Elements of Bayesian Decision Theory

Prior probability (prior)

Classes

ω_1 - sea bass

ω_2 - salmon

a two-class problem

Prior probabilities:

$P(\omega_1)$ - probability of finding sea bass

$P(\omega_2)$ - probability of finding salmon

A simple decision rule (when no other information is available):

$$\begin{cases} \omega_1, if\ P(\omega_1) > P(\omega_2) \\ \omega_2, otherwise \end{cases}$$

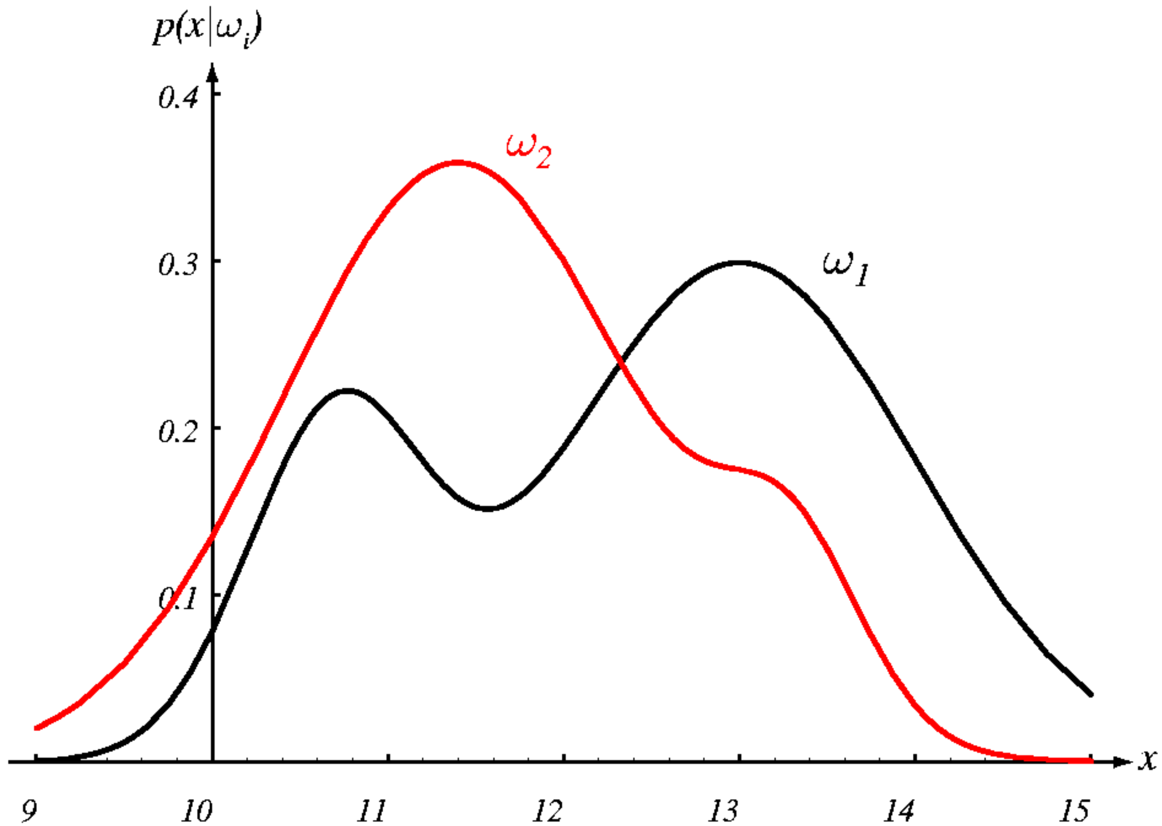
Where do you get the prior probabilities?

Answer: mostly from previous empirical observations or some theoretical considerations (e.g. 0.5 chance to throw tail/head of a fair coin).

Class conditional probability density function and likelihood

$$p(x \mid \omega_1)$$
$$p(x \mid \omega_2)$$

Likelihood –
pdf as a function of the second argument (class) with the first argument (feature value x) fixed



feature x (e.g. lightness)

Bayes formula/rule

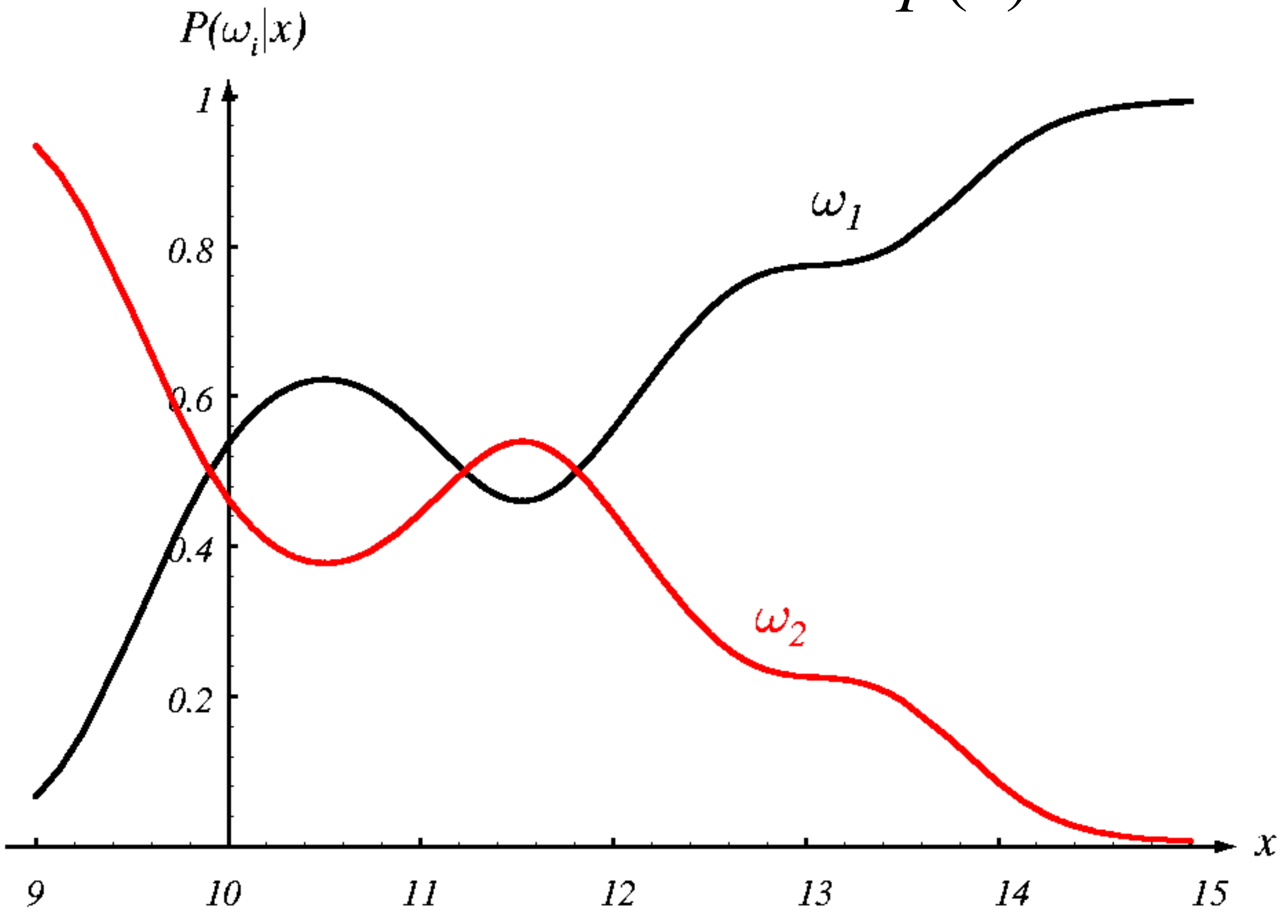
$$P(\omega_j \mid x) = \frac{p(x \mid \omega_j)\ P(\omega_j)}{p(x)}$$

$$p(x) = p(x \mid \omega_1)\ P(\omega_1) + p(x \mid \omega_2)\ P(\omega_2)$$

$$posterior = \frac{likelihood\ \text{ x }\ prior}{evidence}$$

Posterior probability

$$P(\omega_j \mid x) = \frac{p(x \mid \omega_j)\ P(\omega_j)}{p(x)}$$



$$P(\omega_1) = 2/3$$

$$P(\omega_2) = 1/3$$

For any x, use priors as coefficients of likelihoods and normalize so that their sum is 1.

(from Duda, Hart, Stork (2001) Pattern classification)

Bayes decision rule

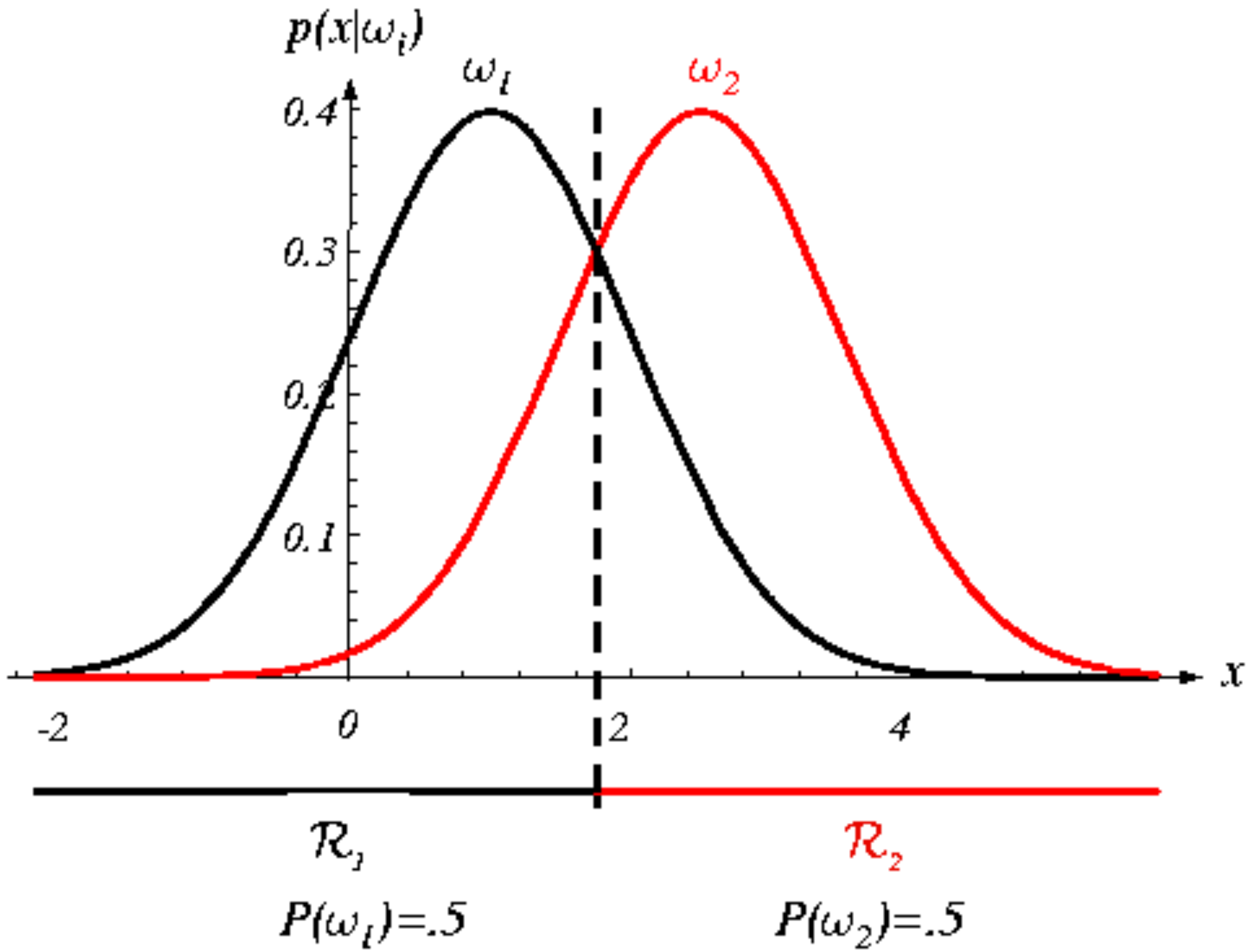
Probability of making an error:

$$P(error \mid x) = \begin{cases} P(\omega_1 \mid x), if \;\; we \; decide \;\; \omega_2 \\ P(\omega_2 \mid x), if \;\; we \; decide \;\; \omega_1 \end{cases}$$

Bayes decision rule:

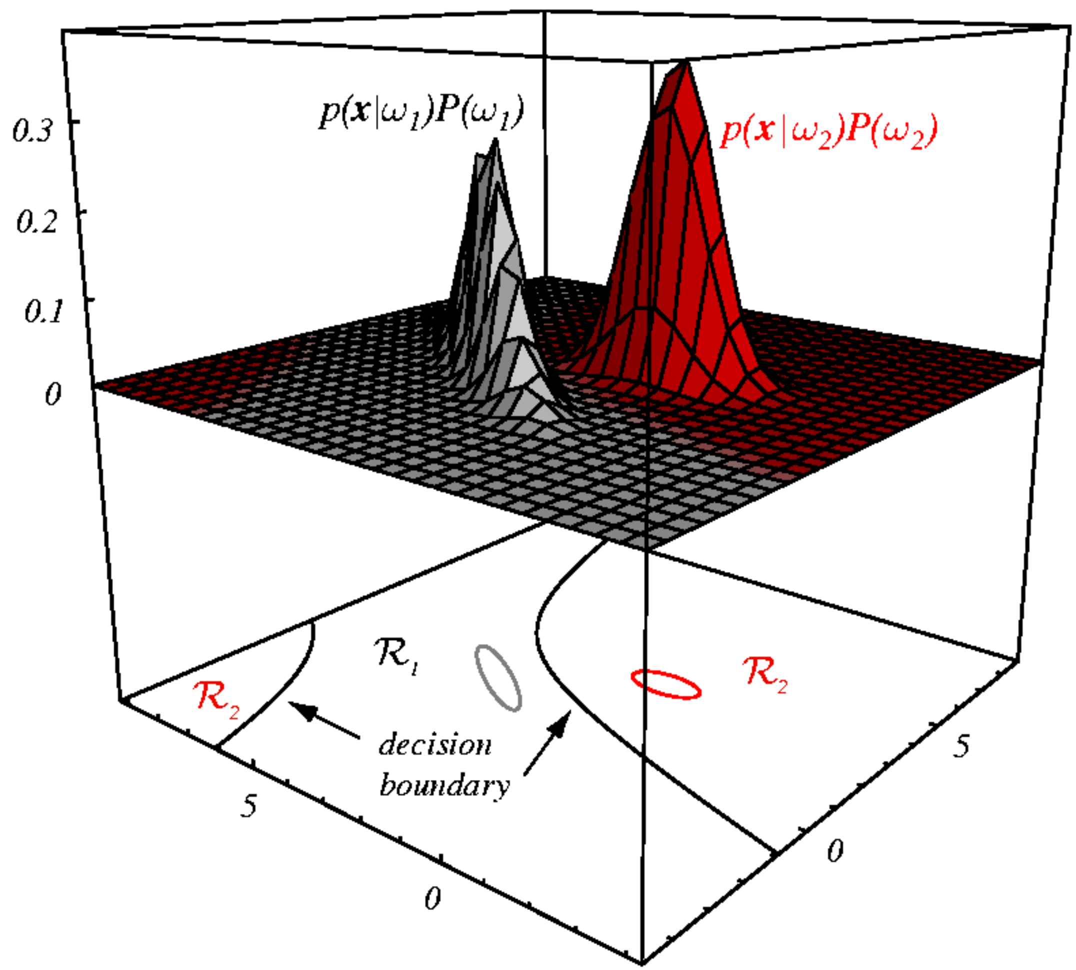
$$\begin{cases} \omega_1, if \;\; P(\omega_1 \mid x) > P(\omega_2 \mid x) \\ \omega_2, otherwise \end{cases}$$

Example of a decision criterion in a one-dimensional feature space



from Duda, Hart, Stork (2001) Pattern classification

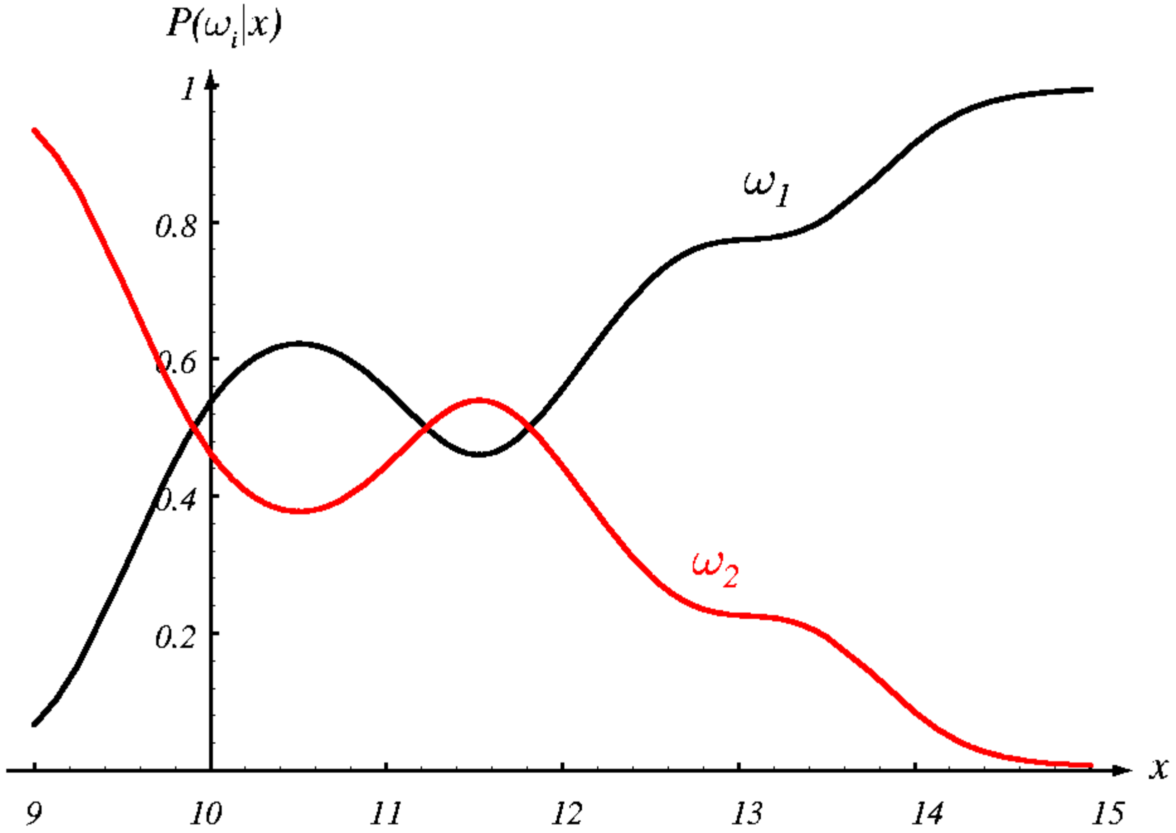
Example of a decision boundary in a two-dimensional feature



from Duda, Hart, Stork (2001) Pattern classification

Error probability of Bayes decision rule

$$P(error \mid x) = \min[P(\omega_1 \mid x), P(\omega_2 \mid x)]$$



(from Duda, Hart, Stork (2001) Pattern classification)

Generalizations of Bayesian Decision Theory

1. We replace the scalar *x* with the *feature vector* $\mathrm{x} \in \mathrm{R}^d$.
2. We introduce a *cost* or a *loss function* λ which states how costly each classification decisions is.

Let

$\{\omega_1, \omega_2, \ldots, \omega_c\}$ - categories (classes)

$\{\alpha_1, \alpha_2, \ldots, \alpha_c\}$ - possible actions

The loss function $\lambda(\alpha_i \mid \omega_j)$ describes the loss incurred for taking action α_i when the category is ω_j .

Bayes formula

$$P(\omega_j \mid \mathrm{x}) = \frac{p(\mathrm{x} \mid \omega_j)\; P(\omega_j)}{p(\mathrm{x})}$$

Evidence

$$p(\mathrm{x}) = \sum_{j=1}^{c} p(\mathrm{x} \mid \omega_j)\; P(\omega_j)$$

Generalized Bayesian decision theory

Taking action α_i , the loss/cost, also called *conditional risk,* is:

$$R(\alpha_i \mid \mathrm{x}) = \sum_{\mathrm{j=1}}^{\mathrm{c}} \lambda(\alpha_{\mathrm{i}} \mid \omega_j)\, P(\omega_j \mid \mathrm{x})$$

Rule to minimize the expected loss/cost:

Select that action which minimizes the conditional risk.

Generalized Bayesian decision theory

Let P(melanoma | x) = 0.1 and P(benign nevus | x) = 0.9

Bayesian classification: benign nevus (due to higher probability)

Let now consider the actions: 1 – remove, 2 – do not remove

With costs in Euro $a_{1,mel} = 50$ $a_{1,nev} = 50$,

$a_{2,mel} = 100000$ $a_{2,nev} = 0$

Expected cost (weighted average over many cases with x):

$C_1 = a_{1,mel}$ P(melanoma | x) + $a_{1,naev}$ P(benign nevus | x) = 50*0.1 + 50*0.9 = 50

$C_2 = a_{2,mel}$ P(melanoma | x) + $a_{2,naev}$ P(benign nevus | x) = 100000*0.1 + 0*0.9 = 10000

-> we choose for the action with lower costs, i.e. ‘remove’

Summary of concepts and facts

- Prior probability
- Class conditional probability density function, likelihood
- Posterior probability
- Bayes formula/rule for posteriors
- Bayes decision rule
- Minimum loss/risk classification