Detection

- Detection
- Unconditional error matrix
- MAP estimators
- Gaussian detection
- Signal-to-noise ratio
- Operating characteristics and SNR
- A-posteriori probabilities
- The sigmoid function
- General Gaussian classification
- Decision regions are quadratic
- Example: Gaussian decision regions
- Equal covariances and linear decision boundaries

Detection

Let's look at the classification problem with *continuous measurements*. This is often called the *detection* problem.

- x is a discrete random variable, taking values in $\Omega^x = \{x_1, \dots, x_n\} \subset \mathbb{R}^m$
- y is a continuous random variable taking values in $\mathbb{R}^m$

We know

- The *hypothesis events*: X_j is the event that $x = x_j$
- The *transition pdfs* $A(y, j)$. For each $j = 1, \dots, n$ we have

$$\mathbf{Prob}(y \in U \mid X_j) = \int_U A(z, j)\, dz \qquad \text{for all } U \subset \mathbb{R}^m$$

- The *prior probabilities*: p is the pmf for x

$$p_j = \mathbf{Prob}(X_j)$$

Detection

We can calculate the *joint pdfs* $J(y, j)$ by

$$\textbf{Prob}(y \in U \text{ and } X_j) = \int_U J(z, j)\, dz$$

and hence

$$J(z, j) = A(z, j) p_j$$

As before, we specify an estimator $f_{\text{est}} : \mathbb{R}^m \to \{1, \ldots, n\}$

> if we observe $y = y_{\text{meas}}$, we estimate $x = x_j$ where $j = f_{\text{est}}(y_{\text{meas}})$

We can also specify f_{est} by the function $K : \mathbb{R}^m \times \{1, \ldots, n\}$ where

$$K(y, j) = \begin{cases} 1 & \text{if } f_{\text{est}}(y) = j \\ 0 & \text{otherwise} \end{cases}$$

The unconditional error matrix

The unconditional error matrix $E \in \mathbb{R}^{n\times n}$ is

$$\begin{aligned} E_{jk} &= \text{probability that } X_j \text{ is estimated and } X_k \text{ occurs} \\ &= \mathbf{Prob}(f_{\text{est}}(y) = j \text{ and } x = x_k) \\ &= \int_{R_j} J(y,k)\,dy \end{aligned}$$

where R_j is the *decision region* for X_j

$$R_j = \{\, y \in \mathbb{R}^m \mid f_{\text{est}}(y) = j \,\}$$

We can also write this as

$$E_{jk} = \int_{R^m} K(y,j) J(y,k)\,dy$$

The MAP estimator

The probability that the estimate is correct is

$$
\begin{aligned}
\sum_{j=1}^{n} E_{jj} &= \mathbf{trace}\, E \\
&= \sum_{j=1}^{n} \int_{R^m} K(y,j) J(y,k)\, dy
\end{aligned}
$$

Hence to maximize the probability of a correct estimate, we pick K to that

$$
K(y,j) = \begin{cases} 1 & \text{if } j = \arg\max_k J(y,k) \\ 0 & \text{otherwise} \end{cases}
$$

This is the *MAP classifier*; pick the estimate with the greatest a-posteriori probability.

Example: Gaussian detection

Suppose we measure y, and would like to determine x, where

$$y = x + w$$

here

- x is a discrete random variable, taking values in $\Omega^x = \{x_1, x_2\} \subset \mathbb{R}$.
- The prior probabilities $p = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}$
- $w \sim \mathcal{N}(0, \sigma^2)$ is a Gaussian random variable

Signal to noise ratio

The *signal-to-noise ratio* (SNR) is

$$\boxed{s = \frac{x_2 - x_1}{2\sigma}}$$

Often $x_1 = -\alpha$ and $x_2 = \alpha$, then $s = \dfrac{\alpha}{\sigma}$

Gaussian detection

We need to find the transition probabilities, i.e., the conditional pdf of y given X_j.

$$\begin{aligned}
\int_a^b A(z,j)\,dz &= \mathbf{Prob}(y \in [a,b] \mid x = x_j) \\
&= \mathbf{Prob}(x + w \in [a,b] \mid x = x_j) \\
&= \mathbf{Prob}(w \in [a - x_j, b - x_j] \mid x = x_j) \\
&= \int_{a-x_j}^{b-x_j} f_w(w)\,dw \\
&= \int_a^b f_w(z - x_j)\,dz
\end{aligned}$$

Since this holds for all intervals $[a,b] \subset \mathbb{R}$, we have

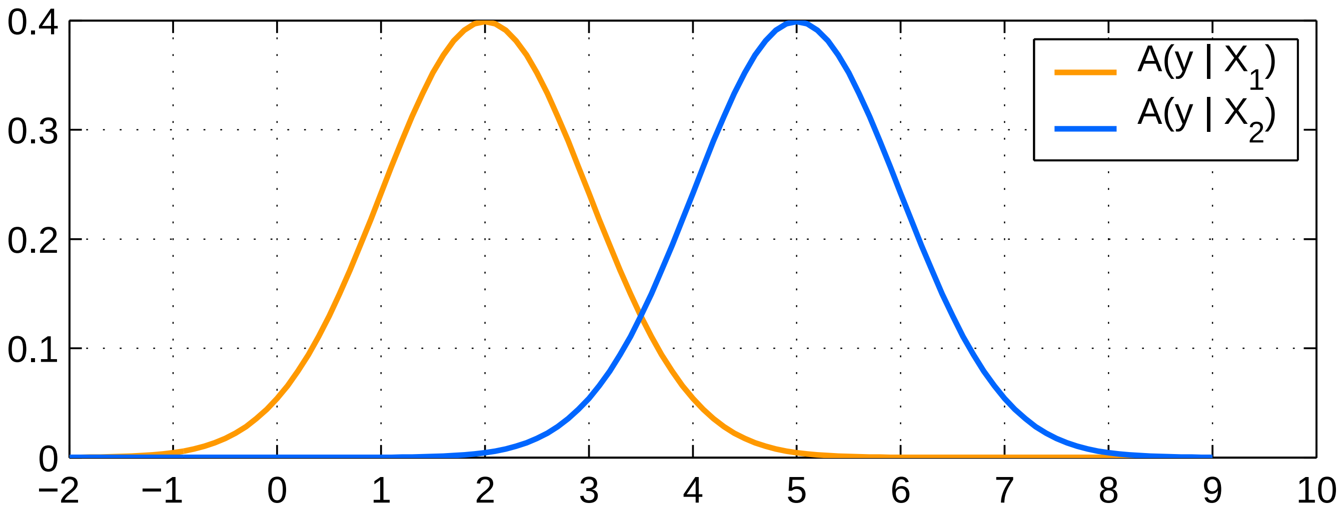
$$A(z,j) = f_w(z - x_j)$$

The MAP estimator for Gaussian detection

With prior $p = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}$, the MAP estimator is

$$\begin{aligned} f_{\text{map}}(y_{\text{meas}}) &= \arg \max_j \frac{1}{2\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y_{\text{meas}} - x_j)^2\right) \\ &= \arg \max_j \exp\left(-\frac{1}{2}(y_{\text{meas}} - x_j)^2\right) \\ &= \arg \min_j (y_{\text{meas}} - x_j)^2 \end{aligned}$$

i.e., estimate $x = x_j$, where x_j is the closest to y_{meas} in $\{x_1, x_2\}$



Error probabilities with equal priors

When $x_1 \le x_2$, the estimator is

$$f_{\text{map}}(y) = \begin{cases} 1 & \text{if } y < \dfrac{x_1 + x_2}{2} \\ 2 & \text{otherwise} \end{cases}$$

Hence the probability that $x = x_1$ and the estimate is correct is

$$\begin{aligned} E_{11} &= \int_{-\infty}^{(x_1+x_2)/2} J(y,1)\,dy \\ &= \int_{-\infty}^{(x_1+x_2)/2} \tfrac{1}{2} f_w(y - x_1)\,dy \\ &= \int_{-\infty}^{(x_2-x_1)/2} \tfrac{1}{2} f_w(z)\,dz \\ &= \frac{1}{2} F_{\mathcal{N}}\Big(\frac{x_2 - x_1}{2\sigma}\Big) \end{aligned}$$

where $F_{\mathcal{N}}(a) = \frac{1}{2} + \frac{1}{2}\,\mathbf{erf}\big(a/\sqrt{2}\big)$ is the cdf for a random variable with $\mathcal{N}(0,1)$ pdf

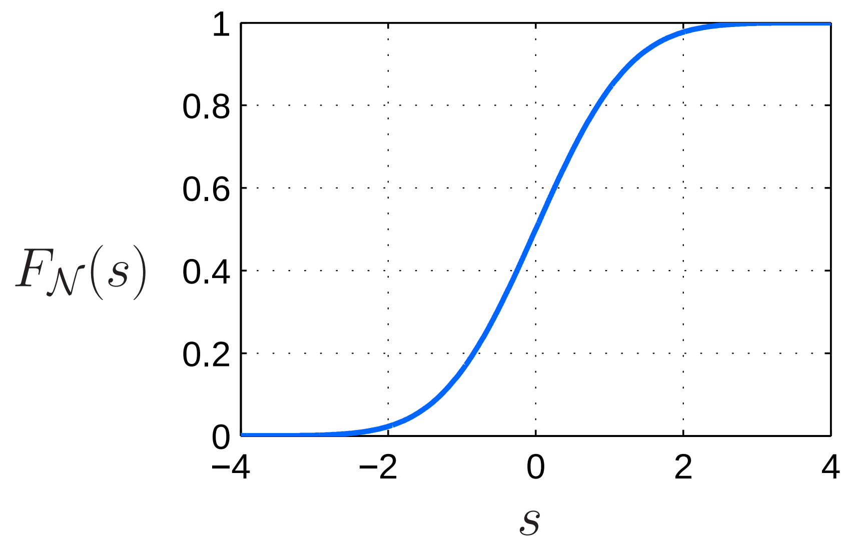
Error probabilities and SNR

With equal priors, similarly $E_{22} = \frac{1}{2} F_{\mathcal{N}}\left(\frac{x_2 - x_1}{2\sigma}\right)$ and so

$$\text{probability that estimate is correct} \; = F_{\mathcal{N}}(s)$$

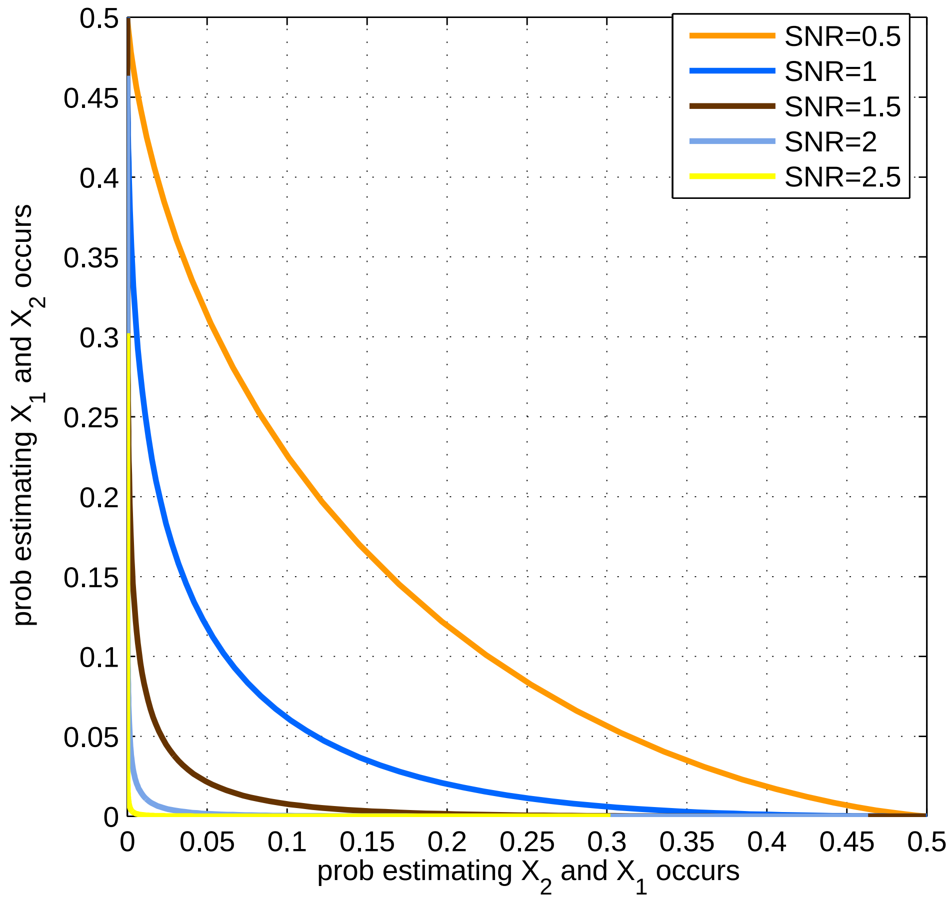
- The probability of correctness is a function solely of SNR s
- For this reason, SNR is also called the *discriminability*

$$F_{\mathcal{N}}(s) = \begin{cases} \frac{1}{2} & \text{if } s = 0 \\ 0.8413 & \text{if } s = 1 \\ 0.9772 & \text{if } s = 2 \\ 0.9987 & \text{if } s = 3 \end{cases}$$



The operating characteristic and SNR

- The operating characteristic is a function only of SNR.
- When the priors are equal, the operating characteristics are below.



Posterior probabilities

We'll drop the assumption that the priors are equal. We have

$$
\begin{aligned}
B(y_{\text{meas}}, 1) = \mathbf{Prob}(X_1 \mid y = y_{\text{meas}}) &= \frac{A(y_{\text{meas}}, 1)p_1}{\sum_{k=1}^2 A(y_{\text{meas}}, k)p_k} \\
&= \frac{1}{1 + \dfrac{A(y_{\text{meas}}, 2)p_2}{A(y_{\text{meas}}, 1)p_1}} \\
&= \frac{1}{1 + \dfrac{p_2 \exp(-(y_{\text{meas}} - x_2)^2/2\sigma^2)}{p_1 \exp(-(y_{\text{meas}} - x_1)^2/2\sigma^2)}} \\
&= \frac{1}{1 + e^{-f(y_{\text{meas}})}}
\end{aligned}
$$

Posterior probabilities

We have

$$\begin{aligned} f(y) &= \log\frac{p_2}{p_1} + \frac{1}{2\sigma^2}(y - x_2)^2 - \frac{1}{2\sigma^2}(y - x_1)^2 \\ &= \log\frac{p_2}{p_1} + \frac{1}{2\sigma^2}(x_2^2 - x_1^2) + \frac{1}{\sigma^2}(x_1 - x_2)y \\ &= ay + b \end{aligned}$$

i.e., the function f is *linear*.

here

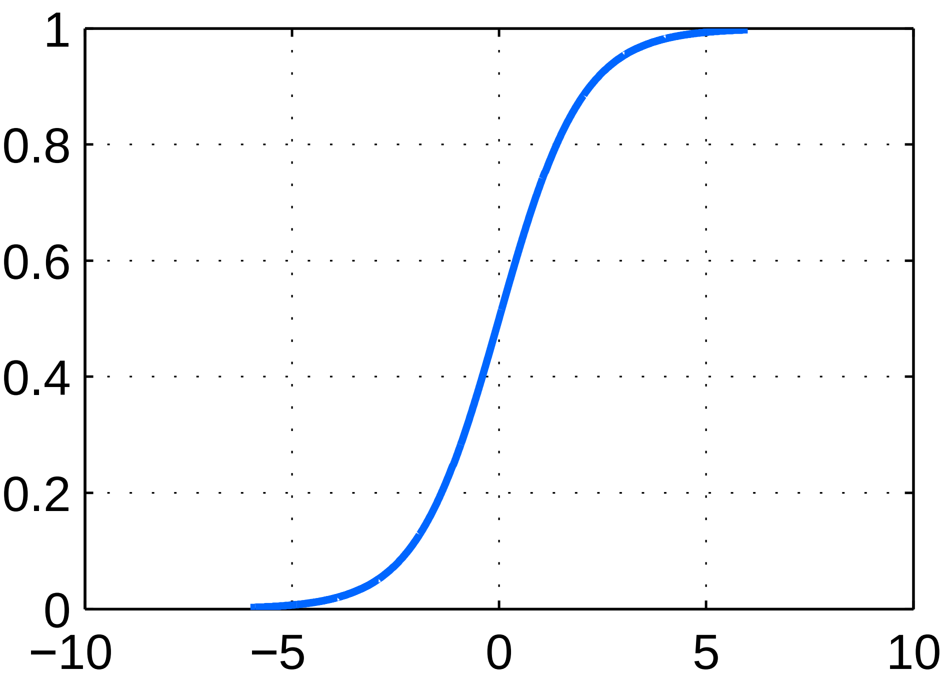
$$a = \frac{1}{\sigma^2}(x_1 - x_2) \qquad b = \log\frac{p_1}{p_2} + \frac{1}{2\sigma^2}(x_2^2 - x_1^2)$$

The sigmoid function

The function $l : \mathbb{R} \to \mathbb{R}$ given by

$$l(x) = \frac{1}{1 + e^{-x}}$$

is called the *logistic* or *sigmoid* function.

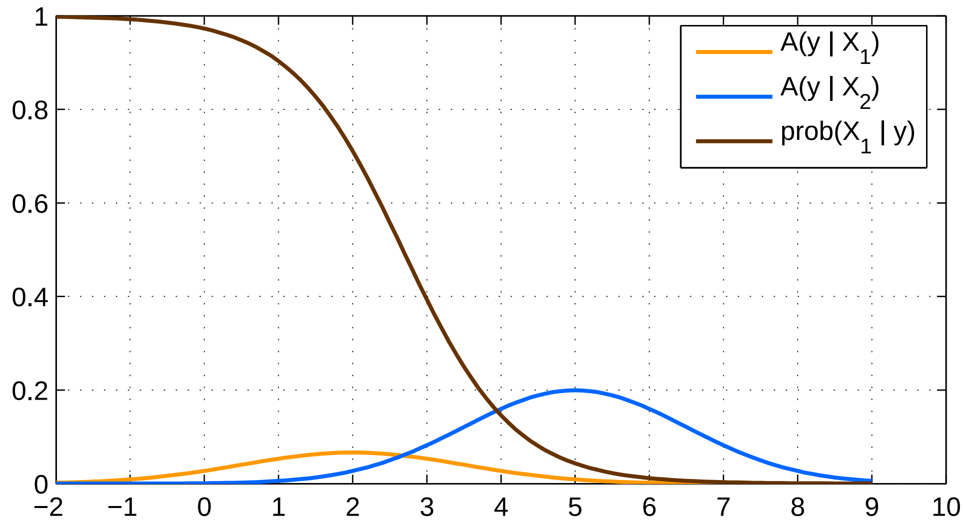


Posterior probabilities

So the posterior probability of X_1 is

$$\begin{aligned}\mathbf{Prob}(X_1 \,|\, y = y_{\text{meas}}) &= l(ay+b) \\ &= \frac{1}{1+e^{-(ay+b)}}\end{aligned}$$

i.e., it is a *stretched and shifted* sigmoid function.



Aside: instead of constructing the a-posterior probabilities based on the distributions, one can instead *fit* the sigmoid to *data*. This is what a *neural network* does.

More general Gaussian classification

Let's look at the general case where we measure y, and would like to determine which of $X_1, X_2, \ldots, X_n$ occurred, where

$$y \mid X_j \sim \mathcal{N}(\mu_j, \Sigma_j)$$

and the prior probabilities are $\mathbf{Prob}(x = x_j) = p_j$

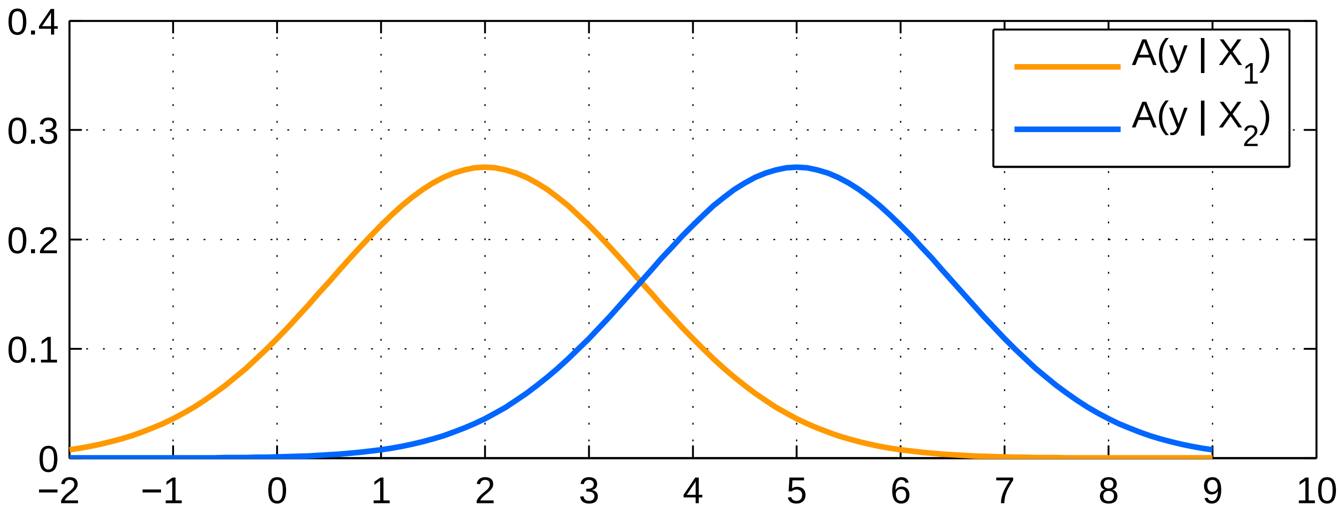
Decision regions

Recall the decision regions are defined to be

$$R_k = \left\{ y \in \mathbb{R}^m \mid f_{\text{est}}(y) = k \right\}$$

that is, R_k is the set of y for which we decided X_k occurred.

for example, with the pdfs below and equal priors we have $R_1 = \left\{ y \in \mathbb{R}^m \mid y \leq 3.5 \right\}$



More general Gaussian classification

The MAP estimator is

$$\begin{aligned} f_{\mathsf{map}}(y_{\mathsf{meas}}) &= \arg\max_k \frac{p_k}{(2\pi)^{\frac{m}{2}}(\det \Sigma_k)^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(y_{\mathsf{meas}} - \mu_k)^T \Sigma_k^{-1}(y_{\mathsf{meas}} - \mu_k)\right) \\ &= \arg\max_k g_k(y_{\mathsf{meas}}) \end{aligned}$$

where we define for convenience

$$g_k(y) = \log p_k - \frac{1}{2}\log\det\Sigma_k - \frac{1}{2}(y - \mu_k)^T\Sigma_k^{-1}(y - \mu_k)$$

Notice that g_k is a *quadratic function* of y

Gaussian decision boundaries are quadratic

Hence the decision regions are

$$R_k = \Big\{ y \in \mathbb{R}^m \mid g_k(y) \geq g_j(y) \text{ for all } y \in \mathbb{R}^m \Big\}$$

when there are ties, we can choose which R_k to put y in.

Hence the boundary between R_1 and R_2 is part of the *separating surface*

$$\Big\{ y \in \mathbb{R}^m \mid g_1(y) = g_2(y) \Big\}$$

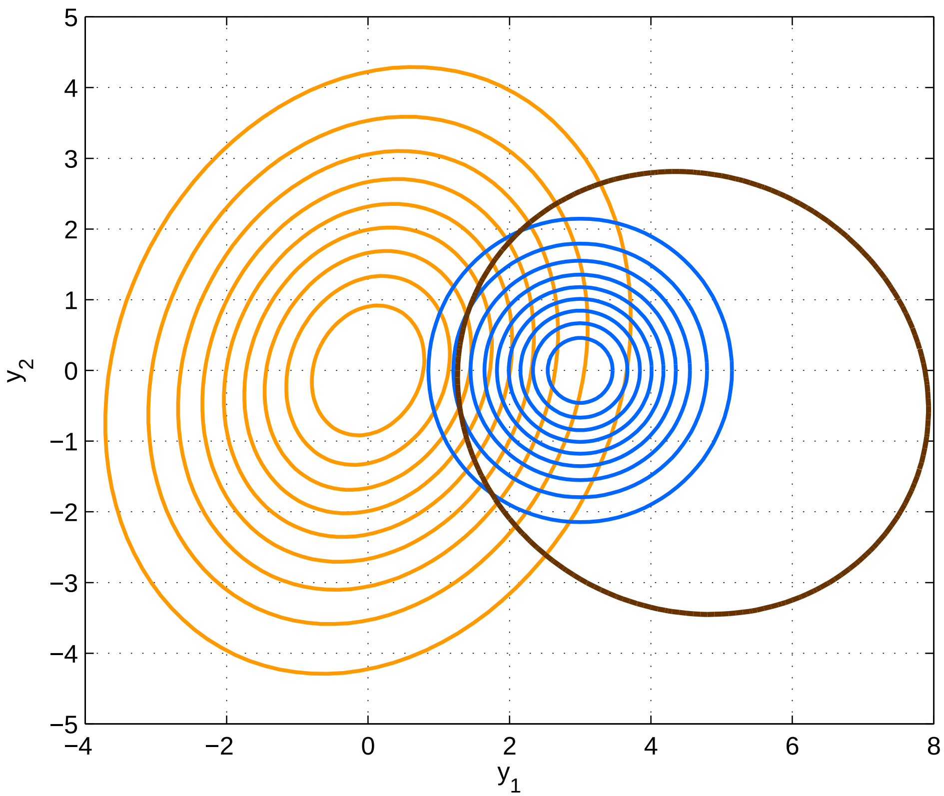
So the separating surfaces are piecewise *level sets* of quadratic functions.

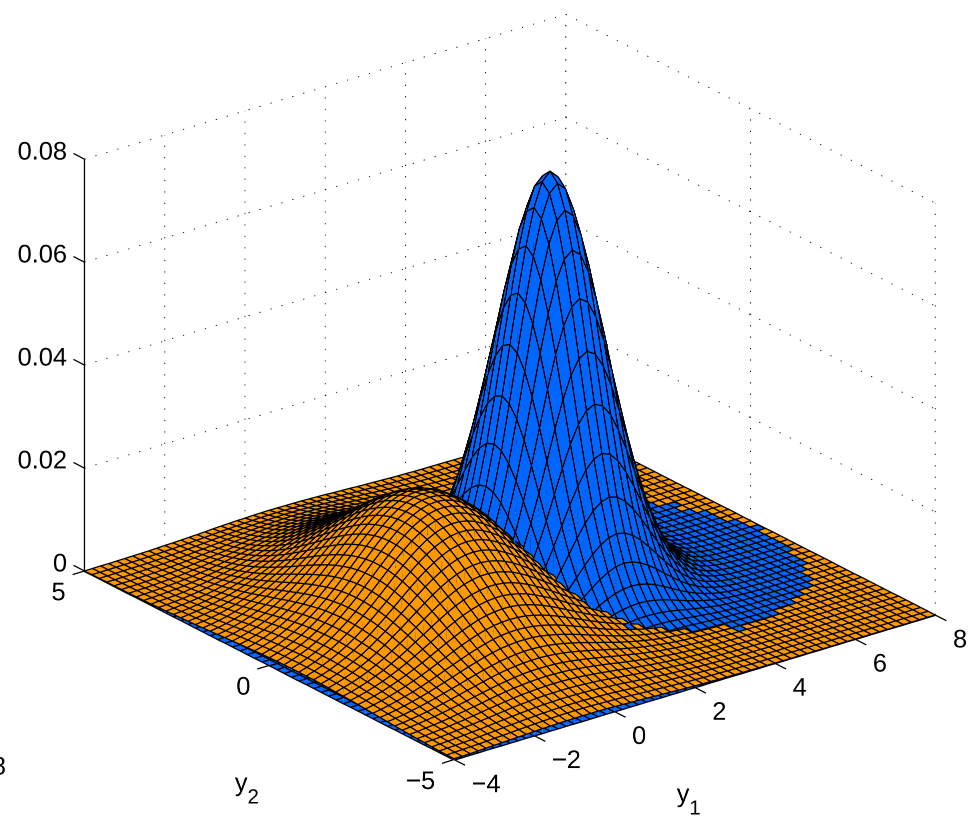
Example: Gaussian decision regions

$$p_1 = 0.5 \quad \mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Sigma_1 = \begin{bmatrix} 3 & 0.6 \\ 0.6 & 4 \end{bmatrix} \qquad p_2 = 0.5 \quad \mu_2 = \begin{bmatrix} 3 \\ 0 \end{bmatrix} \quad \Sigma_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

The decision boundary is the quadratic curve

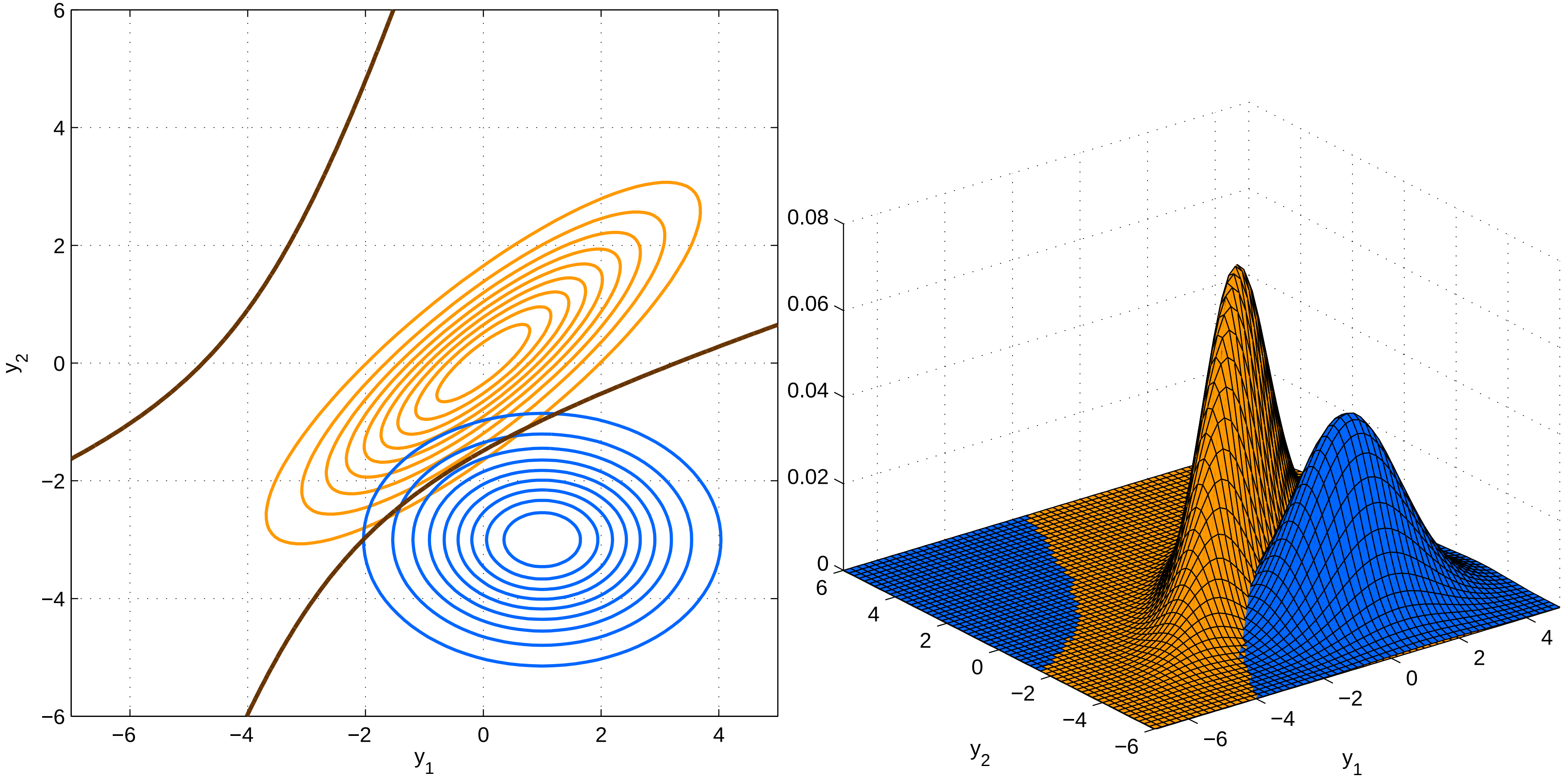
$$\log p_1 - \frac{1}{2}\log\det\Sigma_1 - \frac{1}{2}(y-\mu_1)^T\Sigma_1^{-1}(y-\mu_1) = \log p_2 - \frac{1}{2}\log\det\Sigma_2 - \frac{1}{2}(y-\mu_2)^T\Sigma_2^{-1}(y-\mu_2)$$



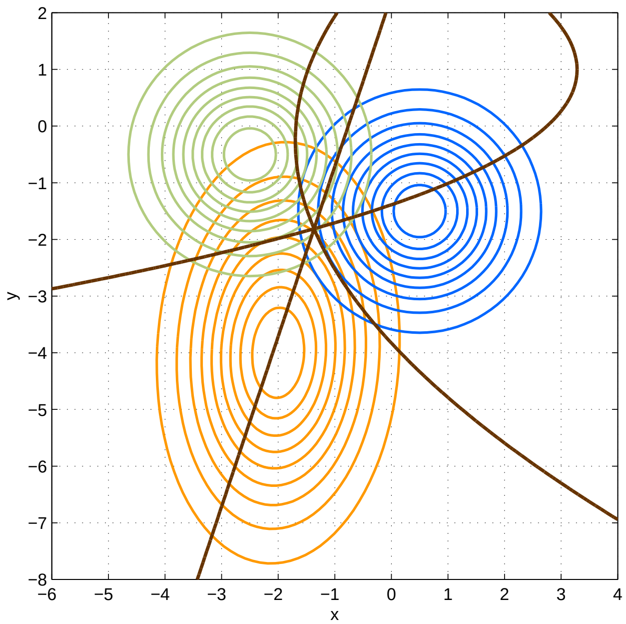


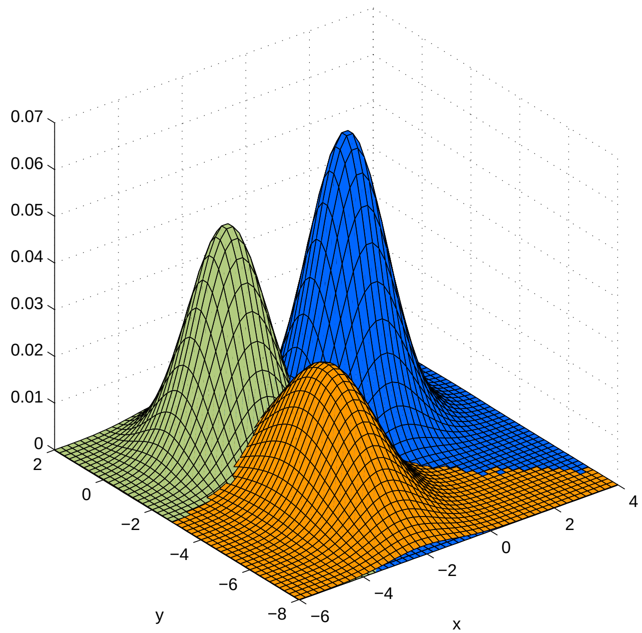
Decision regions

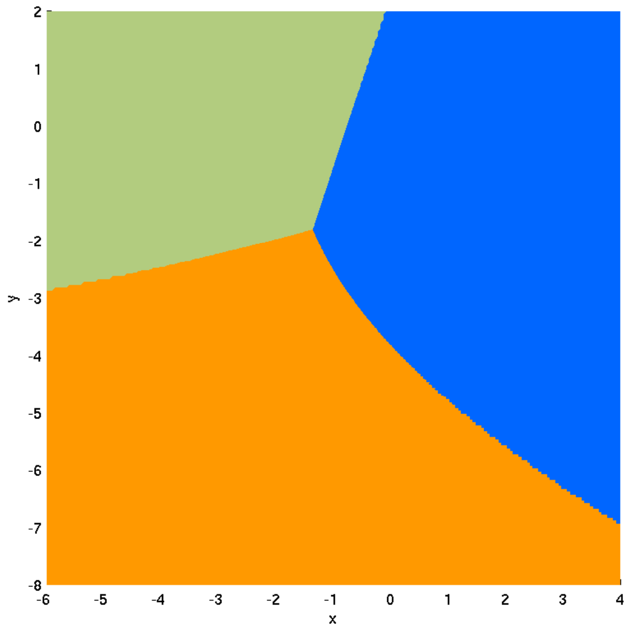
Since the decision boundary is quadratic, it may be hyperbolic; then the decision regions may be disconnected.



Three Gaussians







Equal covariances

An important special case is when the covariances Σ_k are all *equal*; so $\Sigma_k = \Sigma$ for all k.

Then the MAP estimator is

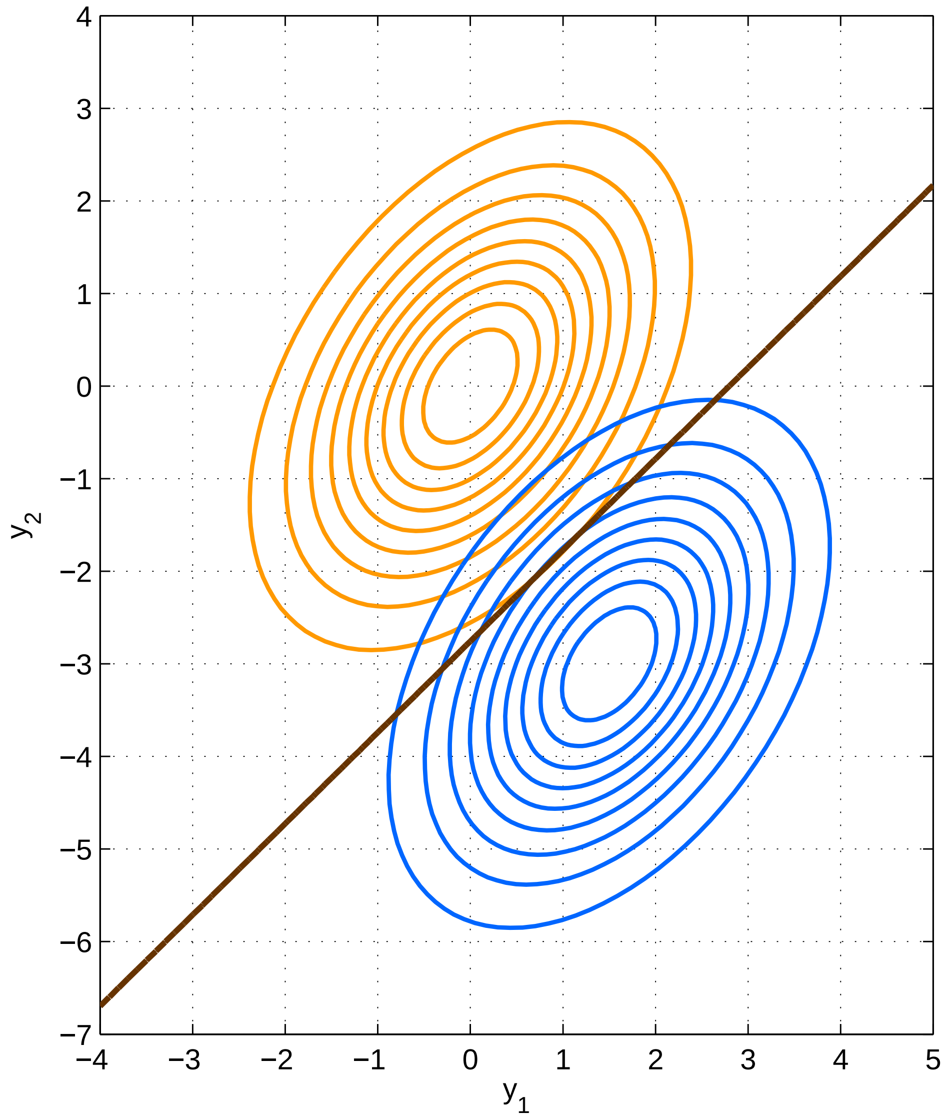
$$\begin{aligned} f_{\mathsf{map}}(y_{\mathsf{meas}}) &= \arg\max_k \left(\log p_k - \frac{1}{2}\log\det\Sigma - \frac{1}{2}(y-\mu_k)^T\Sigma^{-1}(y-\mu_k) \right) \\ &= \arg\max_k \; g_k^{\mathsf{eq}}(y_{\mathsf{meas}}) \end{aligned}$$

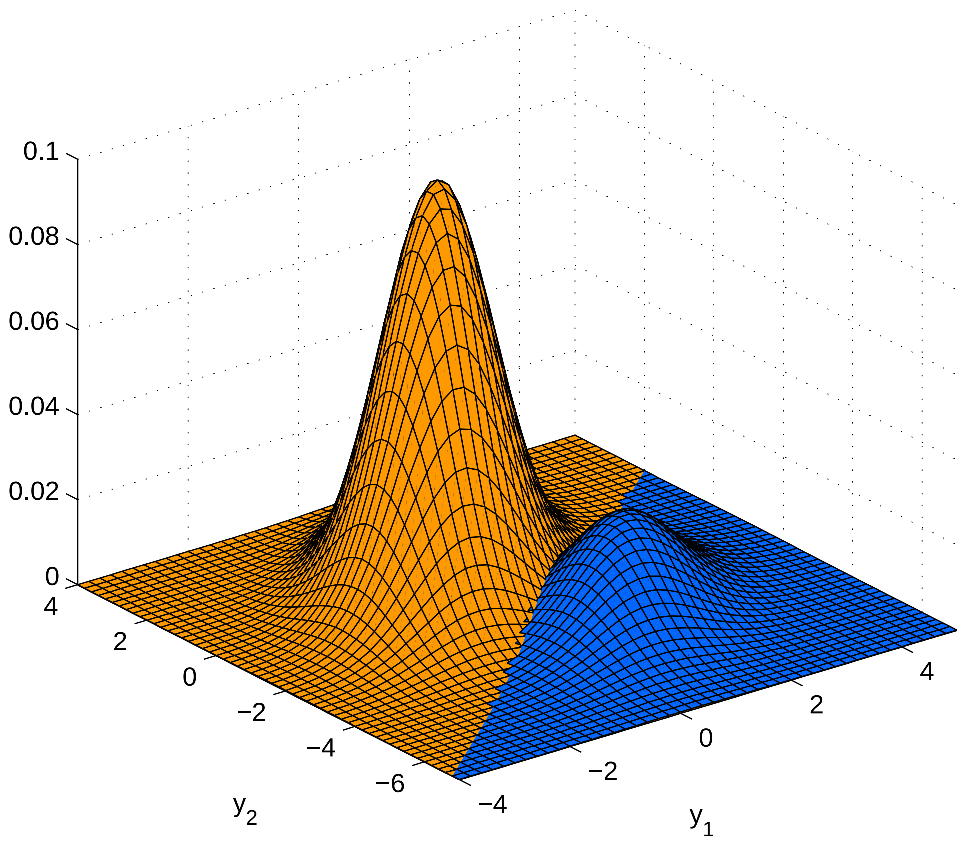
where we define for convenience

$$g_k^{\mathsf{eq}}(y) = \mu_k^T\Sigma^{-1}y + \log p_k - \frac{1}{2}\mu_k^T\Sigma^{-1}\mu_k$$

g^{eq} is a *linear* function of y; so the separating surfaces are *hyperplanes*

Decision regions





The logistic function

When there are *two hypotheses*, and the measurements are Gaussian with *equal covariances* then we can show as before that

$$\mathbf{Prob}(X_1 \mid y = y_{\mathsf{meas}}) = l(a^T y_{\mathsf{meas}} + b)$$

where

$$a = \Sigma^{-1}(\mu_1 - \mu_2) \qquad b = \log \frac{p_1}{p_2} + \frac{1}{2}\mu_2^T \Sigma^{-1} \mu_2 - \frac{1}{2}\mu_1^T \Sigma^{-1} \mu_1$$

- Again, the a-posteriori probabilities are the sigmoid function composed with a linear function
- The decision boundary is the set of y such that $l(a^T y + b) = \frac{1}{2}$.