

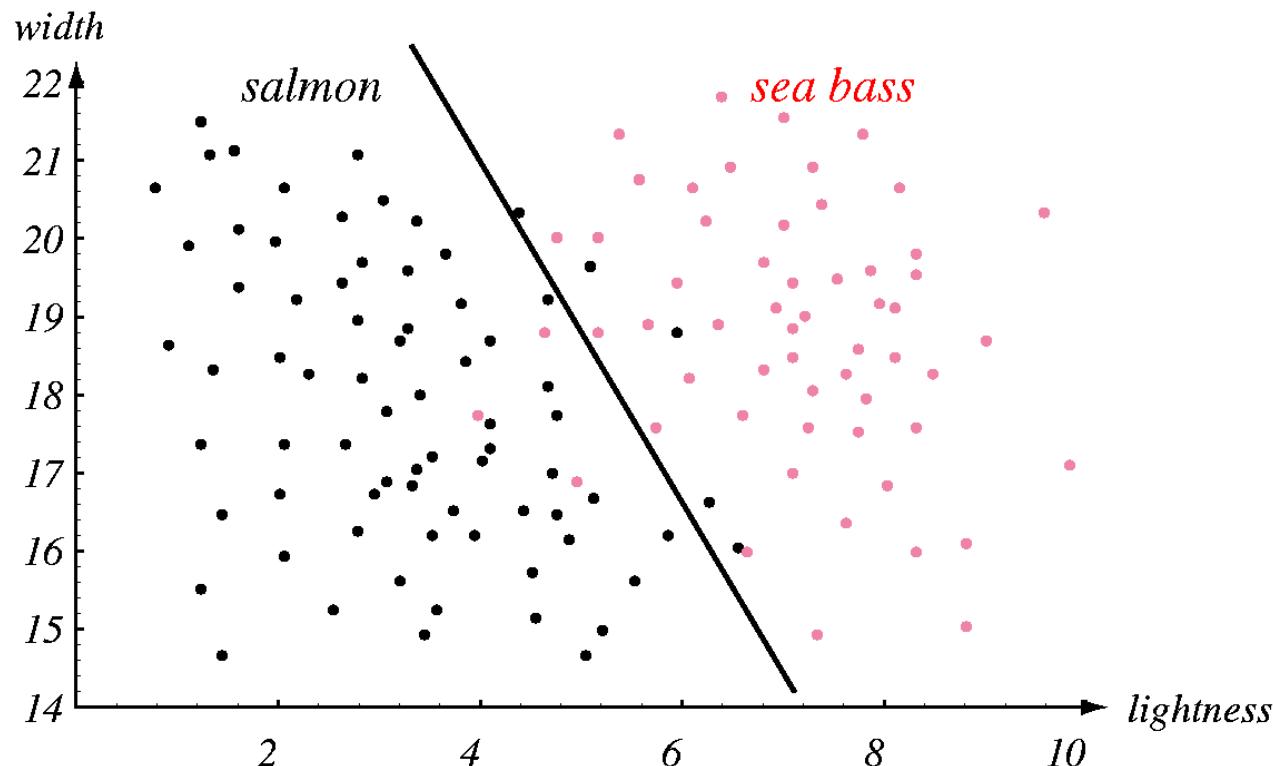
Elements of Bayesian Decision Theory

INTRODUCTION TO INTELLIGENT SYSTEMS 17/18

Probabilistic approach to classification

For each point, estimate the probability for each class.

Choose the class with the highest probability.



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

Priors

Classes

ω_1 - sea bass

ω_2 - salmon

a two-class problem

A priory probabilities (or prior probabilities)

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

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

A simple decision rule

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

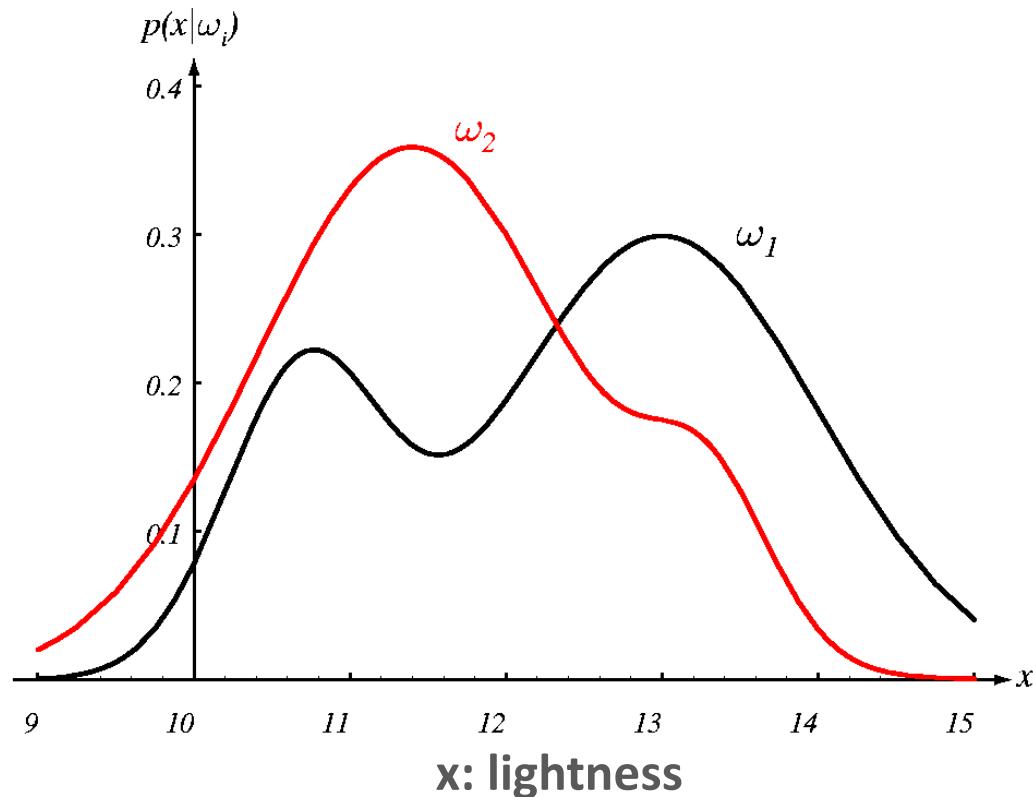
Class conditional probability density function and likelihood

Likelihood

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

$$p(x | \omega_1)$$

$$p(x | \omega_2)$$



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

Bayes formula/rule

$$p(x, \omega_j) = p(x | \omega_j) P(\omega_j) \quad \text{Joint probability}$$

$$p(x, \omega_j) = P(\omega_j | x) p(x)$$

$$P(\omega_j | x) p(x) = p(x | \omega_j) P(\omega_j)$$

$$P(\omega_j | x) = \frac{p(x | \omega_j) P(\omega_j)}{p(x)} \quad \text{Bayes rule}$$

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

$$posterior = \frac{likelihood \times prior}{evidence}$$

Bayes decision rule

Probability of making an error:

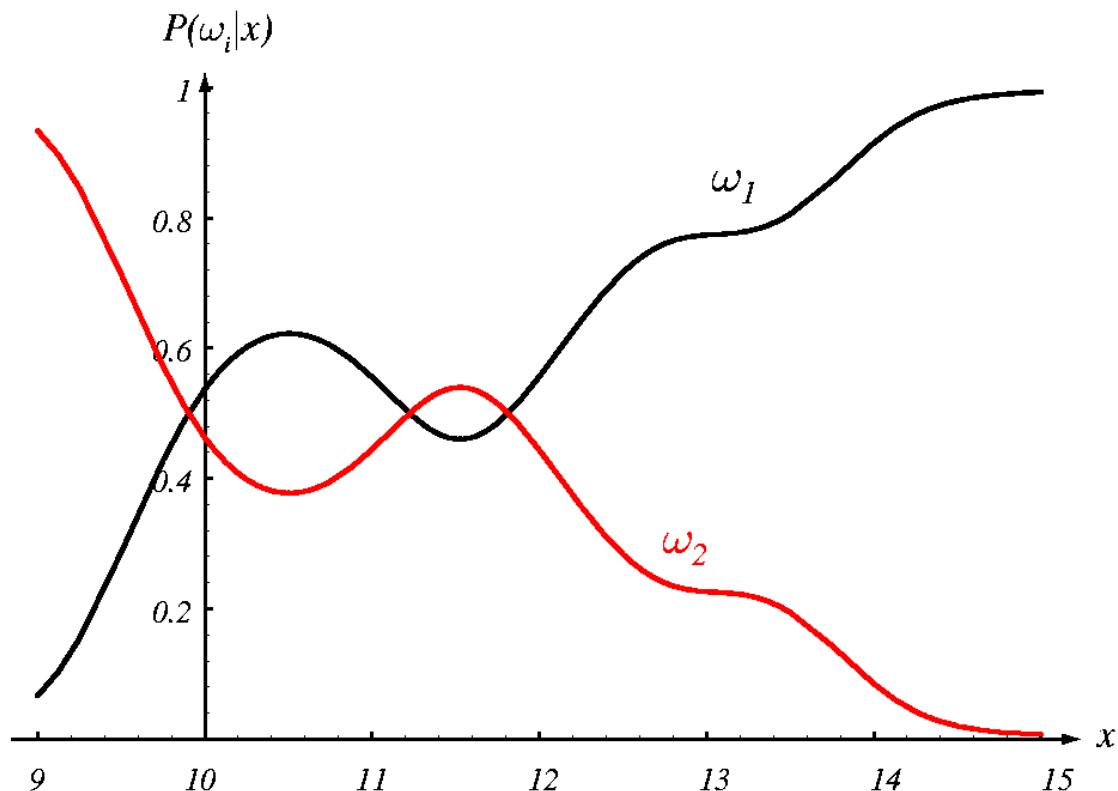
$$P(error | x) = \begin{cases} P(\omega_1 | x), & \text{if we decide } \omega_2 \\ P(\omega_2 | x), & \text{if we decide } \omega_1 \end{cases}$$

Bayes decision rule:

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

Posterior probability plots

Use priors as coefficients of likelihoods and normalize so that their sum is 1 for any x



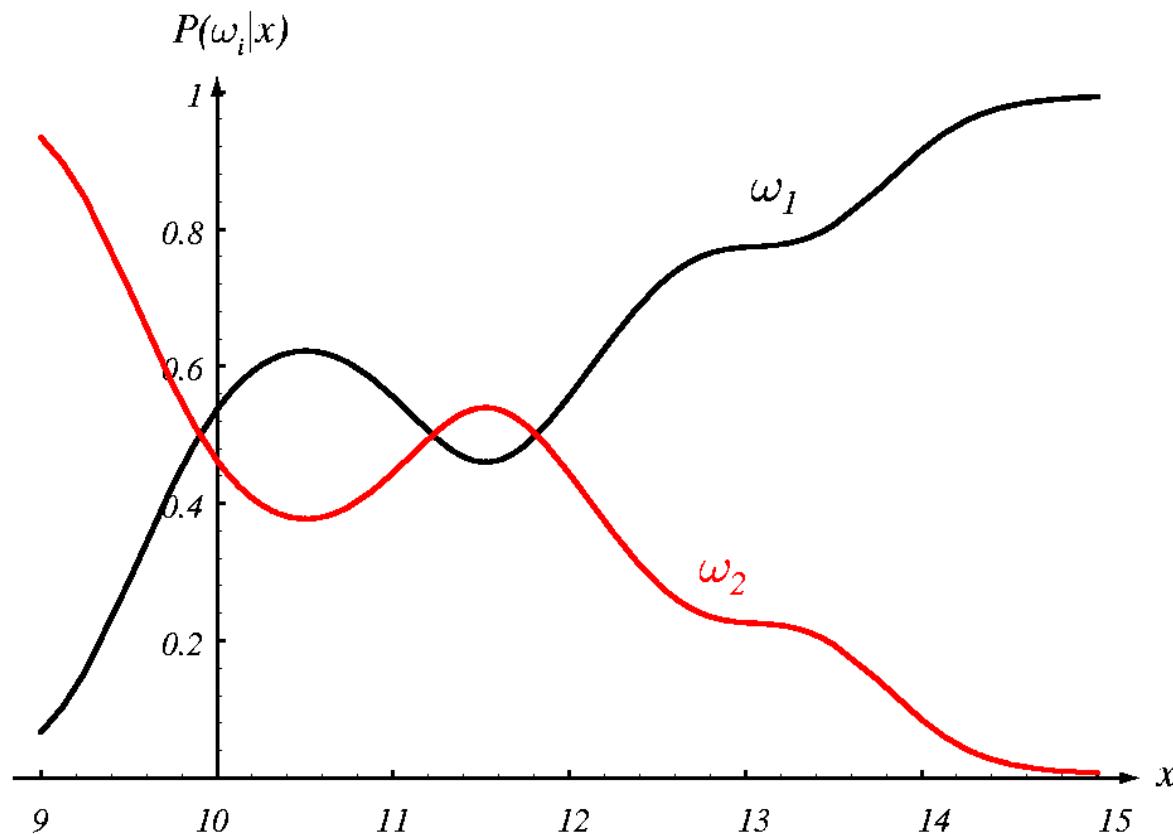
$$P(\omega_1) = 2/3$$

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

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

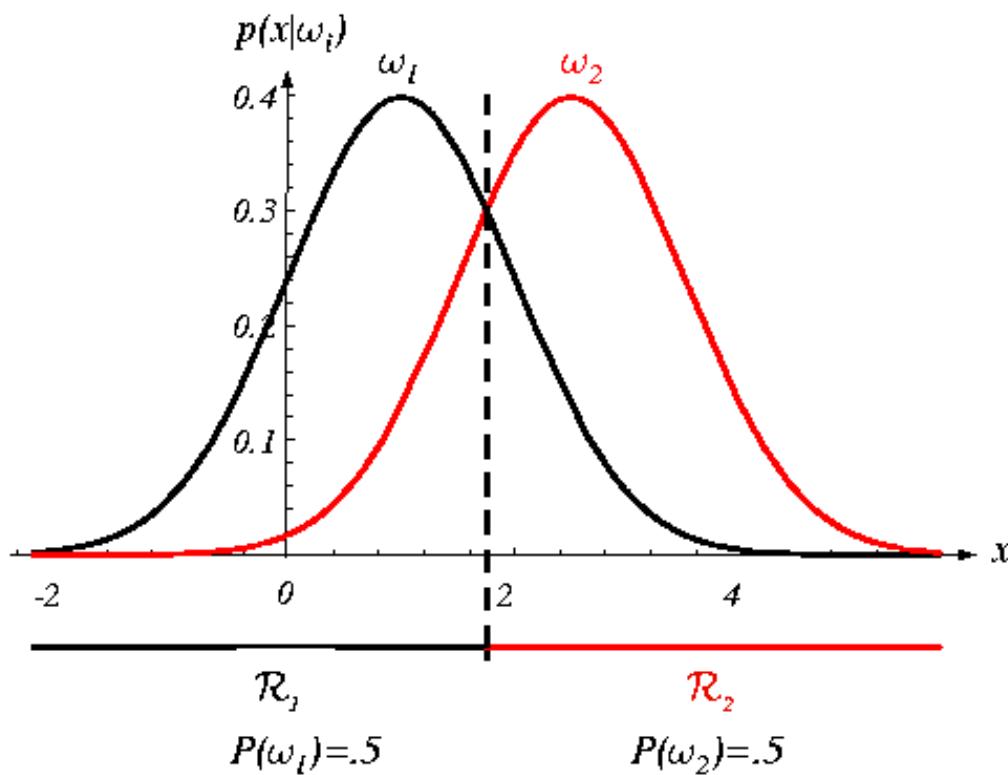
Error probability of Bayes decision rule

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



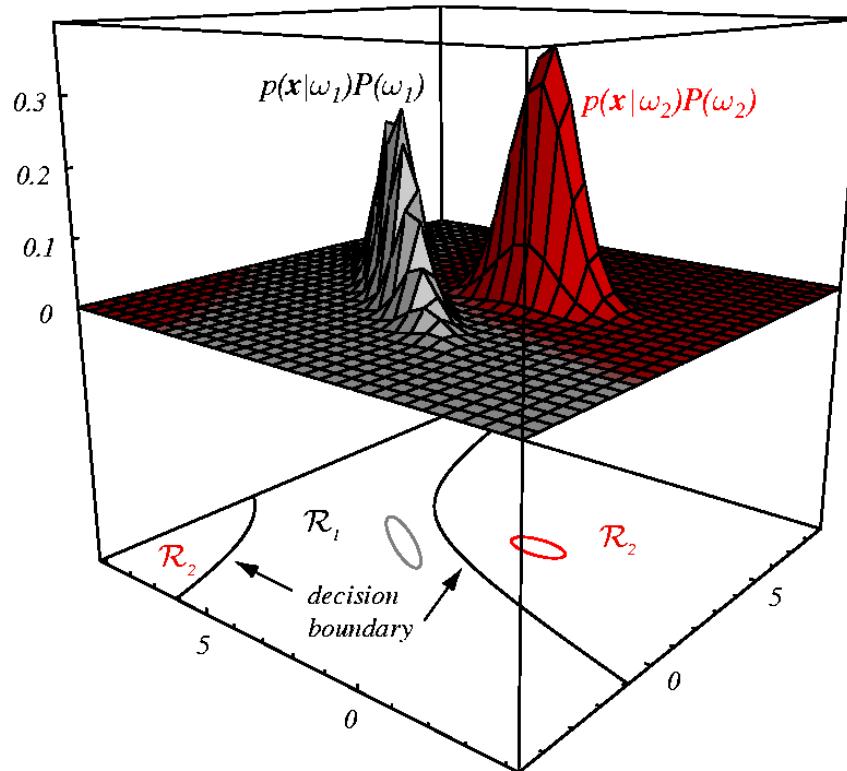
(from Duda, Hart, Stork (2001) Pattern classification)

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 space



from Duda, Hart, Stork (2001) Pattern classification

Generalizations of Bayesian Decision Theory

We replace the scalar x with the feature vector $\underline{x} \in \mathbf{R}^d$

We introduce a cost or a loss function λ which states how costly each classification decisions is.

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

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

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

Bayes formula

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

Evidence

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

Bayesian decision theory

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

$$R(\alpha_i | \underline{x}) = \sum_{j=1}^c \lambda(\alpha_i | \omega_j) P(\omega_j | \underline{x})$$

Rule to minimize the expected loss:

- *Select that action which minimizes the conditional risk.*

Generalized Bayesian decision theory

Let $P(\text{melanoma} | x) = 0.1$ and $P(\text{benign nevus} | x) = 0.9$

Bayesian classification: **benign nevus** (since it has higher probability)

Let now consider the actions: α_1 – remove, α_2 – do not remove, with costs

$$\lambda(\alpha_1 | \text{mel}) = 50 \quad \lambda(\alpha_1 | \text{nev}) = 50$$

$$\lambda(\alpha_2 | \text{mel}) = 100000 \quad \lambda(\alpha_2 | \text{nev}) = 0$$

Expected cost R_i as weighted average over many cases with same x :

$$R_1 = \lambda(\alpha_1 | \text{mel})P(\text{mel} | x) + \lambda(\alpha_1 | \text{nev})P(\text{nev} | x) = 50*0.1 + 50*0.9 = 50$$

$$R_2 = \lambda(\alpha_2 | \text{mel})P(\text{mel} | x) + \lambda(\alpha_2 | \text{nev})P(\text{nev} | x) = 100000*0.1 + 0*0.9 = 10000$$

-> we choose for the action with lower cost: α_1 - ‘remove’

Rule to minimize the expected loss:

Select that action which minimizes the conditional risk.