Pattern Recognition

(from Prof. Bebis)

Introduction to Bayesian Decision Theory

(from Duda et al.)

Bayesian Decision Theory

- Fundamental statistical approach to problem classification.
- Quantifies the tradeoffs between various *classification decisions* using probabilities and the *costs* associated with such decisions.
  - Each action is associated with a cost or risk.
  - The simplest risk is the classification error.
  - Design classifiers to recommend actions that minimize some total expected risk.

Terminology

(using sea bass – salmon classification example)

- State of nature ω *(random variable):*
  - ω_1 for sea bass, ω_2 for salmon.
- Probabilities $P(\omega_1)$ and $P(\omega_2)$ *(priors)*
  - prior knowledge of how likely is to get a sea bass or a salmon
- Probability density function $p(x)$ *(evidence):*
  - how frequently we will measure a pattern with feature value x (e.g., x is a lightness measurement)

Note: if x and y are different measurements, $p(x)$ and $p(y)$ correspond to different *pdfs*: $p_X(x)$ and $p_Y(y)$

Terminology (cont'd)

(using sea bass – salmon classification example)

- Conditional probability density $p(x/\omega_j)$ (*likelihood*):
  - how frequently we will measure a pattern with feature value x given that the pattern belongs to class ω_j

e.g., lightness distributions between salmon/sea-bass populations

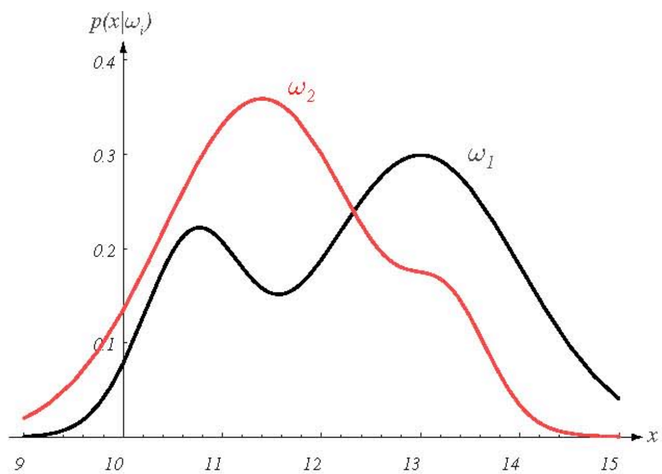


FIGURE 2.1. Hypothetical class-conditional probability density functions show the probability density of measuring a particular feature value x given the pattern is in category ω_i. If x represents the lightness of a fish, the two curves might describe the difference in lightness of populations of two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

Terminology (cont'd)

(using sea bass – salmon classification example)

- Conditional probability $P(\omega_j/x)$ *(posterior)*:
  - the probability that the fish belongs to class ω_j given measurement x.

Note: we will be using an uppercase *P(.)* to denote a probability mass function *(pmf)* and a lowercase *p(.)* to denote a probability density function *(pdf)*.

Decision Rule Using Priors Only

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

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

- Favours the most likely class … (optimum if no other info is available).
- This rule would be making the same decision all the times!
- Makes sense to use for judging just one fish …

Decision Rule Using Conditional pdf

- Using Bayes' rule, the posterior probability of category ω_j given measurement x is given by:

$$P(\omega_j / x) = \frac{p(x/\omega_j)P(\omega_j)}{p(x)} = \frac{likelihood \times prior}{evidence}$$

where $$p(x) = \sum_{j=1}^{2} p(x/\omega_j)P(\omega_j)$$ (scale factor – sum of probs = 1)

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

or

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

Decision Rule Using Conditional pdf (cont'd)

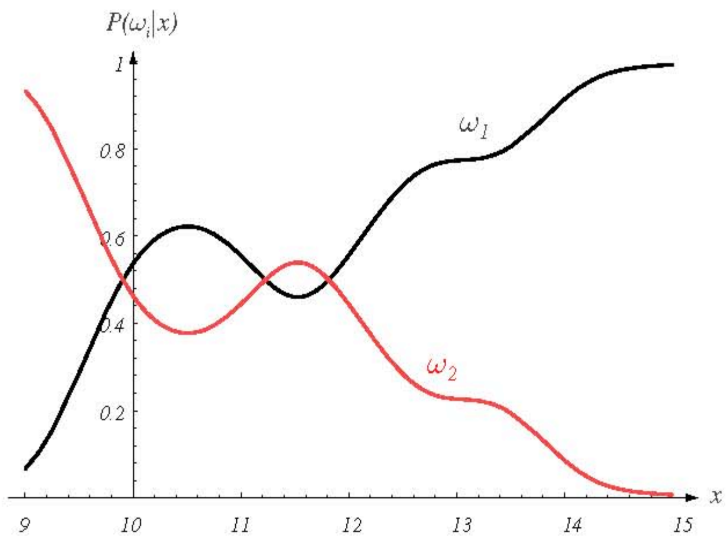


FIGURE 2.2. Posterior probabilities for the particular priors $P(\omega_1) = 2/3$ and $P(\omega_2) = 1/3$ for the class-conditional probability densities shown in Fig. 2.1. Thus in this case, given that a pattern is measured to have feature value $x = 14$, the probability it is in category ω_2 is roughly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

$$P(w_1) = \frac{2}{3} \qquad P(w_2) = \frac{1}{3}$$

Probability of Error

- The probability of error is defined as:

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

- The average probability error is given by:

$$P(error) = \int_{-\infty}^{\infty} P(error, x)dx = \int_{-\infty}^{\infty} P(error/x)\, p(x)dx$$

- The Bayes rule is *<u>optimum</u>*, that is, it minimizes the average probability error since:

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

Where do Probabilities Come From?

- The Bayesian rule is optimal if the *pmf* or *pdf* is known.
- There are two competitive answers to the above question:

(1) **Relative frequency** (objective) approach.

– Probabilities can only come from experiments.

(2) **Bayesian** (subjective) approach.

– Probabilities may reflect degree of belief and can be based on opinion as well as experiments.

Example

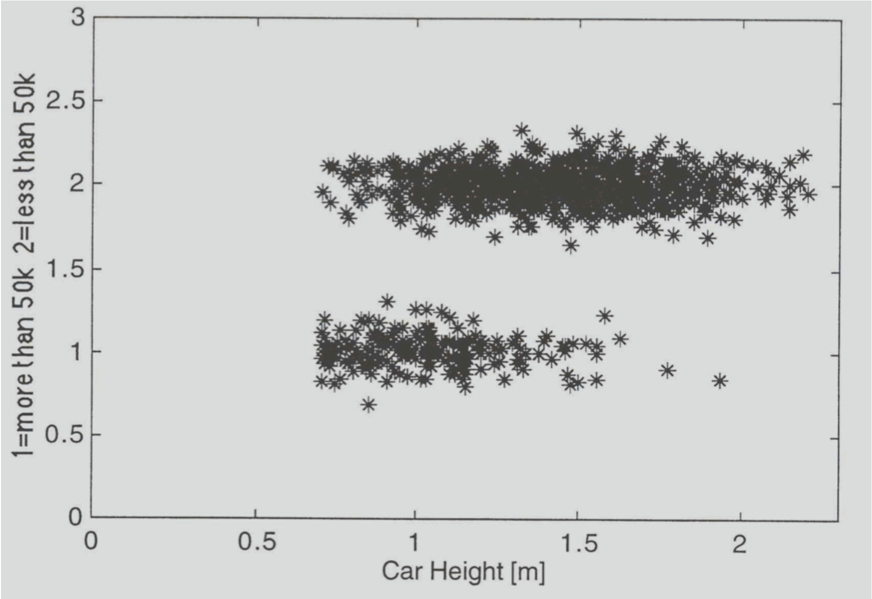
- Classify cars on UNR campus whether they are more or less than \$50K:
  - C1: price > \$50K
  - C2: price < \$50K
  - Feature x: height of car
- From Bayes' rule, we know how to compute the posterior probabilities:

$$P(C_i / x) = \frac{p(x / C_i)P(C_i)}{p(x)}$$

- Need to compute $p(x/C_1)$, $p(x/C_2)$, $P(C_1)$, $P(C_2)$

Example (cont'd)

- Determine prior probabilities
  - Collect data: ask drivers how much their car was and measure height.
  - e.g., 1209 samples: $\#C_1=221$ $\#C_2=988$

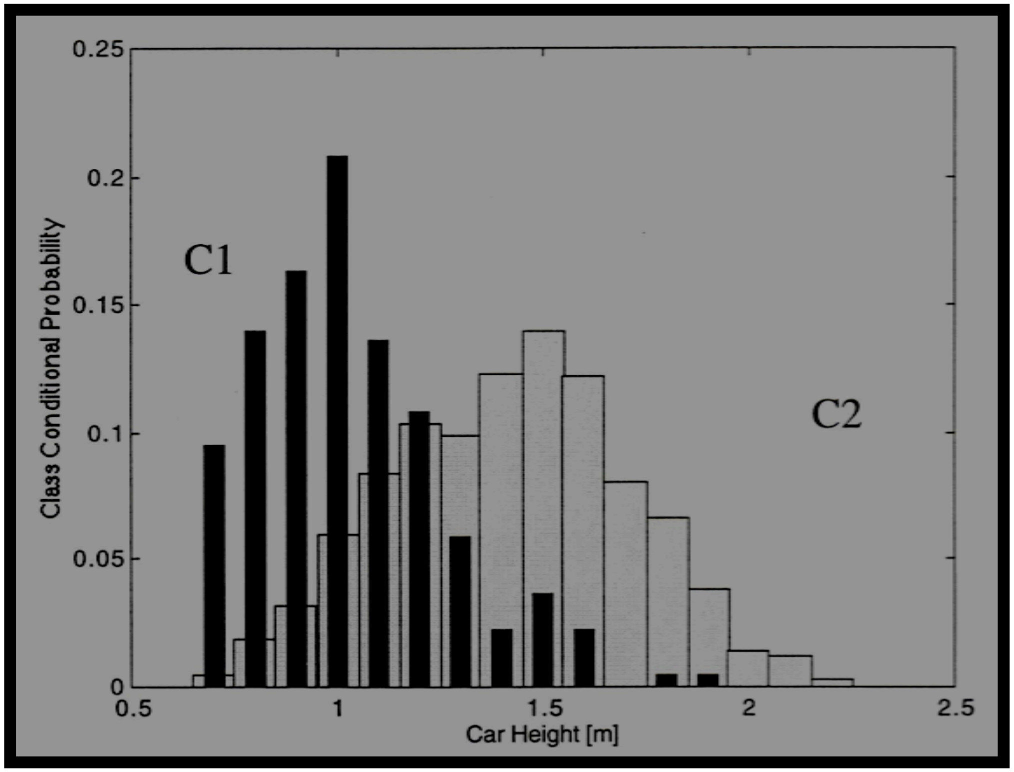


$$P(C_1) = \frac{221}{1209} = 0.183$$

$$P(C_2) = \frac{988}{1209} = 0.817$$

Example (cont'd)

- Determine class conditional probabilities (*likelihood*)
  - Discretize car height into bins and use normalized histogram



Example (cont'd)

- Calculate the posterior probability for each bin:

$$P(C_1 / x = 1.0) = \frac{p(x = 1.0 / C_1)\, P(C_1)}{p(x = 1.0 / C_1)\, P(C_1) + p(x = 1.0 / C_2)\, P(C_2)} =$$

$$= \frac{0.2081 * 0.183}{0.2081 * 0.183 + 0.0597 * 0.817} = 0.438$$