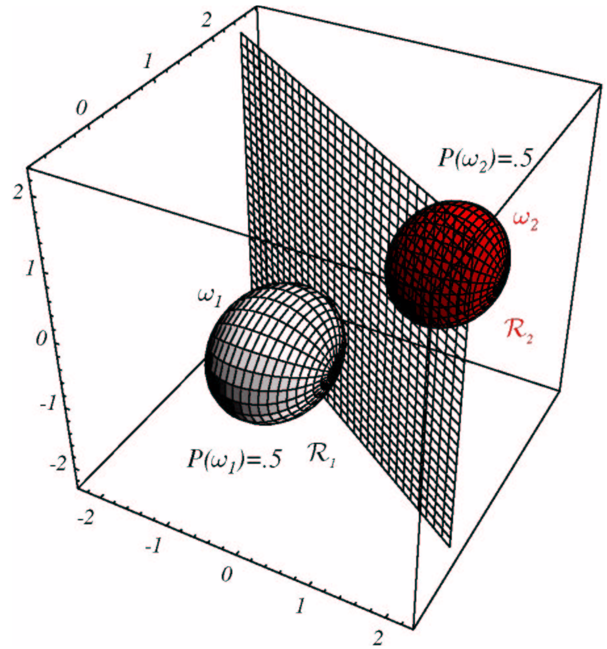
Examples (Case 1)



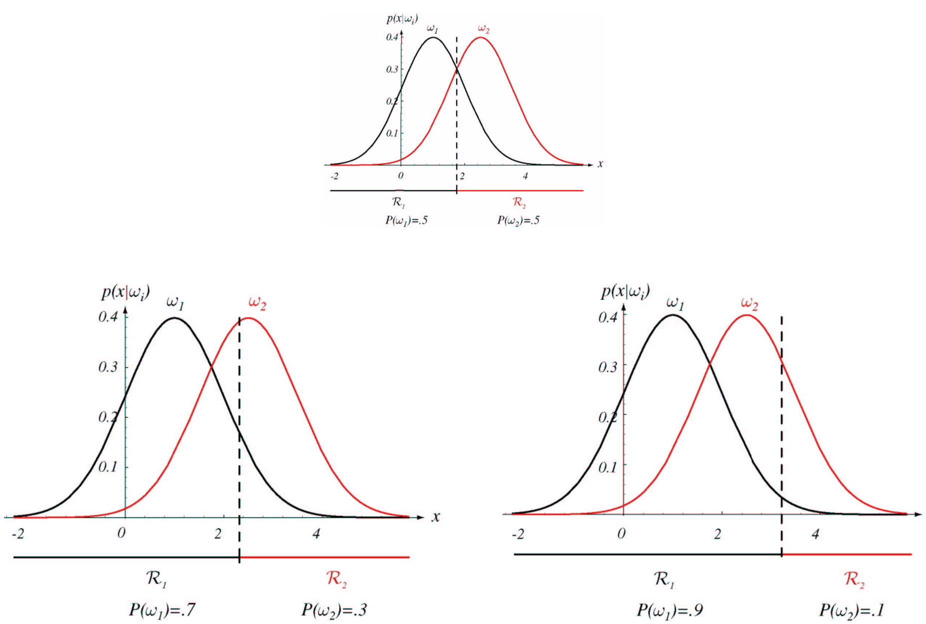
Examples (Case 1)



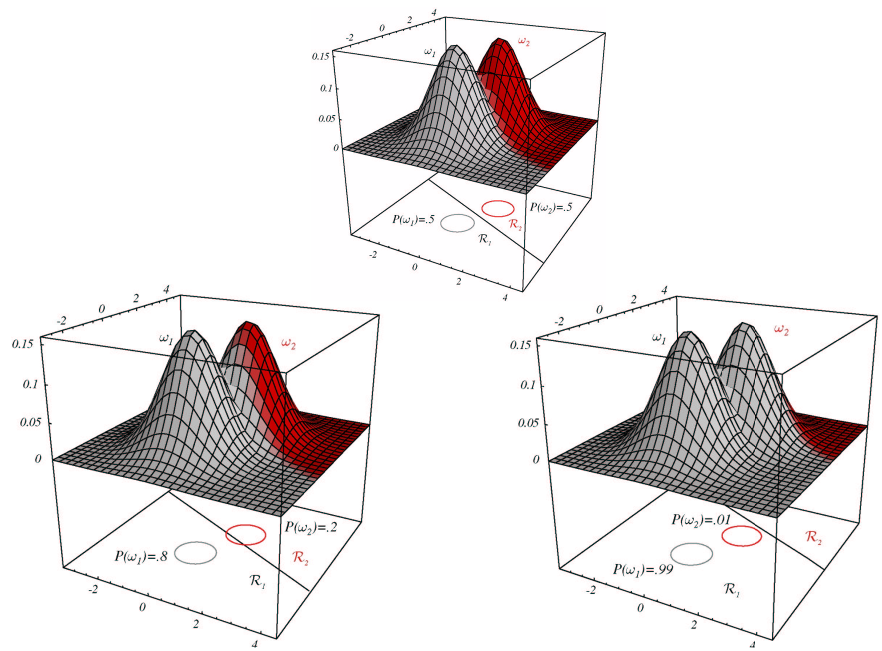
Examples (Case 1)



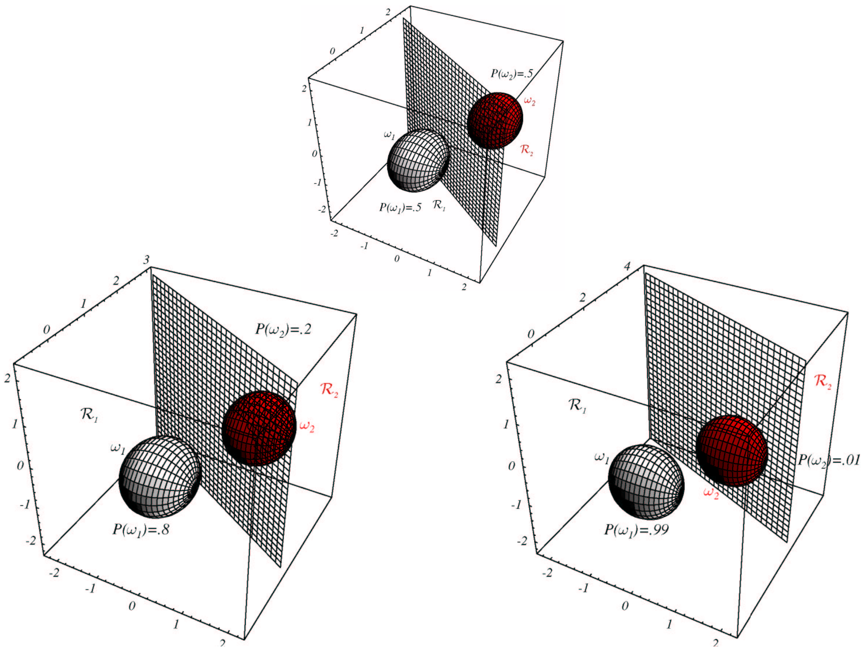
Non-equal priors (Case 1)



Non-equal priors (Case 1)



Non-equal priors (Case 1)



Case 2: $\Sigma_i = \Sigma$

Equal covariance matrices:

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

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

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

Discriminant function (Case 2)

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

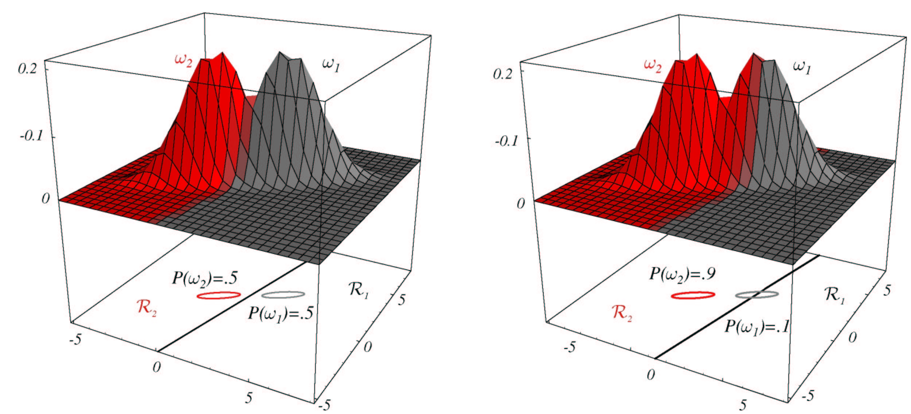
$$g_i(\mathbf{x}) = -\frac{1}{2}[\mathbf{x}^T\Sigma^{-1}\mathbf{x} - 2\mu_i^T\Sigma^{-1}\mathbf{x} + \mu_i^T\Sigma^{-1}\mu_i] + \ln P(\omega_i)$$

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

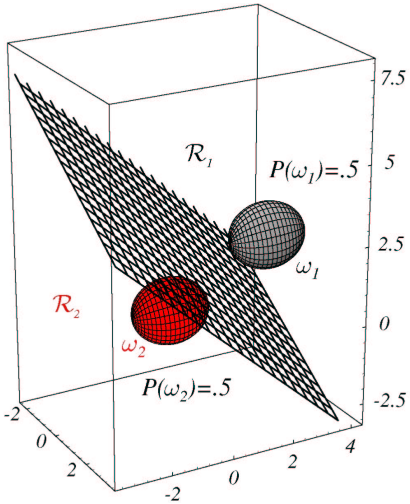
$g_i(\mathbf{x}) = \mathbf{w}_i^T\mathbf{x} + w_{i0}$ **=> linear discriminant function**

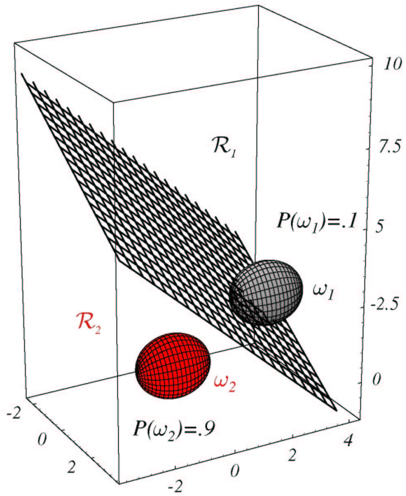
$$\mathbf{w}_i = \Sigma^{-1}\mu_i, \quad w_{i0} = -\frac{1}{2}\mu_i^T\Sigma^{-1}\mu_i + \ln P(\omega_i)$$

Examples (Case 2)



Examples (Case 2)





Case 3: Σ_i = arbitrary

Arbitrary covariance matrices:

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

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

Discriminant function (Case 3)

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

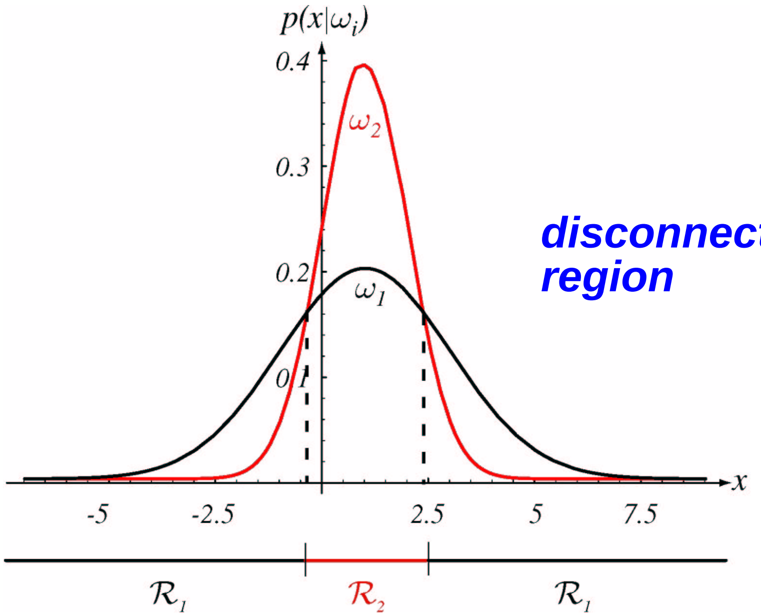
Quadratic discriminant function:

$$g_i(\mathbf{x}) = \mathbf{x}^T\mathbf{W}_i\mathbf{x} + \mathbf{w}_i^T\mathbf{x} + w_{i0}$$

$$\mathbf{W}_i = -\frac{1}{2}\Sigma_i^{-1}, \quad \mathbf{w}_i = \Sigma_i^{-1}\mu_i$$

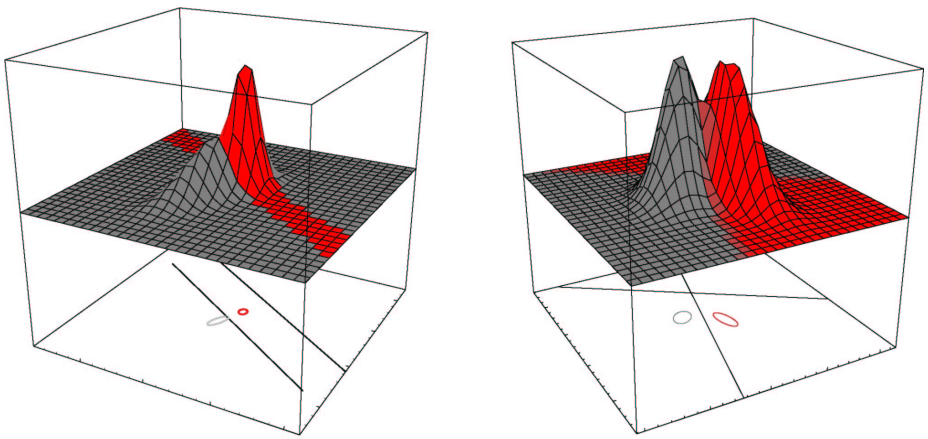
$$w_{i0} = -\frac{1}{2}\mu_i^T\Sigma_i^{-1}\mu_i - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i)$$

Case 3 examples

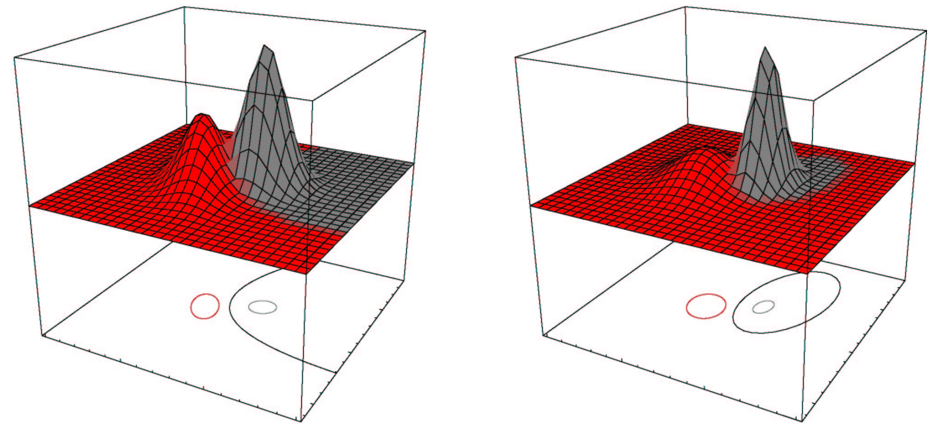


disconnected region

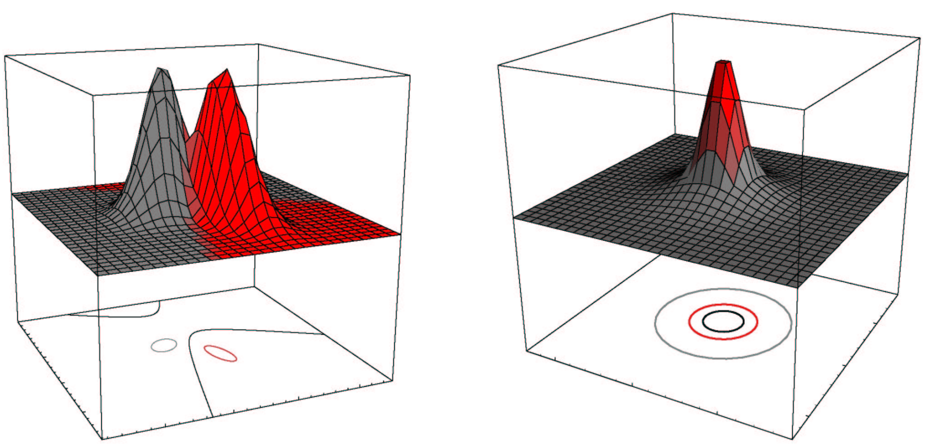
Case 3 examples (2d)



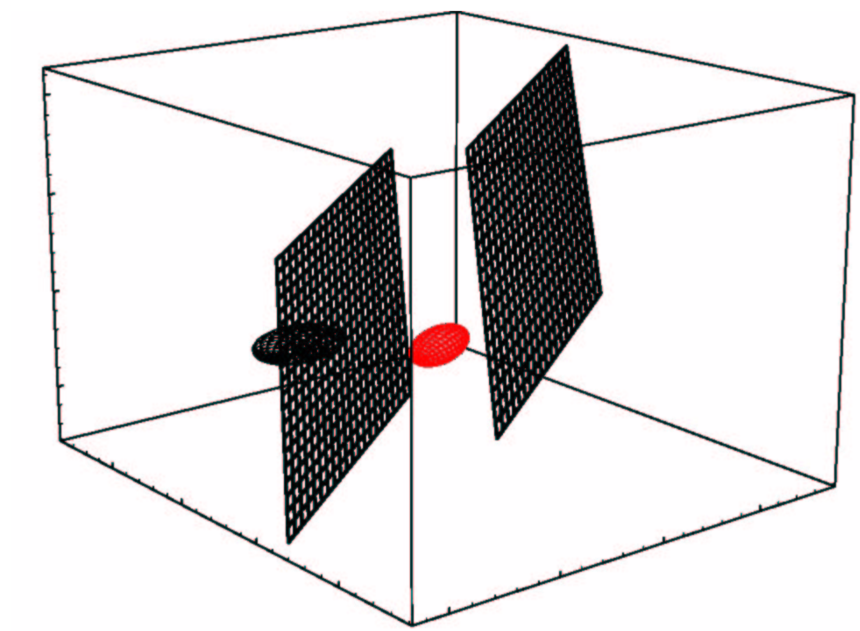
Case 3 examples (2d)



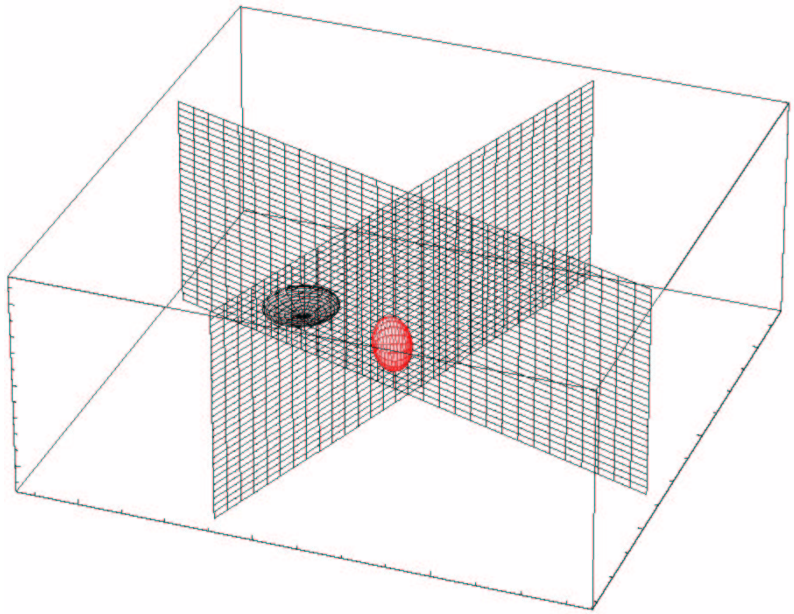
Case 3 examples (2d)



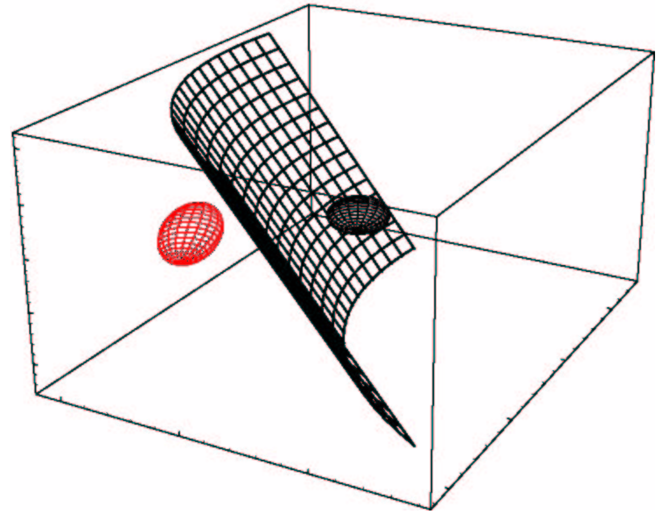
Case 3 examples (3d)



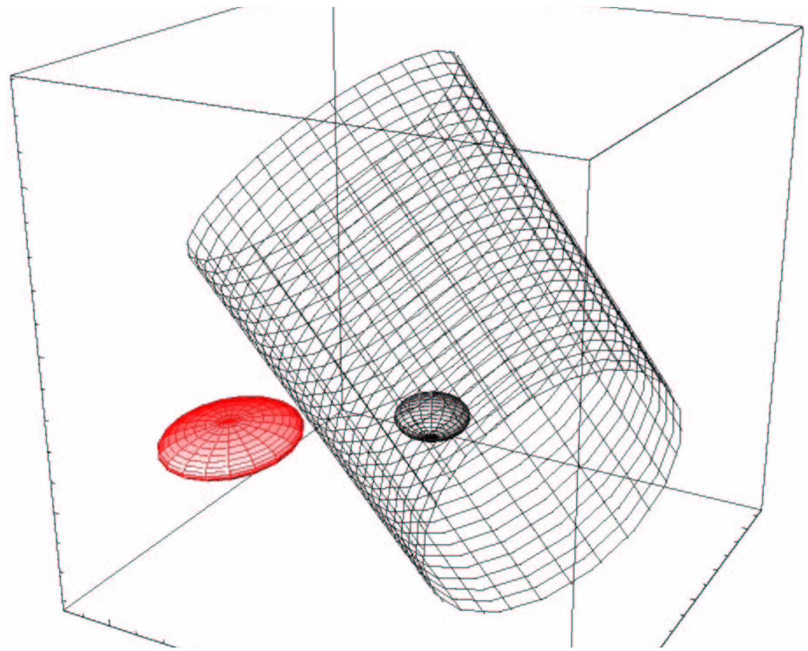
Case 3 examples (3d)



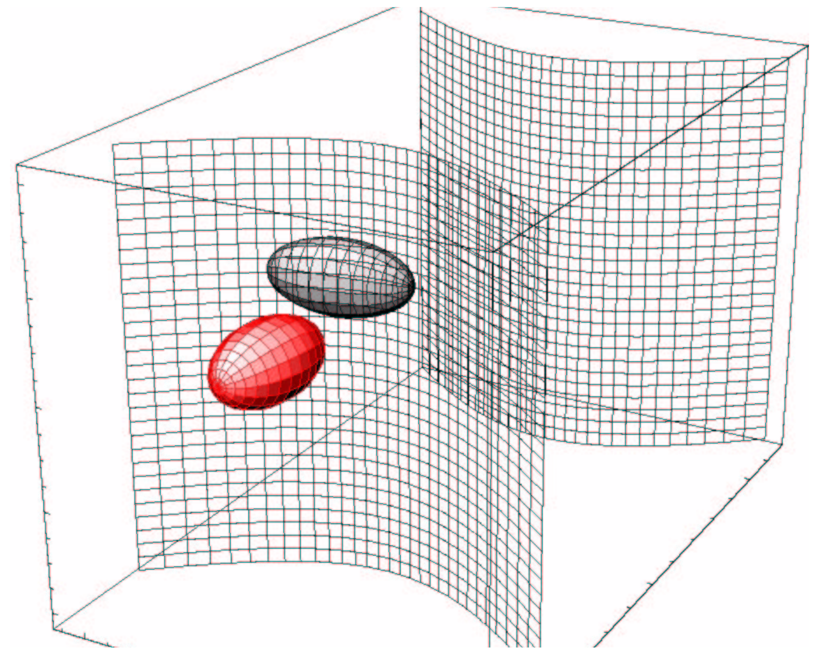
Case 3 examples (3d)



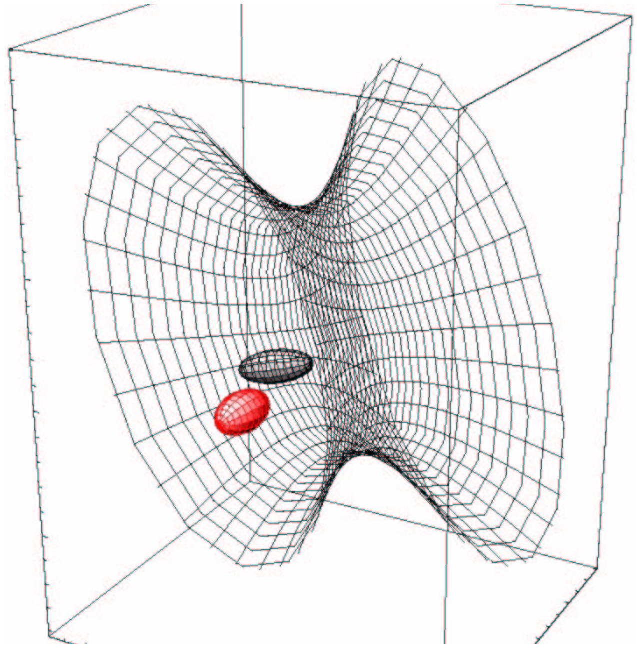
Case 3 examples (3d)



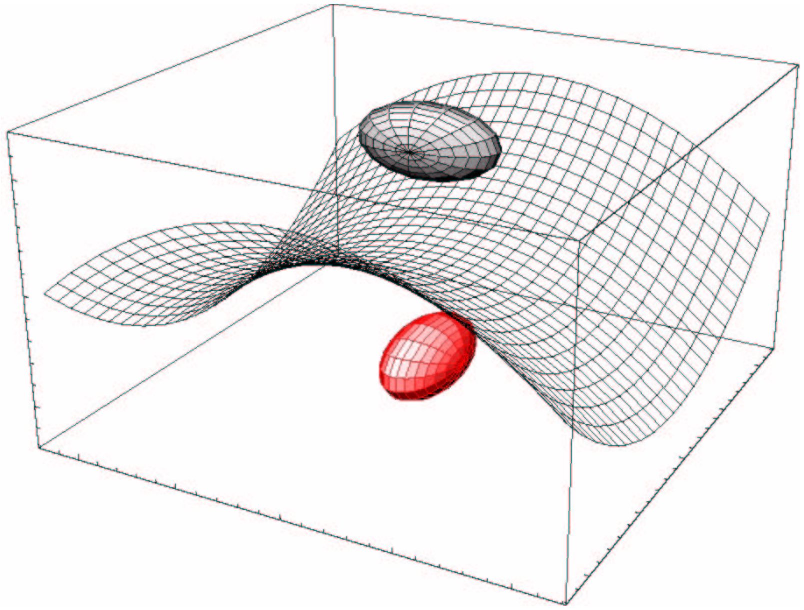
Case 3 examples (3d)



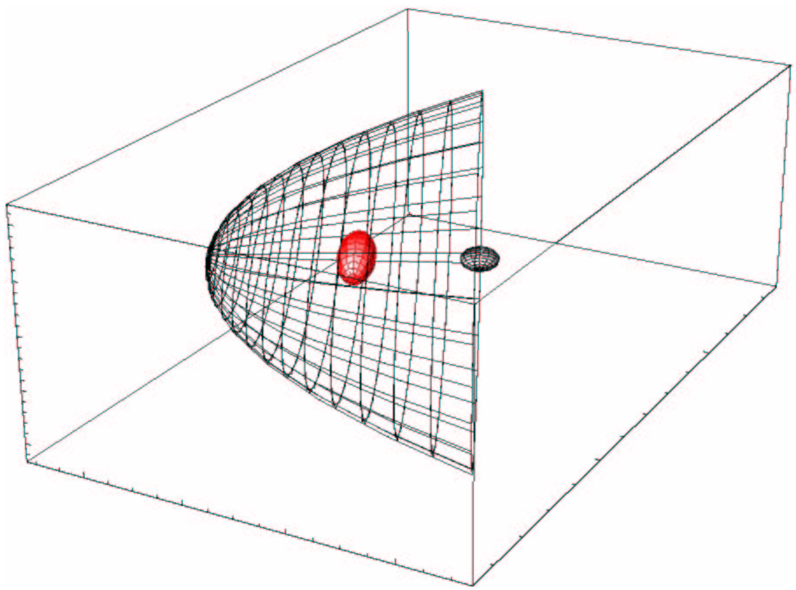
Case 3 examples (3d)



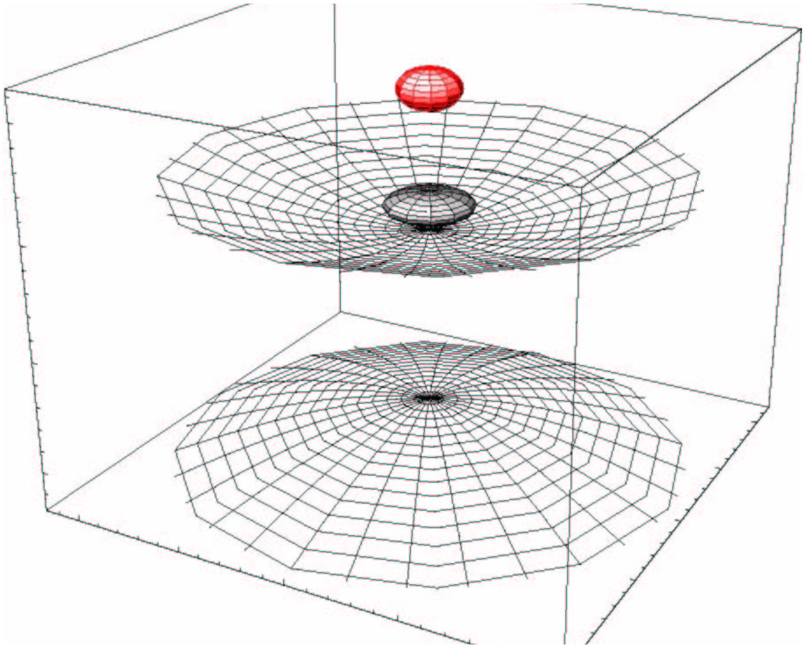
Case 3 examples (3d)



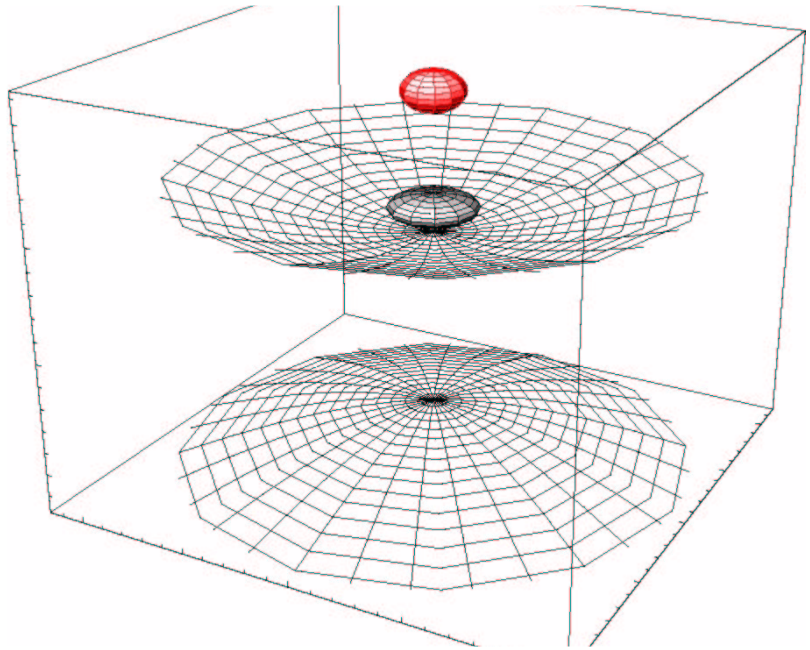
Case 3 examples (3d)



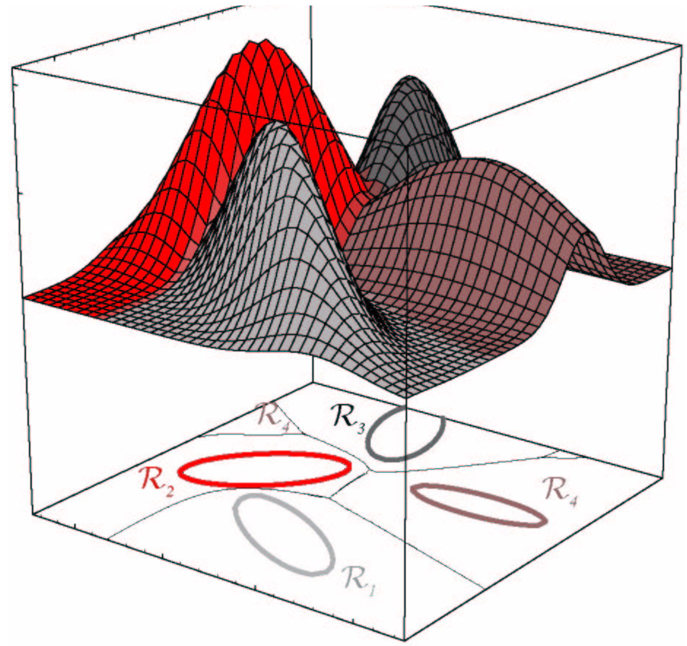
Case 3 examples (3d)



Case 3 examples (3d)



Multiple regions



Example of decision boundary derivation

Two normal-distributed classes: ω_1 **,** ω_2 **(equal priors)**

$$\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix},\ \Sigma_1 = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \quad — \quad \mu_2 = \begin{bmatrix} 0 \\ 2 \end{bmatrix},\ \Sigma_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

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

$$g_1(\mathbf{x}) = -\frac{1}{2}\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} 1/2 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} - \frac{1}{2}\ln 2$$

$$g_2(\mathbf{x}) = -\frac{1}{2}\begin{bmatrix} x & (y-2) \end{bmatrix}\begin{bmatrix} x \\ (y-2) \end{bmatrix}$$

Example of decision boundary derivation

$$g_1(\mathbf{x}) = g_2(\mathbf{x})$$

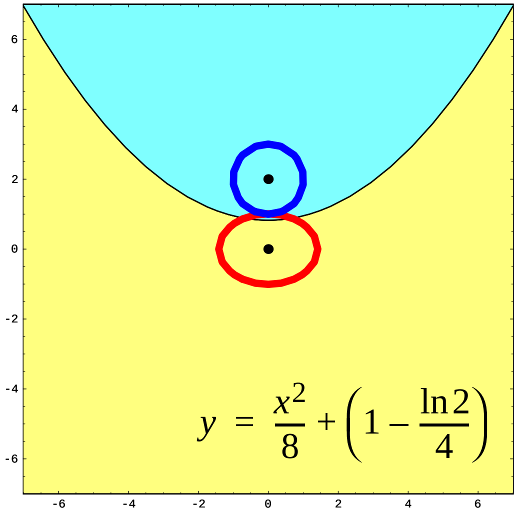
$$-\frac{1}{2}\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} 1/2 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} - \frac{1}{2}\ln 2 = -\frac{1}{2}\begin{bmatrix} x & (y-2) \end{bmatrix}\begin{bmatrix} x \\ (y-2) \end{bmatrix}$$

$$\begin{bmatrix} x & y \end{bmatrix}\begin{bmatrix} 1/2 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} + \ln 2 = \begin{bmatrix} x & (y-2) \end{bmatrix}\begin{bmatrix} x \\ (y-2) \end{bmatrix}$$

$$x^2/2 + y^2 + \ln 2 = x^2 + (y^2 - 4y + 4)$$

$$y = \frac{x^2}{8} + \left(1 - \frac{\ln 2}{4}\right)$$ **(parabola)**

Example of decision boundary derivation



Other derivations: *Mathematica*...