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 More Information: 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

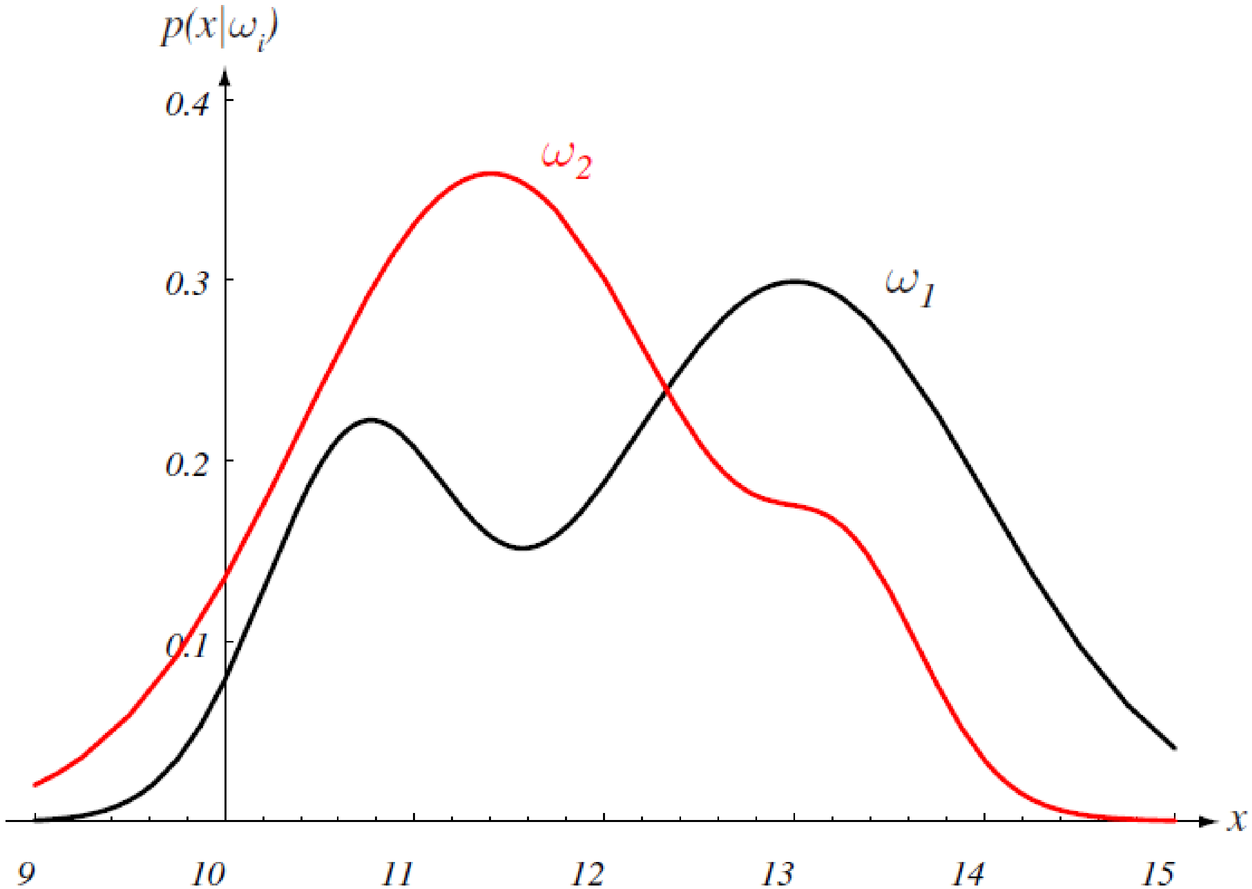


FIGURE 2.1. Hypothetical class-conditional probability density functions show the probability density of measuring a particular feature value x given the pattern is in category ω_i. If x represents the lightness of a fish, the two curves might describe the difference in lightness of populations of two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

  - $P(\omega_j \mid x) = p(x \mid \omega_j) \,.\, P(\omega_j) / p(x)$
  - 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 can be viewed as a scale factor

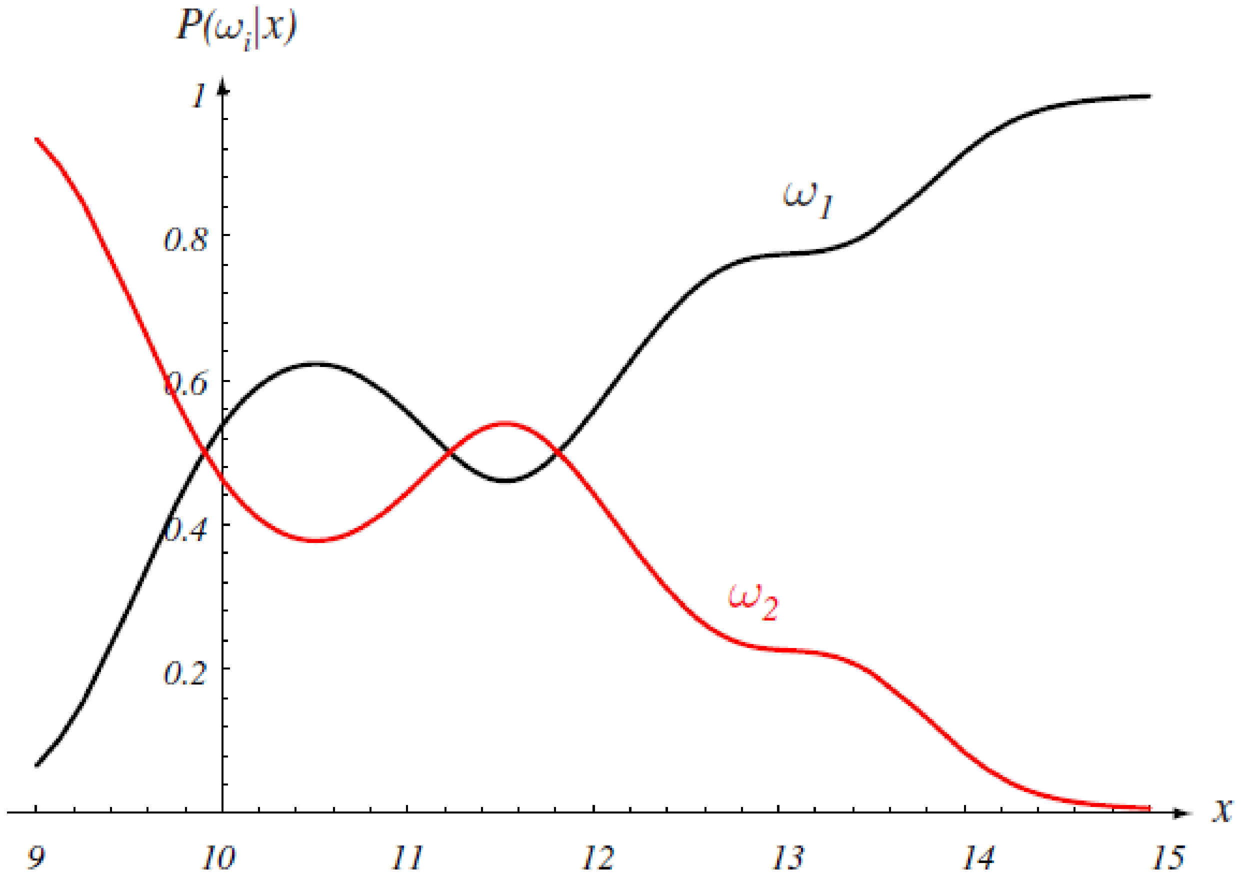
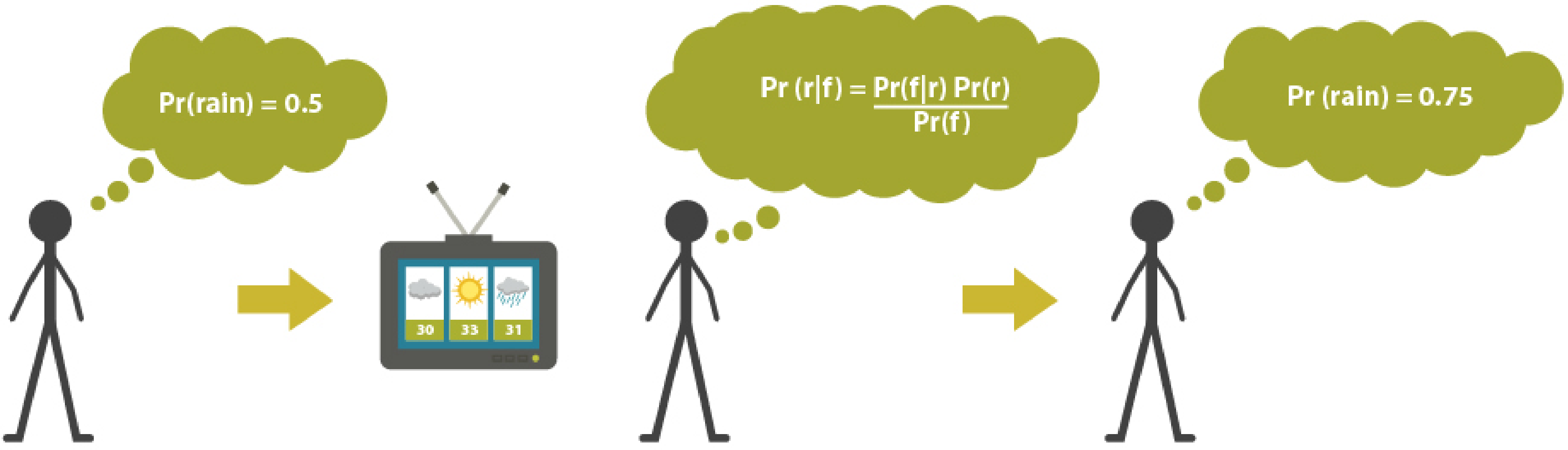


FIGURE 2.2. Posterior probabilities for the particular priors $P(\omega_1) = 2/3$ and $P(\omega_2) = 1/3$ for the class-conditional probability densities shown in Fig. 2.1. Thus in this case, given that a pattern is measured to have feature value $x = 14$, the probability it is in category ω_2 is roughly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

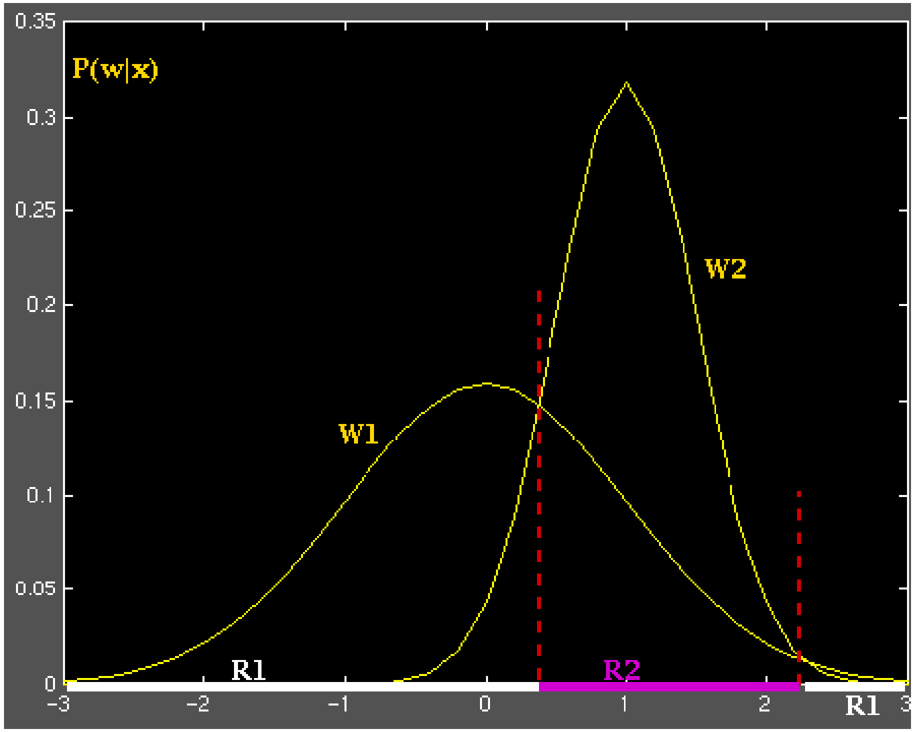
Posterior is a modification of prior

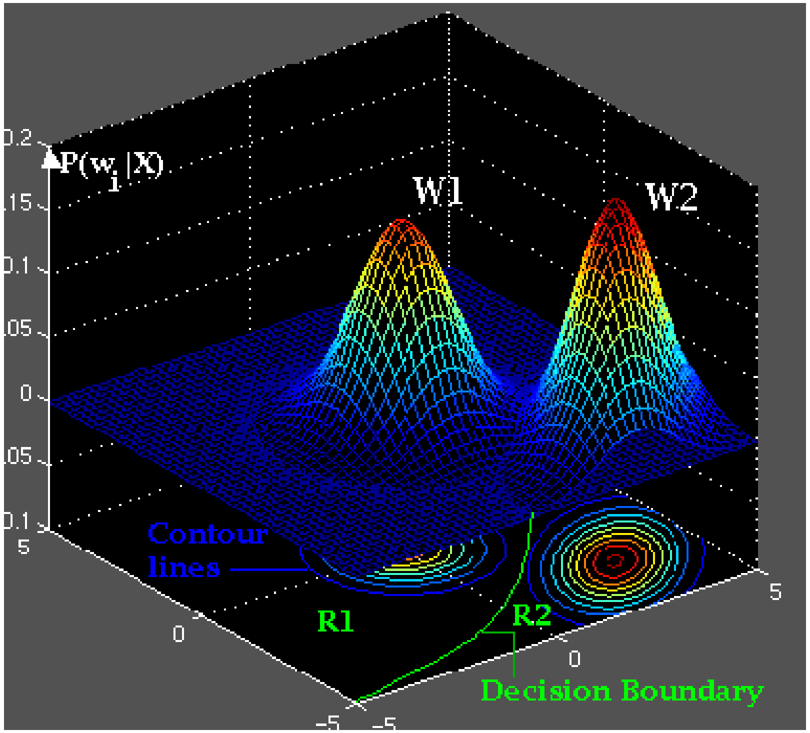
- The modification is caused by the likelihood



Pr(rain) = 0.5
30
33
31
Pr (r|f) = Pr(f|r) Pr(r) / Pr(f)
Pr (rain) = 0.75

Decision region





- Decision given the posterior probabilities

  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

  Multi-class ?

  Therefore，

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

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

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

$$\Downarrow$$

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

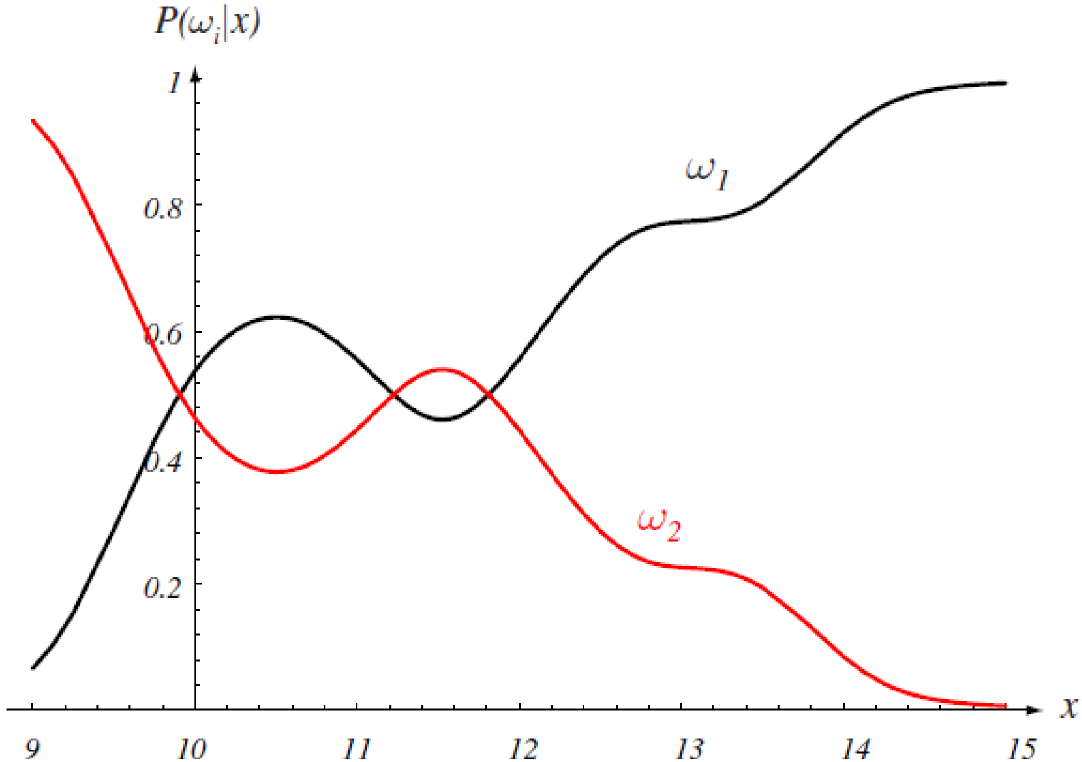


FIGURE 2.2. Posterior probabilities for the particular priors $P(\omega_1) = 2/3$ and $P(\omega_2) = 1/3$ for the class-conditional probability densities shown in Fig. 2.1. Thus in this case, given that a pattern is measured to have feature value $x = 14$, the probability it is in category ω_2 is roughly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

Multi-class

if $P(\omega_j \mid x) > P(\omega_i \mid x)$

Then the true state of nature = ω_j

- 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)

Decide ω_1 if $p(x \mid \omega_1) \mathrm{P}(\omega_1) > p(x \mid \omega_2) P(\omega_2)$

- Special Case:

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

$$\mathrm{P}(\omega_1) = P(\omega_2)$$

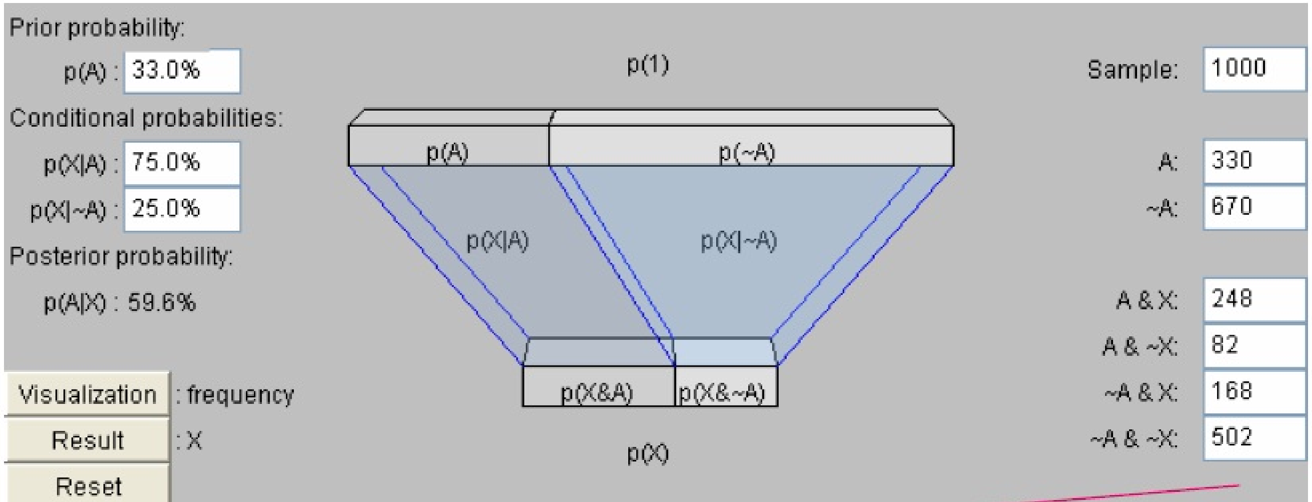
Interesting video

- http://weike.enetedu.com/play.asp?vodid=148661

- http://weike.enetedu.com/play.asp?vodid=141126

One example

Known



Data

By Conditional Probability Rule,

$$p(X/A) = \frac{p(X \,\&\, A)}{p(A)} = \frac{.248}{.330} = 0.7515$$

$$p(X/\sim A) = \frac{p(X \,\&\sim A)}{p(\sim A)} = \frac{.168}{.670} = 0.2507$$

By Bayes Rule,

$$P(A/X) = \frac{P(X/A)P(A)}{P(X)}$$

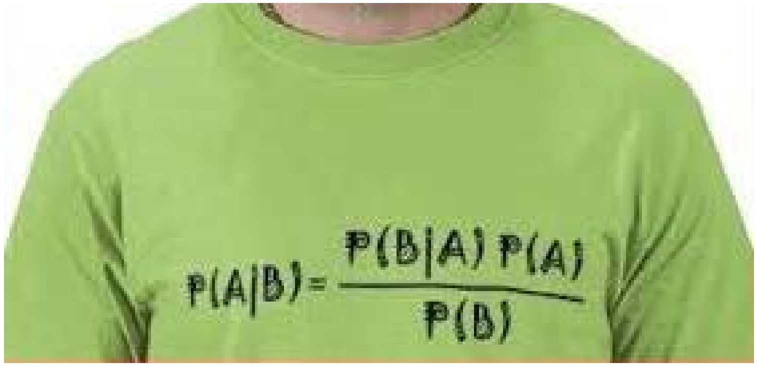
$$= \frac{P(X/A)P(A)}{P(X \,\&\, A) + P(X \,\&\sim A)}$$

$$= \frac{P(X/A)P(A)}{P(X/A)P(A) + P(X/\sim A)P(\sim A)}$$

$$= \frac{0.75 \times 0.33}{0.75 \times 0.33 + 0.25 \times 0.67}$$

$$= \frac{.2475}{.2475 + .1675} = \frac{.2475}{.415} = 0.596$$

Exercise

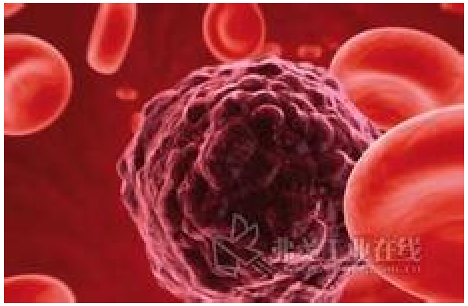


- **Problem**:

- A patient takes a lab test and the result is positive. The test returns a correct positive result in 98% of the cases in which the cancer disease is actually present, and a correct negative result in 97% of the cases in which the disease is not present. Furthermore, 0.8% of the entire population have this cancer disease. Does the patient suffers from the cancer?

- **Solution**:
  - Given: P(+|cancer)=0.98

    P(-|no cancer)=0.97



    P(cancer)=0.008

    P(-cancer)=0.992

  - Compute:

    P(no cancer | +) , P(cancer | +),

- $P(cancer \mid +)=P(+|cancer)* P(cancer)/p(+)$

- $P(cancer \mid +)=0.98X0.008/p (+)$

- $P(cancer \mid +)= 0.00784 /p(+)$

P(no cancer|+)=P(+|no cancer)*P(no cancer)/P(+)

P(no cancer|+)=(1-0.97)*(1-0.008)/ P(+)

– P(no cancer|+)=0.02976 / P(+)

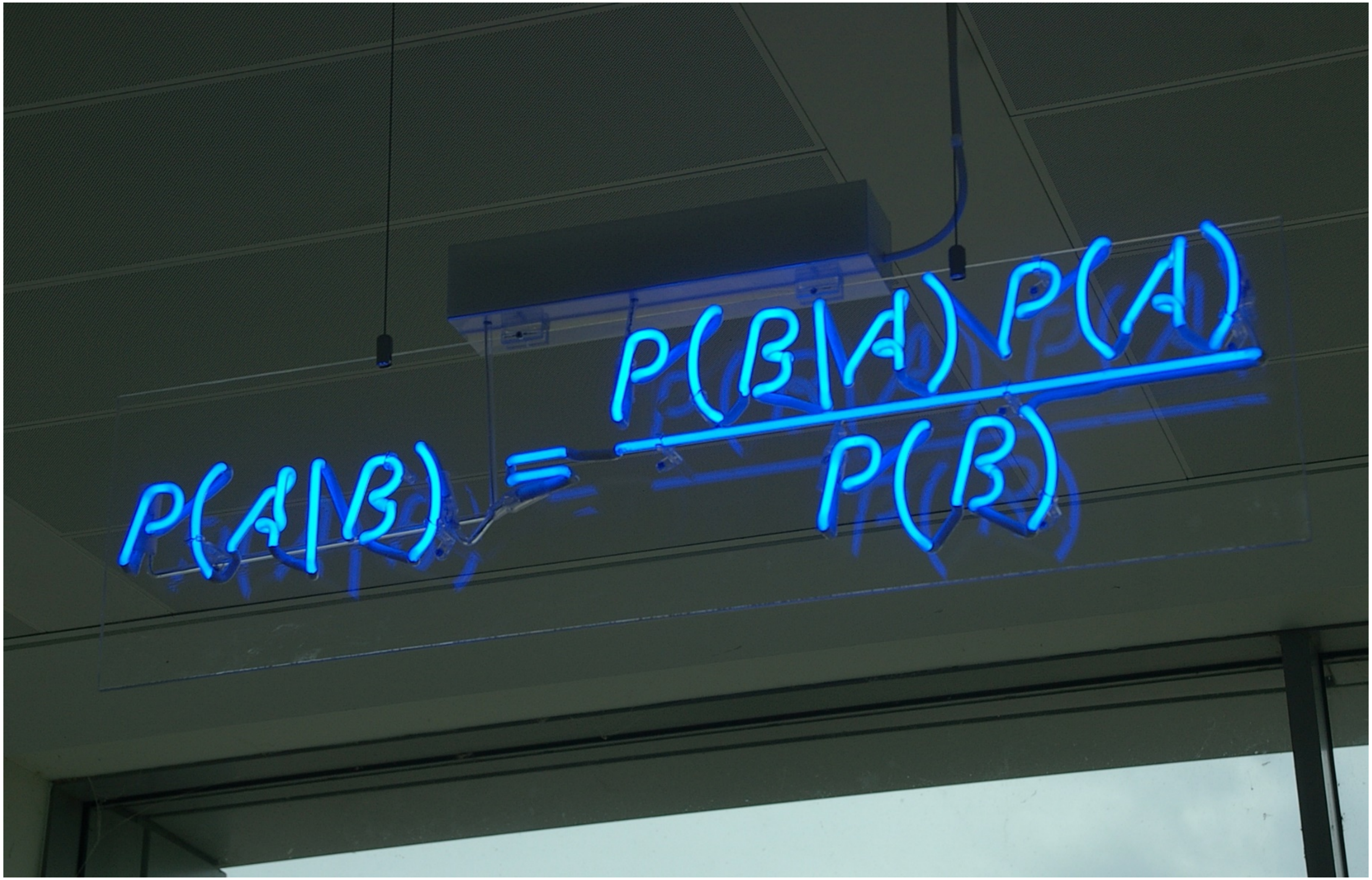
– Since P(no cancer | +) > P(cancer | +), we decide that the patient does not have cancer

– (Bayesian decision rule)

Exercise

- **Another Problem**:

- A person takes a lab test of nuclear radiation and the result is positive. The test returns a correct positive result in 99% of the cases in which the nuclear radiation is actually present, and a correct negative result in 95% of the cases in which the nuclear radiation is not present. Furthermore, 30% of the entire population are radioactively contaminated. Is this person contaminated ?


P(A|B) = P(B|A)P(A) / P(B)

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

Shortcoming of simple Bayesian decision

It have to let

X $\rightarrow$ ω_i

If the sample does not belong to any class, it will still be assigned to a class

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

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

- The loss function states how costly each action taken is

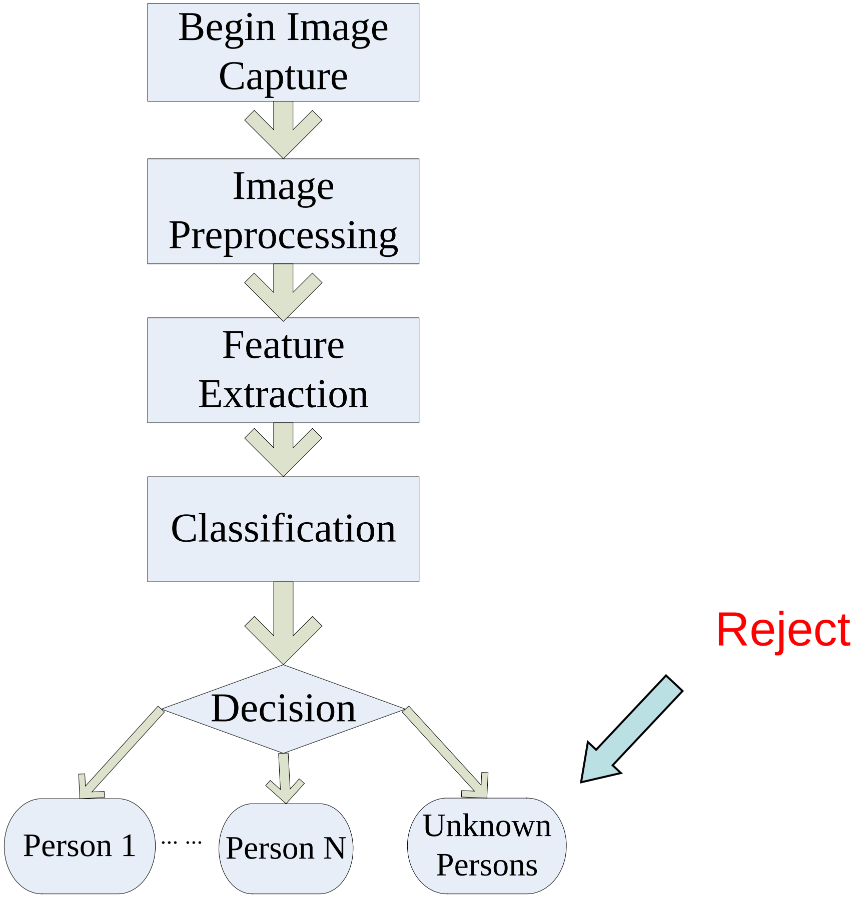
Examples of classification with rejection

Consequence of no rejection: if a person (the user) is not one of the registered users, he will be also erroneously recognized as a registered user!

Consequently this user will be erroneously allowed to pass the system!!

Personal identification & rejection

Face recognition flowchart



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

A simple case

- $\omega_1, \ \omega_2, \ldots, \ \omega_c$: *C classes*
- $\alpha_1, \ \alpha_2, \ldots, \ \alpha_c$: *C actions*

α_{c+1}: *do not assign the sample into any class--- reject*

Overall risk

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

Minimizing R $\Longleftrightarrow$ Minimizing $R(\alpha_i \mid x)$ for $i = 1,\ldots, a$

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

for i = 1,…,a

Fail to declare and error declaration

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 reasonable result that can be achieved!

- Two-category classification

$$\alpha_1 : deciding\ \omega_1$$

$$\alpha_2 : deciding\ \omega_2$$

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

λ_{ij} :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:

$$\text{if } R(\alpha_1 \mid x) < R(\alpha_2 \mid x)$$

action α_1: "decide ω_1" is taken

This results in the equivalent rule :

decide ω_1 if:

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

- and decide ω_2 otherwise

默多克"断腕求生"成效有待观察。

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

- is equal to

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

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"

Bayesian Decision Theory
Loss Function

- $\lambda(\alpha_i | \omega_j)$: cost incurred for taking action α_i (i.e., classification or rejection) when the state of nature is ω_j

- Example
  - *x:* financial characteristics of firms applying for a bank loan
  - ω_0 – company did not go bankrupt
    ω_1 – company failed
  - $P(\omega_i|\mathbf{x})$ – predicted probability of bankruptcy
  - Confusion matrix:

	Algorithm: ω_0	Algorithm: ω_1
Truth: ω_0	TN	FP
Truth: ω_1	FN	TP

  - FN are 10 times as costly as FP

$$\Rightarrow \lambda(\alpha_0 | \omega_1) = \lambda_{01} = 10 \times \lambda(\alpha_1 | \omega_0) = 10 \times \lambda_{10}$$

- Simplest λ_{ij}

- λ_{ij}=1, i<>j
- λ_{ij}=0

- Then minimum risk Bayesian decision will be equivalent to Minimum error Bayesian decision

Exercise

Select the optimal decision where:

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

$p(+ \mid \omega_1) \longrightarrow 0.9$

$p(+ \mid \omega_2) \longrightarrow 0.001$

$P(\omega_1) = 0.01$

$P(\omega_2) = 0.99$

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

$\lambda = \begin{bmatrix} 1 & 4 \\ 2 & 3 \end{bmatrix}$ means λ_{11}=**1**, λ_{12}=**4**, λ_{21}=**2**, λ_{22}=**3**

Example: earthquake forecast; typhoon forecast

Example

- 例：已知正常细胞先验概率为$P(\omega_1)=0.9$，异常为$P(\omega_2)=0.1$，
  从类条件概率密度分布曲线上查的$P(x/\omega_i)=0.2, P(x/\omega_i)=0.4$，
  $\lambda_{11}=0, \lambda_{12}=6, \lambda_{21}=1, \lambda_{22}=0$
  由上例中计算出的后验概率：$P(\omega_1/x)=0.818, P(\omega_2/x)=0.182$
  条件风险：$R(\alpha_1/x)=\sum_{j=1}^{2}\lambda_{1j}P(\omega_j/x)=\lambda_{12}P(\omega_2/x)=1.092$
  $R(\alpha_2/x)=\lambda_{21}P(\omega_1/x)=0.818$
  因为$R(\alpha_1/x)>R(\alpha_2/x)\ \therefore x\in$异常细胞，因决策$\omega_1$类风险大。
  因λ_{12}=6较大，决策损失起决定作用。