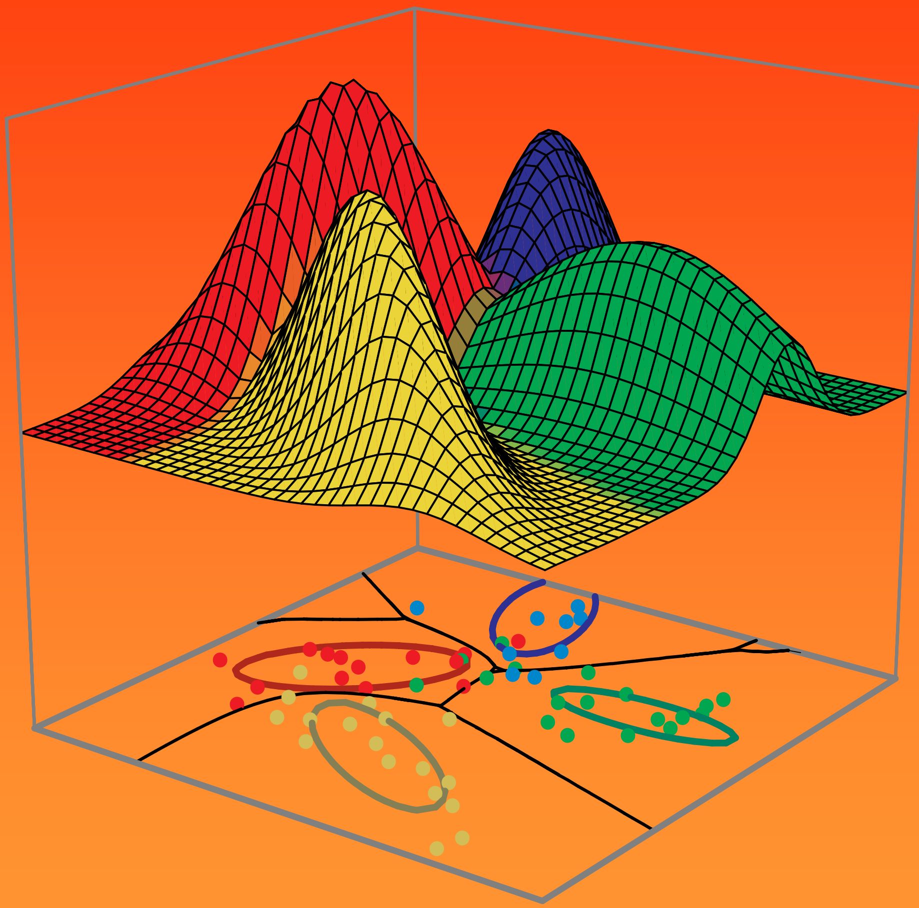

Pattern Classification

All materials in these slides were taken from Pattern Classification (2nd ed) by R. O. Duda, P. E. Hart and D. G. Stork, John Wiley & Sons, 2000
with the permission of the authors and the publisher

Chapter 2 (Part 1): Bayesian Decision Theory (Sections 2.1-2.2)

- Introduction
- Bayesian Decision Theory–Continuous Features

Introduction

- The sea bass/salmon example
  - State of nature, prior
    - State of nature is a random variable
    - The catch of salmon and sea bass is equiprobable
      - $P(\omega_1) = P(\omega_2)$ (uniform priors)
      - $P(\omega_1) + P(\omega_2) = 1$ (exclusivity and exhaustivity)

- Decision rule with only the prior information
  - Decide ω_1 if $P(\omega_1) > P(\omega_2)$ otherwise decide ω_2

- Use of the class –conditional information

- $P(x \mid \omega_1)$ and $P(x \mid \omega_2)$ describe the difference in lightness between populations of sea and salmon

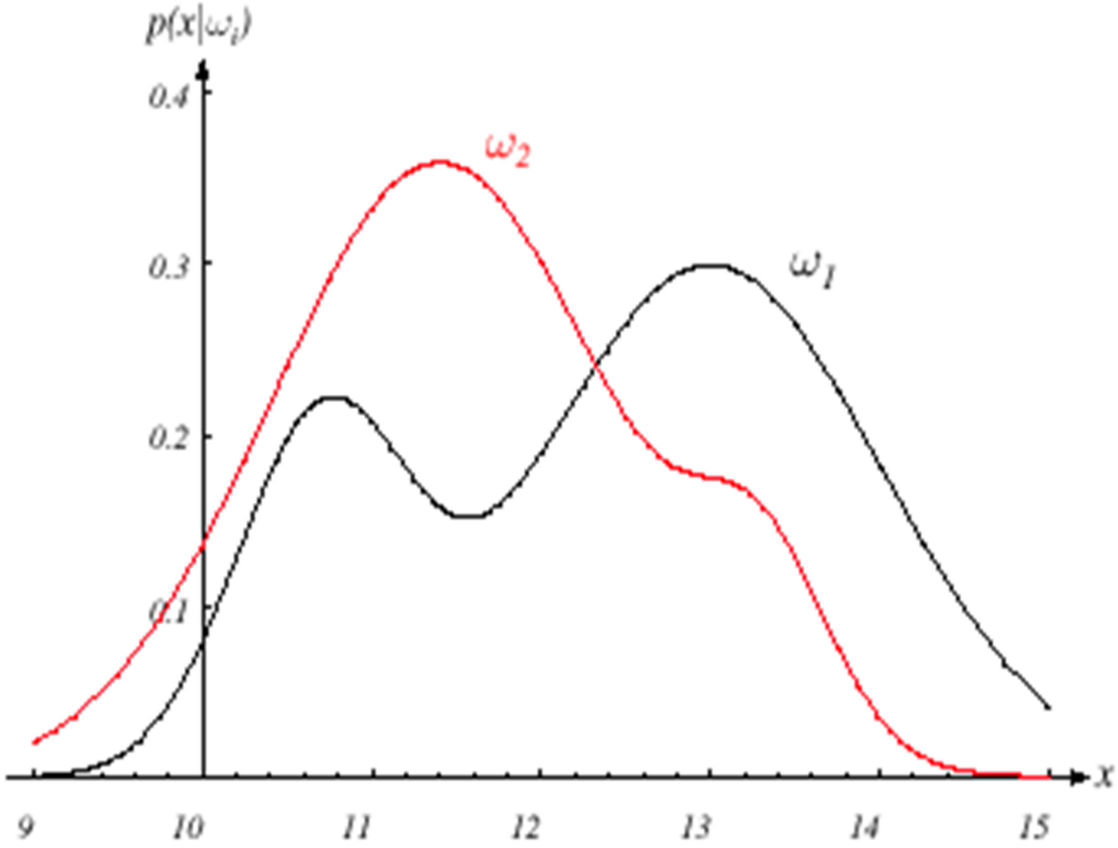

FIGURE 2.1. Hypothetical class-conditional probability density functions show the probability density of measuring a particular feature value x given the pattern is in category ω_i. If x represents the lightness of a fish, the two curves might describe the difference in lightness of populations of two types of fish. Density functions are normalized, and thus the area under each curve is 1.0. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

- Posterior, likelihood, evidence
  - $P(\omega_j \mid x) = P(x \mid \omega_j)P(\omega_j) / P(x)$ (Bayes formula)
  - Where in case of two categories

$$P(x) = \sum_{j=1}^{j=2} P(x \mid \omega_j)P(\omega_j)$$

  - Posterior = (Likelihood * Prior) / Evidence

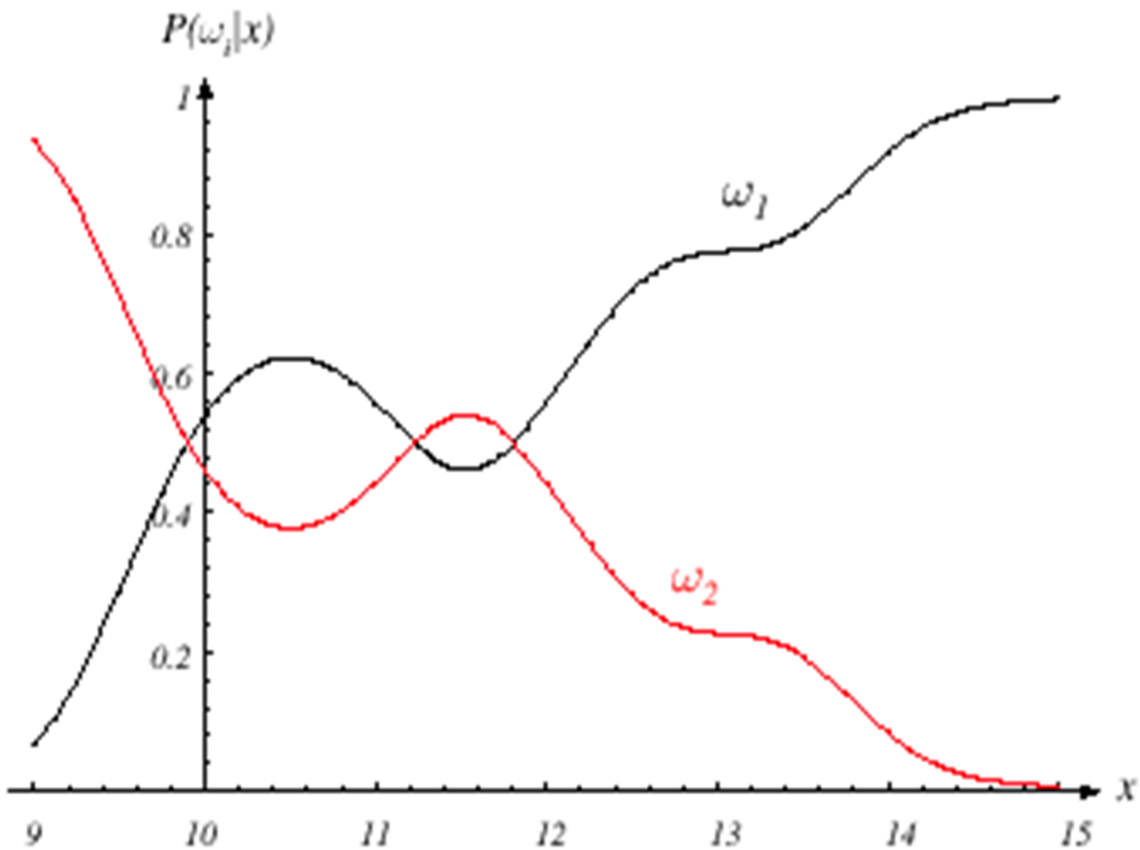

FIGURE 2.2. Posterior probabilities for the particular priors $P(\omega_1) = 2/3$ and $P(\omega_2) = 1/3$ for the class-conditional probability densities shown in Fig. 2.1. Thus in this case, given that a pattern is measured to have feature value $x = 14$, the probability it is in category ω_2 is roughly 0.08, and that it is in ω_1 is 0.92. At every x, the posteriors sum to 1.0. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*.

- Decision given the posterior probabilities

X is an observation for which:

if $P(\omega_1 \mid x) > P(\omega_2 \mid x)$ ⇒ True state of nature = ω_1
if $P(\omega_1 \mid x) < P(\omega_2 \mid x)$ ⇒ True state of nature = ω_2

Therefore:

whenever we observe a particular x, the probability of error is :

$$P(error \mid x) = P(\omega_1 \mid x) \text{ if we decide } \omega_2$$
$$P(error \mid x) = P(\omega_2 \mid x) \text{ if we decide } \omega_1$$

- Minimizing the probability of error

- Decide ω_1 *if* $P(\omega_1 \mid x) > P(\omega_2 \mid x)$; otherwise decide ω_2

  Therefore:

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

(Bayes decision)

Bayesian Decision Theory – Continuous Features

- Generalization of the preceding ideas
  - Use of more than one feature
  - Use more than two states of nature
  - Allowing actions and not only decide on the state of nature
  - Introduce a loss of function which is more general than the probability of error

- Allowing actions other than classification primarily allows the possibility of rejection
  - Rejection in the sense of abstention
  - Don’t make a decision if the alternatives are too close
  - This must be tempered by the cost of indecision

- The loss function states how costly each action taken is

Let $\{\omega_1, \omega_2,\ldots, \omega_c\}$ be the set of c states of nature (or “categories”)

Let $\{\alpha_1, \alpha_2,\ldots, \alpha_a\}$ be the set of possible actions

Let $\lambda(\alpha_i \mid \omega_j)$ be the loss incurred for taking

action α_i when the state of nature is ω_j

Overall risk

R = Sum of all $R(\alpha_i \mid x)$ *for i = 1,…,a* ***and all x***

Conditional risk

Minimizing R ⟺ Minimizing $R(\alpha_i \mid x)$ *for i = 1,…, a*

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

for each action α_i (i = 1,…,a)

Note: This is the risk specifically for observation *x*

Select the action α_i for which $R(\alpha_i \mid x)$ is minimum

R is minimum and R in this case is called the Bayes risk = best performance that can be achieved!

- Two-category classification

α_1 : deciding ω_1

α_2 : deciding ω_2

$\lambda_{ij} = \lambda(\alpha_i \mid \omega_j)$

loss incurred for deciding ω_i when the true state of nature is ω_j

Conditional risk:

$$R(\alpha_1 \mid x) = \lambda_{11}P(\omega_1 \mid x) + \lambda_{12}P(\omega_2 \mid x)$$
$$R(\alpha_2 \mid x) = \lambda_{21}P(\omega_1 \mid x) + \lambda_{22}P(\omega_2 \mid x)$$

Our rule is the following:

if $R(\alpha_1 \mid x) < R(\alpha_2 \mid x)$

action α_1: "decide ω_1" is taken

Substituting the def. of R() we have :

decide ω_1 if:

$$\lambda_{11} P(\omega_1 \mid x) + \lambda_{12} P(\omega_2 \mid x) < \lambda_{21} P(\omega_1 \mid x) + \lambda_{22} P(\omega_2 \mid x)$$

and decide ω_2 otherwise

We can rewrite

$$\lambda_{11} P(\omega_1 \mid x) + \lambda_{12} P(\omega_2 \mid x) < \lambda_{21} P(\omega_1 \mid x) + \lambda_{22} P(\omega_2 \mid x)$$

As

$$(\lambda_{21} - \lambda_{11}) P(\omega_1 \mid x) > (\lambda_{12} - \lambda_{22}) P(\omega_2 \mid x)$$

Finally, we can rewrite

$$(\lambda_{21} - \lambda_{11})\, P(\omega_1 \mid x) > (\lambda_{12} - \lambda_{22})\, P(\omega_2 \mid x)$$

using Bayes formula and posterior probabilities to get:

decide ω_1 if:

$$(\lambda_{21} - \lambda_{11})\, P(x \mid \omega_1)\, P(\omega_1) > (\lambda_{12} - \lambda_{22})\, P(x \mid \omega_2)\, P(\omega_2)$$

and decide ω_2 otherwise

If $\lambda_{21} > \lambda_{11}$ then we can express our rule as a Likelihood ratio:

The preceding rule is equivalent to the following rule:

$$\text{if } \frac{P(x \mid \omega_1)}{P(x \mid \omega_2)} > \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \cdot \frac{P(\omega_2)}{P(\omega_1)}$$

Then take action α_1 (decide ω_1)

Otherwise take action α_2 (decide ω_2)

Optimal decision property

"If the likelihood ratio exceeds a threshold value independent of the input pattern x, we can take optimal actions"

Exercise

Select the optimal decision where:

$\Omega = \{\omega_1, \omega_2\}$

$P(x \mid \omega_1)$ → N(2, 0.5) (Normal distribution)

$P(x \mid \omega_2)$ → N(1.5, 0.2)

$P(\omega_1) = 2/3$

$P(\omega_2) = 1/3$

$$\lambda = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$