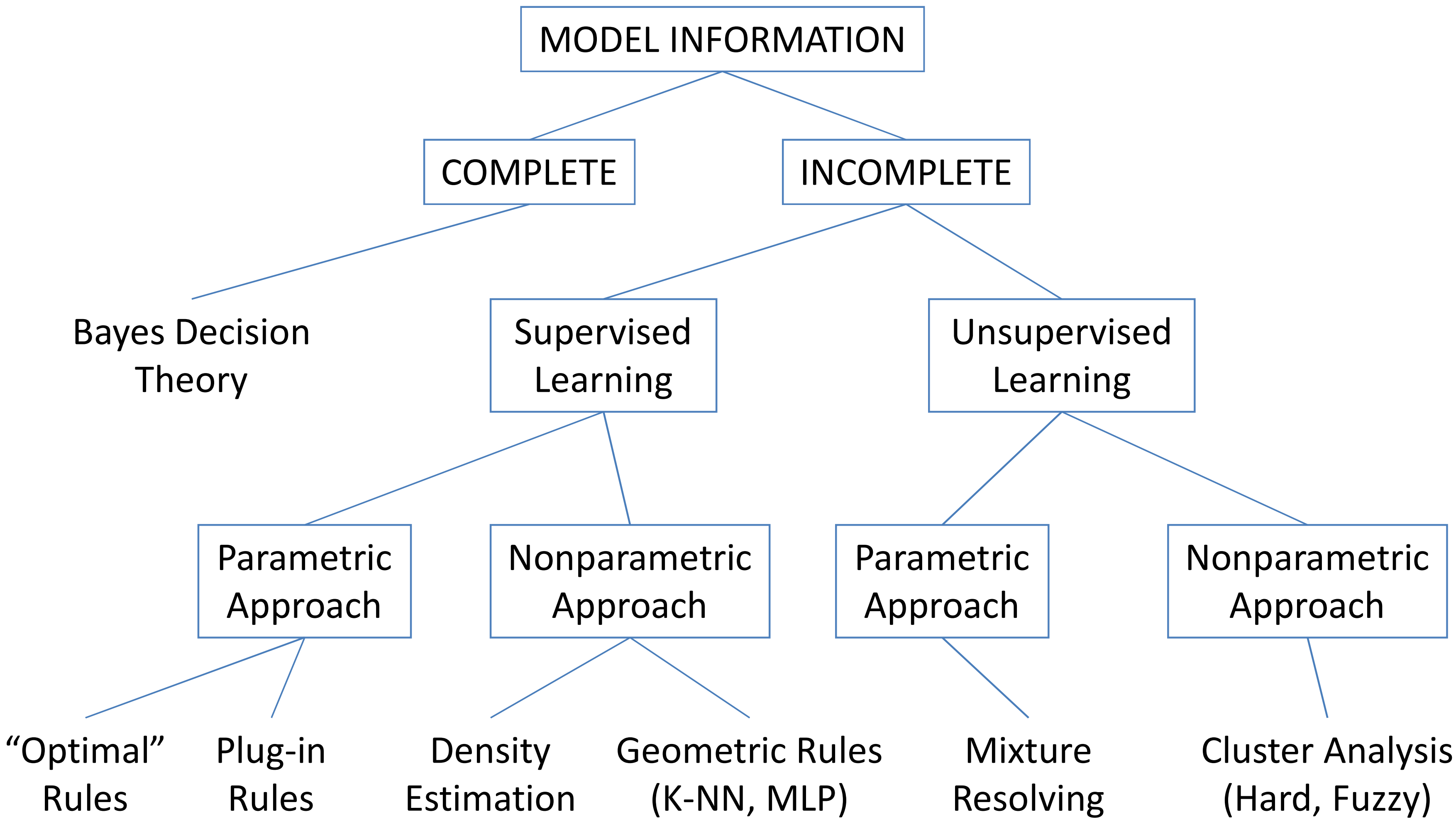
Course Outline



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

Decision problem posed in probabilistic terms

Bayesian Decision Theory–Continuous Features

All the relevant probability values are known

Introduction

- The sea bass/salmon example
  - State of nature, a priori (prior) probability
    - State of nature is unpredictable, so it 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)
    - Prior prob. reflects our prior knowledge about how likely we are to observe a sea bass or salmon; these probabilities may depend on time of the year or the fishing area!

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

- Suppose now we have a measurement or feature on the state of nature - say the fish lightness value

- Use of the class –conditional probability density

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

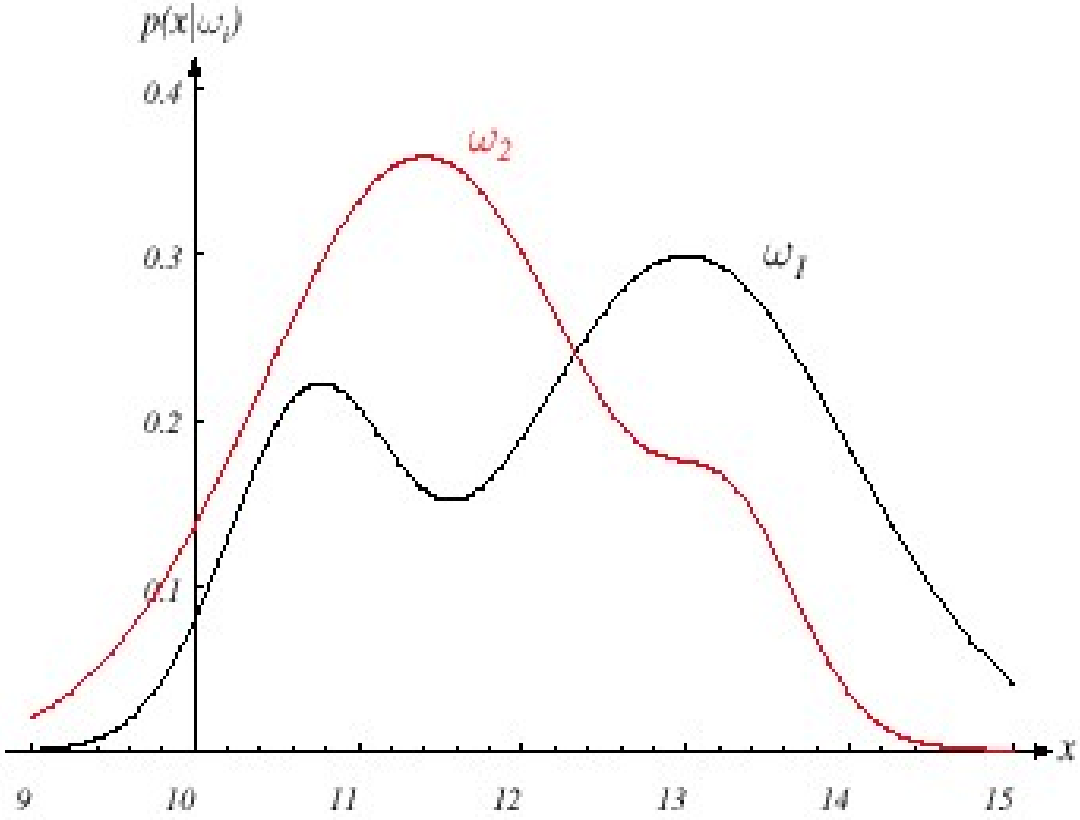


FIGURE 2.1. Hypothetical class-conditional probability density functions show the probability density of measuring a particular feature value x given the pattern is in category ω_i. If x represents the lightness of a fish, the two curves might describe the difference in lightness of populations of two 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.

Amount of overlap between the densities determines the "goodness" of feature

- Maximum likelihood decision rule
  - Assign input pattern x to class ω_1 if

$$P(x \mid \omega_1) > P(x \mid \omega_2), \text{ otherwise } \omega_2$$

- How does the feature x influence our attitude (prior) concerning the true state of nature?

- Posterior, likelihood, evidence
  - Using *Bayes formula*
  - $P(\omega_j \mid x) = \{P(x \mid \omega_j) \,.\, P(\omega_j)\} / P(x)$

    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
  - Evidence *P(x)* can be viewed as a scale factor that guarantees that the posterior probabilities sum to 1
  - $P(x \mid \omega_j)$ is called the likelihood of ω_j *with respect to x; the category* ω_j *for which* $P(x \mid \omega_j)$ is large is more likely to be the true category

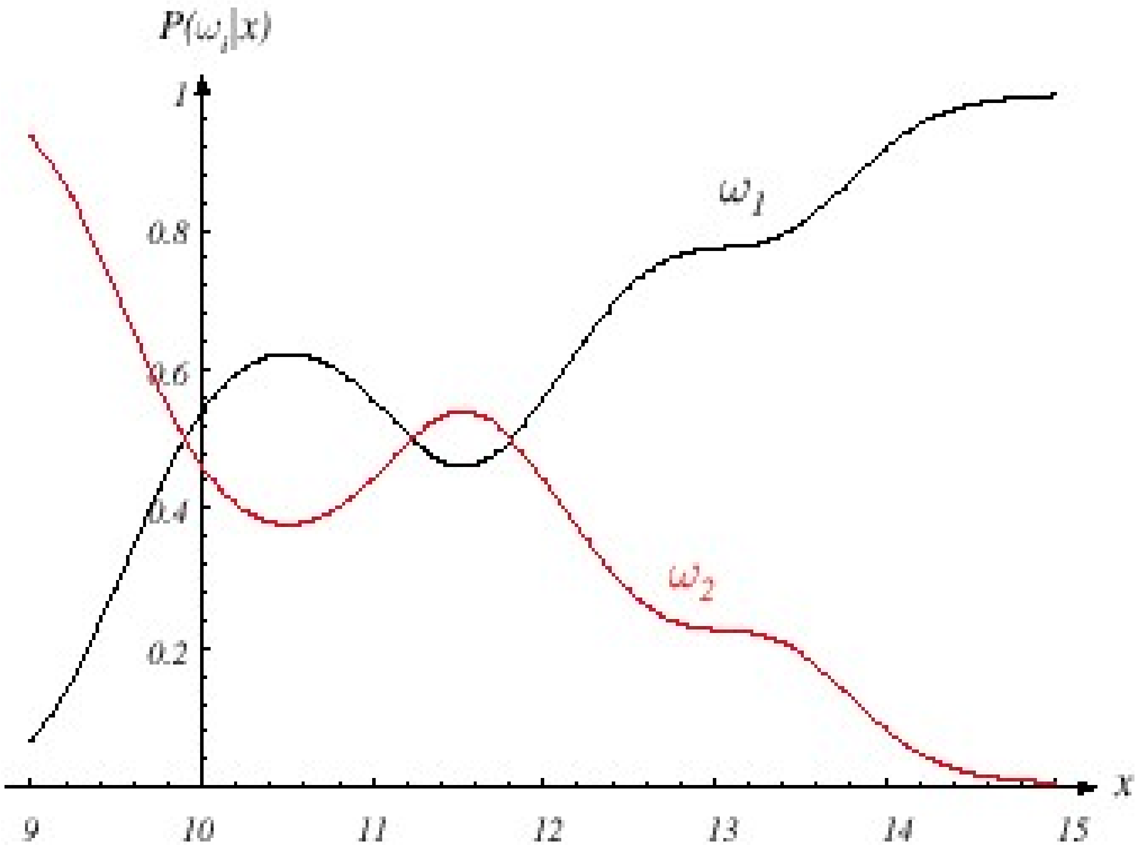


FIGURE 2.2. Posterior probabilities for the particular priors $P(\omega_1) = 2/3$ and $P(\omega_2) = 1/3$ for the class-conditional probability densities shown in Fig. 2.1. Thus in this case, given that a pattern is measured to have feature value $x = 14$, the probability it is in category ω_2 is roughly 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*. Copyright © 2001 by John Wiley & Sons, Inc.

- $P(\omega_1 \mid x)$ is the probability of the state of nature being ω_1 given that feature value x has been observed

- Decision given the posterior probabilities, Optimal Bayes Decision rule

X is an observation for which:

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

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

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

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

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

- Bayes decision rule minimizes the probability of error

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

  Therefore:

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

- Unconditional error, *P(error) obtained by integration over all x w.r.t. p(x)*

- Optimal Bayes decision rule

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

- Special cases:

  (i) $P(\omega_1) = P(\omega_2)$; *Decide* ω_1 if

  $P(x \mid \omega_1) > P(x \mid \omega_2)$, otherwise ω_2

  (ii) $P(x \mid \omega_1) = P(x \mid \omega_2)$; *Decide* ω_1 if

  $P(\omega_1) > P(\omega_2)$, otherwise ω_2

Bayesian Decision Theory – Continuous Features

- Generalization of the preceding ideas
  - Use of more than one feature (d features)
  - Use of more than two states of nature (c classes)
  - Allowing other actions besides deciding on the state of nature
  - Introduce a loss function which is more general than the probability of error

- Allowing actions other than classification primarily allows the possibility of rejection

- Refusing to make a decision in close or noisy cases!

- 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 *a* possible actions

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

  action α_i when the true state of nature is ω_j

- General decision rule $\alpha_{(x)}$ *specifies which action to take for every possible observation x*

Overall risk

R = Sum of all $\underbrace{R(\alpha_i \mid x)}_{\textbf{Conditional risk}}$ *for i = 1,…,a*

Minimizing R $\Longleftrightarrow$ 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 i = 1,…,a

Select the action for which the conditional risk $R(\alpha_i \mid x)$ *is minimum*

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

$\Rightarrow$ Risk 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:

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

Bayes rule is the following:

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

action α_1: "decide ω_1" is taken

This results in the equivalent rule:

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

Likelihood ratio:

The preceding rule is equivalent to the following rule:

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

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 that is independent of the input pattern x, we can take optimal actions"

For the maximum likelihood rule, the threshold value is 1.

Exercise

Select the optimal decision where:

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

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

$P(x \mid \omega_2) \longrightarrow$ N(1.5, 0.2)

$P(\omega_1) = 2/3$

$P(\omega_2) = 1/3$

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