Authentication

Prof. Dr. Helene Dörksen

helene.doerksen@th-owl.de

Lecture 5:

Supervised Learning

Supervised Learning

Parametric methods of (statistical) classification:
- require probability distributions
- estimate parameters (e.g. mean and standard deviation) and provide compact representation of the classes

Examples:
- *Bayes' decision rule* (based on probability distributions)
- *discriminant analysis* based on functions which separate classes

Non-parametric methods:
- probability distributions unknown

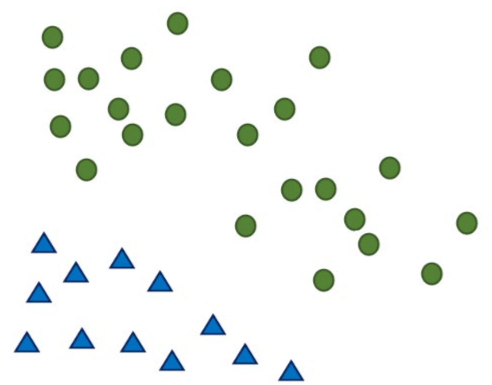
Examples:
- estimate density functions (*Parzen window* approach)
- directly construct decision boundaries (*k-NN, SVM, Neural Network*)

Bayesian Decision Theory

Single feature

class 1 ($\boldsymbol{\omega_1}$)

~ 70%



class 2 ($\boldsymbol{\omega_2}$)

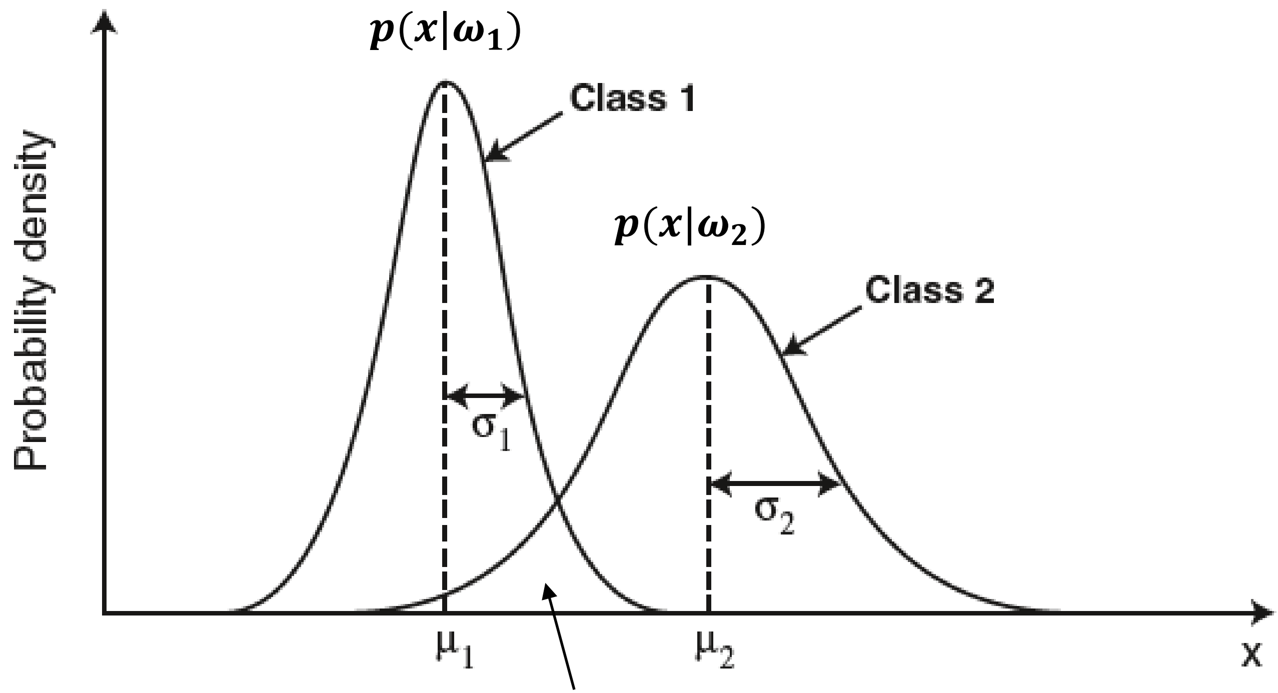
~ 30%

(prior) probabilities: P ($\boldsymbol{\omega_1}$)=0.7   P($\boldsymbol{\omega_2}$)=0.3 (0.7+03=1.0)

$$P(\text{error}) = P(\text{choose } \omega_2|\omega_1) \cdot P(\omega_1) + P(\text{choose } \omega_1|\omega_2) \cdot P(\omega_2)$$

Bayesian Decision Theory

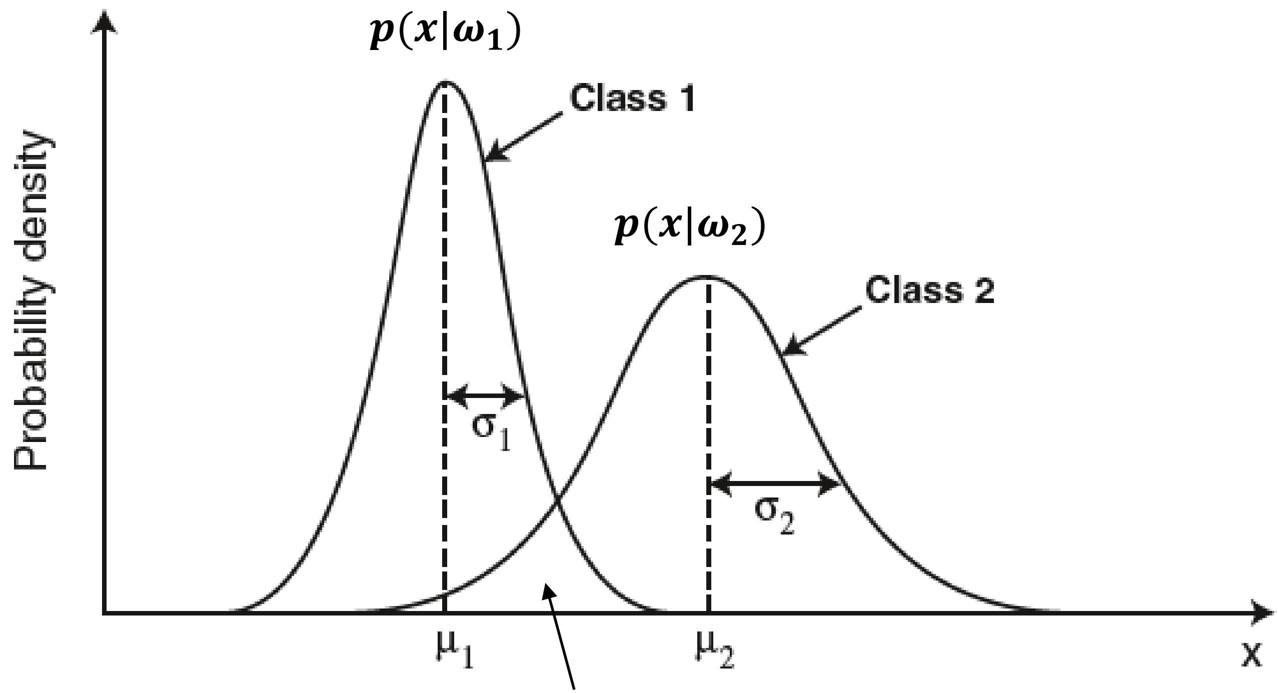
Single feature



? two probability density functions overlap

Bayesian Decision Theory

Single feature



! two probability density functions overlap

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Bayesian Decision Theory

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

$$\text{If} \quad \frac{p(x|\omega_1) \cdot P(\omega_1)}{p(x)} > \frac{p(x|\omega_2) \cdot P(\omega_2)}{p(x)} \quad \text{choose } \omega_1$$

$$\text{else} \quad \text{choose } \omega_2$$

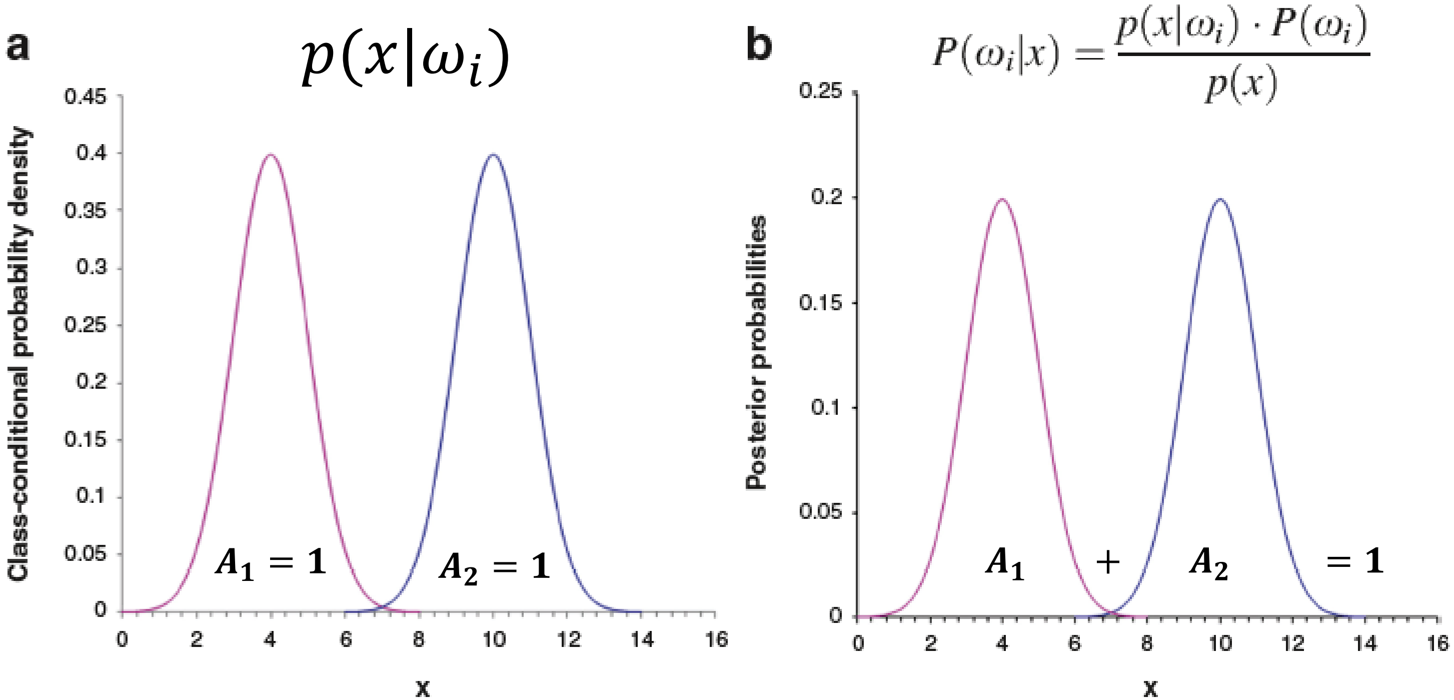
might be ignored in classification task

likelihood ratio test

$$\left\{ \begin{array}{ll} \underbrace{\frac{p(x|\omega_1)}{p(x|\omega_2)}}_{\text{likelihood ratio}} > \frac{P(\omega_2)}{P(\omega_1)} & \text{choose } \omega_1 \\ \text{else} & \text{choose } \omega_2 \end{array} \right.$$

Bayesian Decision Theory

Example 1: Suppose that a single measurement has Gaussian class-conditional densities of equal variance ($\boldsymbol{\sigma^2 = 1}$) but different means ($\boldsymbol{\mu_1 = 4}$ and $\boldsymbol{\mu_2 = 10}$) and that the prior probabilities are equally likely [$\boldsymbol{P(\omega_1) = P(\omega_2) = 0.5}$].

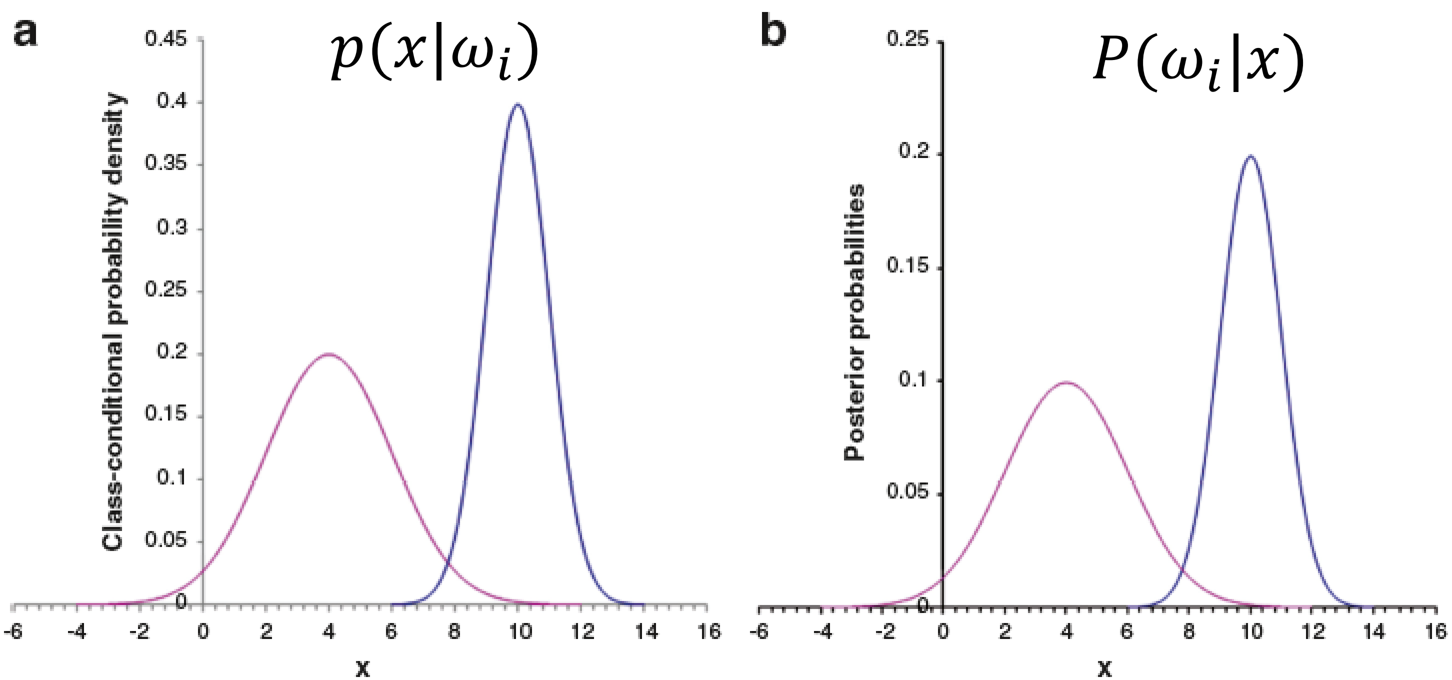


likelihood ratio test

$$\begin{cases} \dfrac{\exp(-\frac{1}{2}(x-4)^2)}{\exp(-\frac{1}{2}(x-10)^2)} > \dfrac{0.5}{0.5} > 1 & \text{choose } \omega_1 \\ \text{else} & \text{choose } \omega_2 \end{cases}$$

Bayesian Decision Theory

Example 2: Consider Example 1 with different variances $\boldsymbol{\sigma_1^2 = 4}$ and $\boldsymbol{\sigma_2^2 = 1}$



likelihood ratio test

$$\begin{cases} 8\ln\frac{1}{2} - (x-4)^2 > -4(x-10)^2 \quad \text{choose } \omega_1 \\ 3x^2 - 72x + (384 + 8\ln\frac{1}{2}) > 0 \end{cases}$$

Bayesian Decision Theory

Example 2: Consider Example 1 with different variances $\boldsymbol{\sigma_1^2 = 4}$ and $\boldsymbol{\sigma_2^2 = 1}$

likelihood ratio test

$$\begin{cases} 8\ln\frac{1}{2} - (x-4)^2 > -4(x-10)^2 \quad \text{choose } \omega_1 \\ 3x^2 - 72x + (384 + 8\ln\frac{1}{2}) > 0 \end{cases}$$

can be solved by:

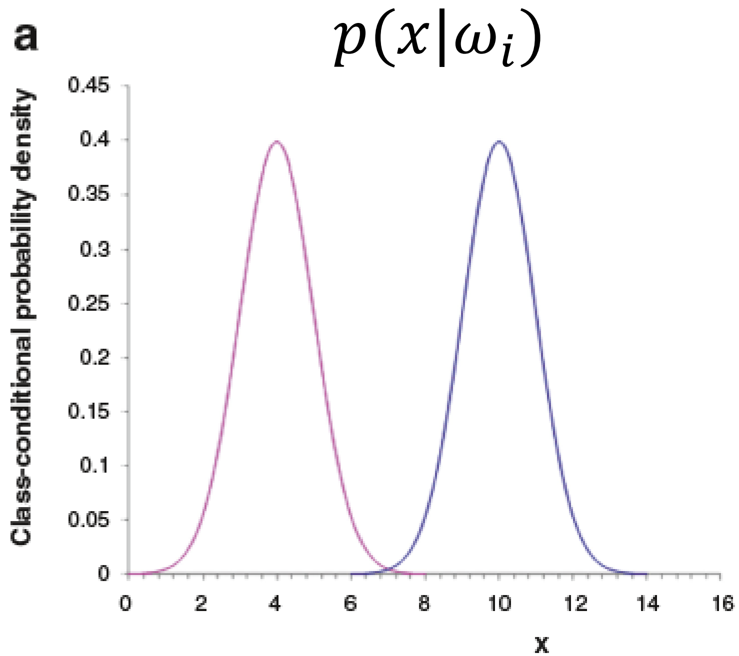
$$\begin{aligned} x &= (72 \pm 25.3)/6 \\ &= 7.78 \text{ and } 16.22 \end{aligned}$$

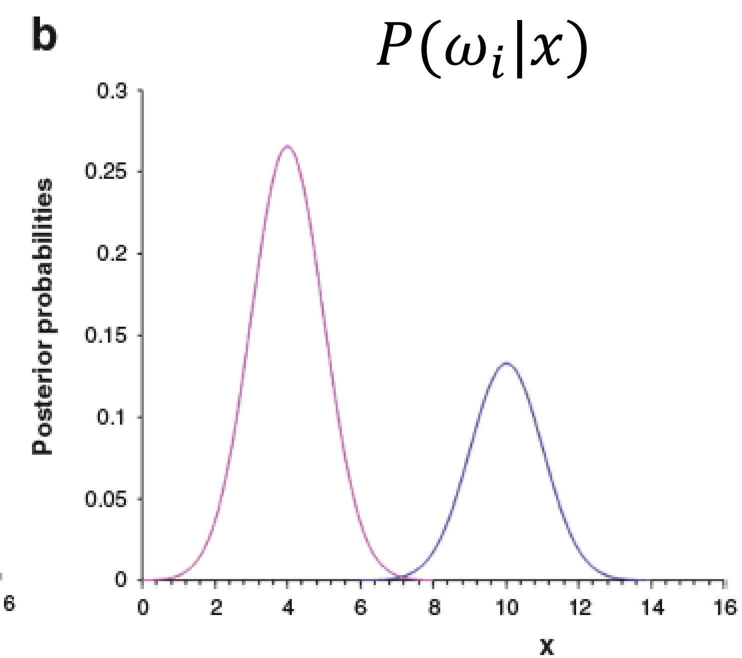
likelihood ratio test

$$\begin{cases} x < 7.78 \text{ or } x > 16.22 \quad \text{choose } \omega_1 \\ 7.78 < x < 16.22 \quad \text{choose } \omega_2 \end{cases}$$

Bayesian Decision Theory

Example 3: Suppose $P(\omega_1) = 2/3$ and $P(\omega_2) = 1/3$ $(\sigma_1^2 = \sigma_2^2 = 1)$

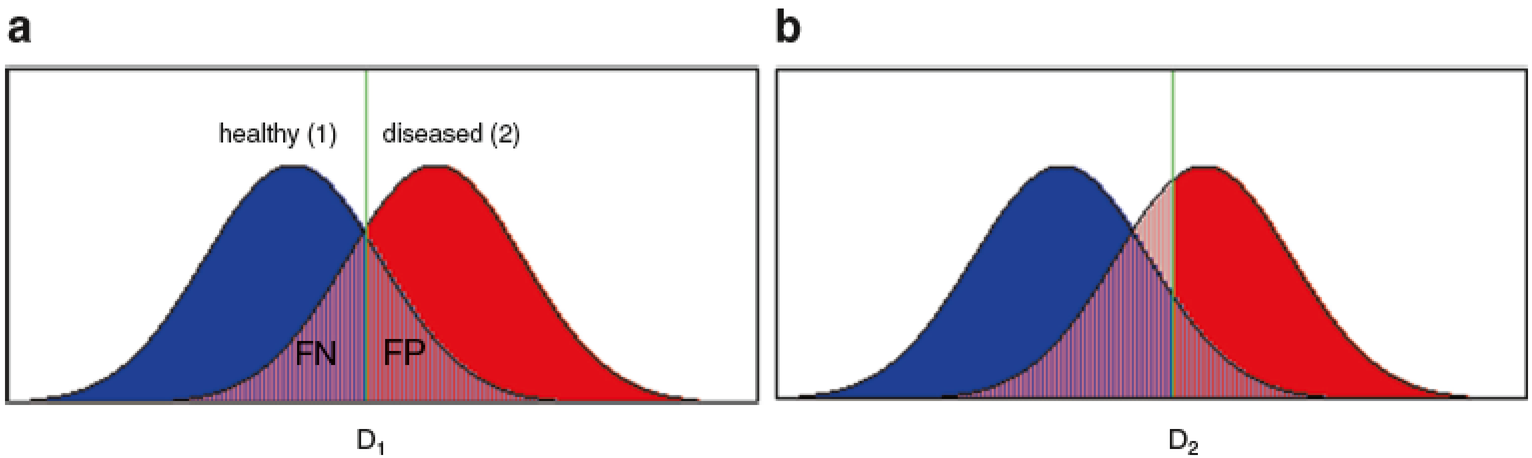




likelihood ratio test

$$\begin{cases} \dfrac{\exp(-\frac{1}{2}(x-4)^2)\cdot 2/3}{\exp(-\frac{1}{2}(x-10)^2)\cdot 1/3} > 1 & \text{choose } \omega_1 \\ \text{else} & \text{choose } \omega_2 \end{cases}$$

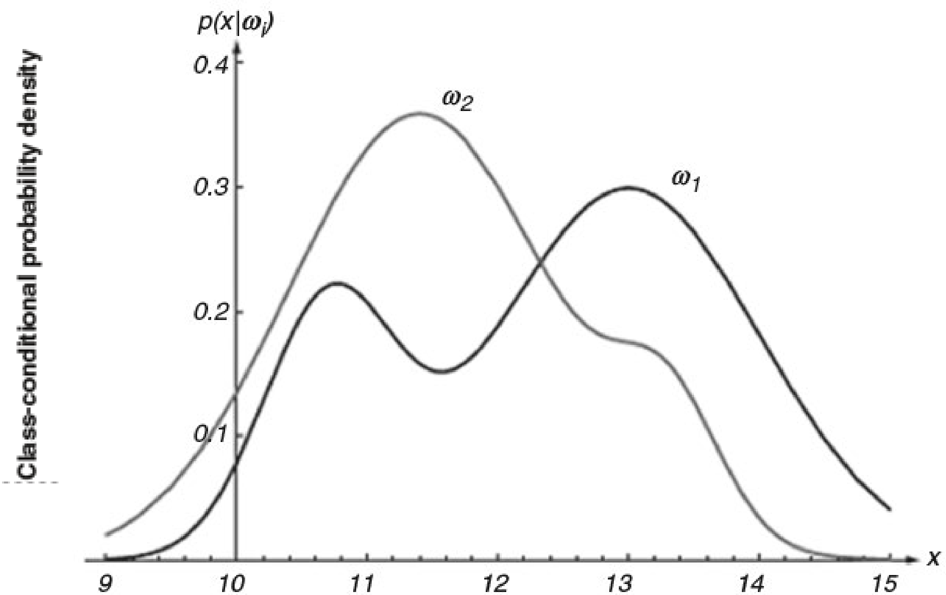
Bayesian Decision Theory

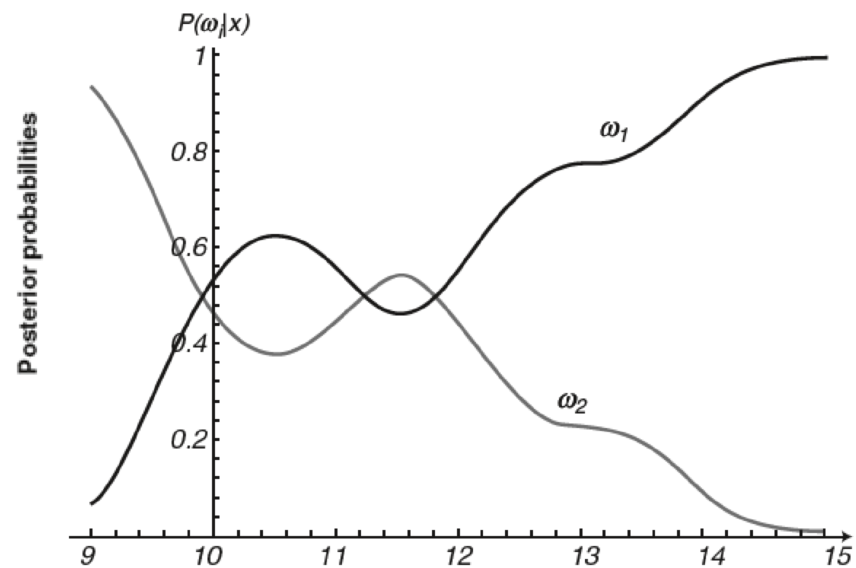


Intersecting distributions with a decision threshold
(a) at the intersection point and
(b) higher than the intersection point

Bayesian Decision Theory

More complicated example: Suppose $\boldsymbol{P(\omega_1) = 2/3}$ and $\boldsymbol{P(\omega_2) = 1/3}$





Bayesian Decision Theory

In previous examples we have assumed that the two types of classification errors are of equal importance.

However, this may not always be true!

Solution: *loss matrix* $$\lambda = \begin{vmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{vmatrix}$$

$\boldsymbol{\lambda_{ij}}$ is the loss associated with deciding $\boldsymbol{\omega_i}$ when the correct state is $\boldsymbol{\omega_j}$

diagonal terms are usually set to zero

generalised likelihood ratio test

$$\begin{cases} \dfrac{p(x|\omega_1)}{p(x|\omega_2)} > \dfrac{(\lambda_{12} - \lambda_{22}) \cdot P(\omega_2)}{(\lambda_{21} - \lambda_{11}) \cdot P(\omega_1)} & \text{choose } \omega_1 \\ \text{else} & \text{choose } \omega_2 \end{cases}$$

Bayesian Decision Theory

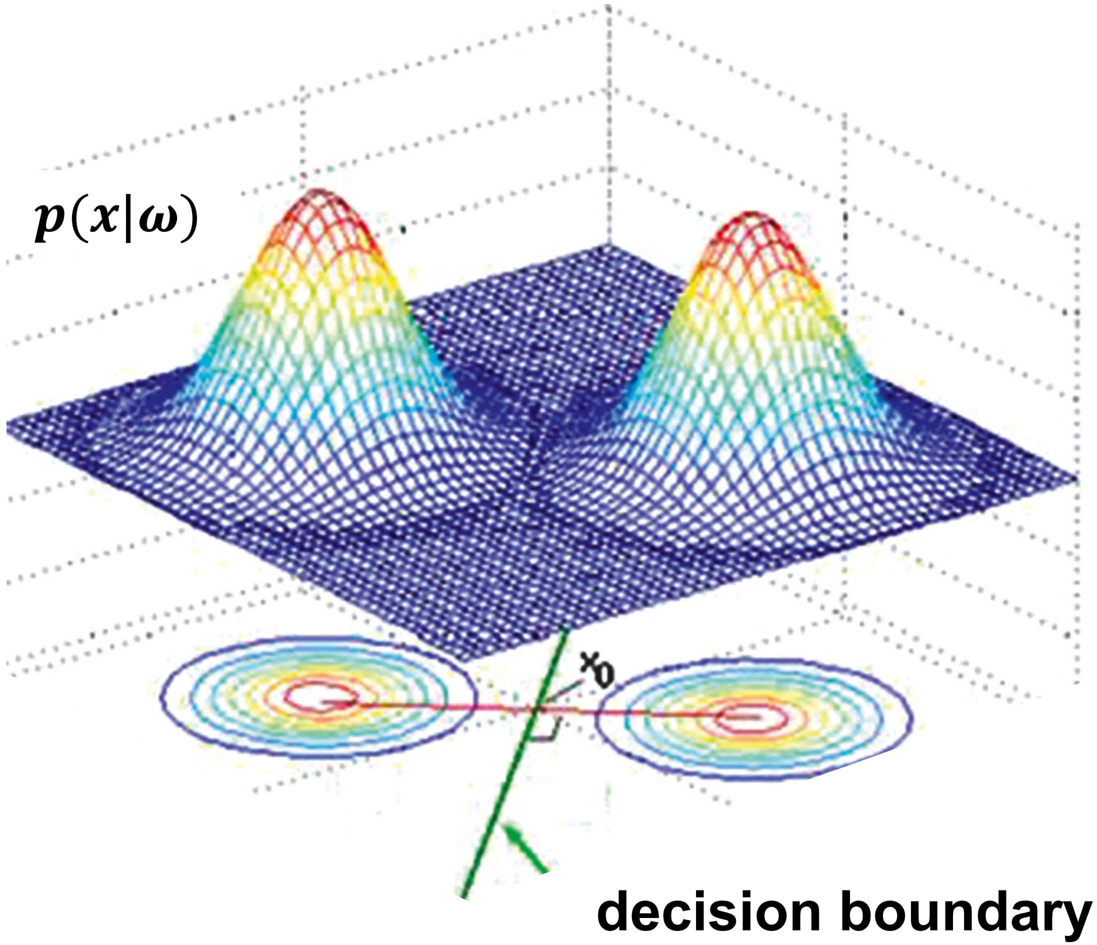
If the class-conditional probabilities are **Gaussian distributions**,

$\boldsymbol{N(\mu_1, \sigma_1^2)}$ and $\boldsymbol{N(\mu_2, \sigma_2^2)}$

substitutions, taking (natural) logs and simplifying gives:

$$((x-\mu_2)/\sigma_2)^2 - ((x-\mu_1)/\sigma_1)^2 > 2\ln\frac{(\sigma_1 \cdot k)}{(\sigma_2)} \quad \text{choose } \omega_1 \text{ : else choose } \omega_2$$

Discriminant Functions and Decision Boundaries



Discriminant Functions and Decision Boundaries

Discriminant functions might be any appropriate function for considered application, i.e.

$g_i(\boldsymbol{x})$ for each class $i = 1,2,\ldots,$ where $\boldsymbol{x}$ is the feature vector

Example: discriminant function obtained from Bayes‘ rule

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

For Bayes’ classification, it is convenient to use the natural log:

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

Summary

We learned **classification rules (for one feature)** based on:

- **likelihood ratio test** (errors are of equal importance)
- loss matrix and **generalised likelihood ratio test** (errors are of **non equal** importance)

Classification rules for more general considerations (e.g. more than one features):

- **Decision boundaries** obtained from **discriminant functions**

Homework: Exercises and Labs

for the next week prepare practical exercises and labs from **Exercises Lec 5** (you will find it in the donwload area)