# **Bayes Decision Theory**

Sargur Srihari

CSE 555
Introduction to Pattern Recognition

# Reverend Thomas Bayes



1702-1761

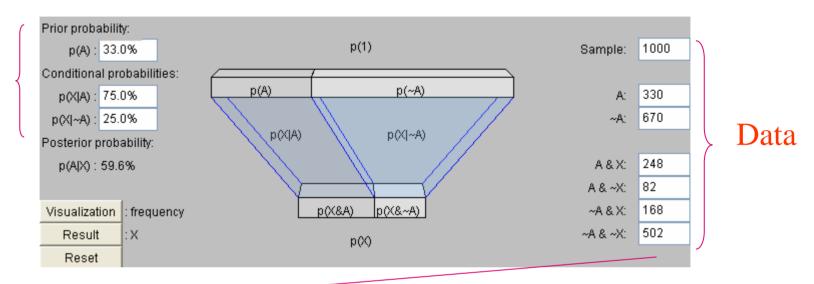
Bayes set out his theory of probability in Essay towards solving a problem in the doctrine of chances published in the Philosophical Transactions of the Royal Society of London in 1764. The paper was sent to the Royal Society by Richard Price, a friend of Bayes', who wrote:-I now send you an essay which I have found among the papers of our deceased friend Mr Bayes, and which, in my opinion, has great merit... In an introduction which he has writ to this Essay, he says, that his design at first in thinking on the subject of it was, to find out a method by which we might judge concerning the probability that an event has to happen, in given circumstances, upon supposition that we know nothing concerning it but that, under the same circumstances, it has happened a certain number of times, and failed a certain other number of times.

CSE 555: Srihari

## Bayes Rule

#### Two Classes(A, ~A), Single Binary-Valued Feature (X,~X)





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

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

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

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

$$= \frac{.2475}{.2475 + .1675} = \frac{.2475}{.415} = 0.596_{2}$$

# **Bayes Decision Theory**

- Fundamental statistical approach to statistical pattern classification
- Quantifies trade-offs between classification using probabilities and costs of decisions
- Assumes all relevant probabilities are known

## **Prior Probabilities**

### State of nature, prior

State of nature is a random variable

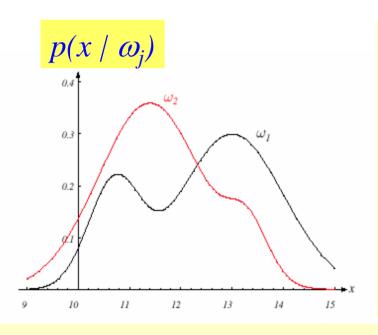
$$P(\omega_1) + P(\omega_2) = 1$$
 (exclusivity and exhaustivity)

Decision rule with only the prior information

Decide  $\omega_1$  if  $P(\omega_1) > P(\omega_2)$  otherwise decide  $\omega_2$ 

## Class-conditional Probabilities

$$p(x / \omega_1)$$
 and  $p(x / \omega_2)$ 



Pdfs show the probability of measuring a particular feature value given category  $\omega_i$ .

If *x* is a feature value, the two curves describe the difference in populations of two types of classes.

Density functions are normalized-- thus area under each curve is 1.0

Feature *x* 

CSE 555: Srihari

# Bayes formula to combine prior and class-conditional probabilities

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

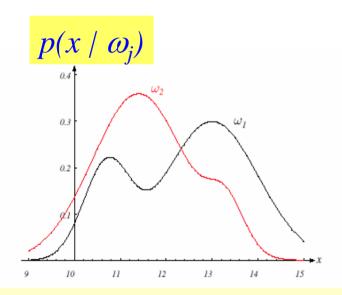
In the case of two categories

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

• Informally, Bayes rule says:
posterior = likelihood x prior / evidence

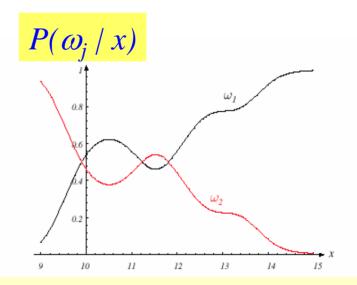
# Posterior probabilities

Class-conditional p.d.f.s



Feature *x* 

Posterior probabilities for the priors  $P(\omega_1) = 2/3$ ,  $P(\omega_2)=1/3$ For x = 14,  $P(\omega_1/x) = 0.08$ ,  $P(\omega_2/x) = 0.92$ 



Feature *x* 

# **Bayes Decision Rule**

x is an observation for which:

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

#### Therefore:

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

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

# Bayes Decision Rule minimizes probability of error

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

#### Therefore:

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

# Bayes 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

CSE 555: Srihari

## Loss Function

- 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

# Loss Function Definition

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

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

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

action  $\alpha_i$  when the state of nature is  $\omega_i$ 

## **Overall Risk**

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

#### **Conditional risk**

Minimizing R  $\longleftarrow$  Minimizing  $R(\alpha_i \mid x)$  for i = 1,..., a

Expected Loss with action i

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

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

R is minimum and R in this case is called the Risk

Bayes risk = best performance that can be achieved

# Two-category classification

 $\alpha_1$  : deciding  $\omega_1$ 

 $\alpha_2$ : deciding  $\omega_2$ 

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

loss incurred for deciding  $\omega_i$  when the true state of nature is  $\omega_j$ 

#### Conditional risk:

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

$$R(\alpha_2 \mid \mathbf{x}) = \lambda_{21} P(\omega_1 \mid \mathbf{x}) + \lambda_{22} P(\omega_2 \mid \mathbf{x})$$
CSE 555; Sribari

### Minimum Risk Decision Rule

Our rule is the following:

if 
$$R(\alpha_1 \mid x) < R(\alpha_2 \mid x)$$
  
action  $\alpha_1$ : "decide  $\omega_1$ " is taken

This results in the equivalent rule : decide  $\omega_1$  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  $\omega_2$  otherwise

# Likelihood ratio Decision Rule

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  $\alpha_1$  (decide  $\omega_1$ )
Otherwise take action  $\alpha_2$  (decide  $\omega_2$ )

### 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)

$$P(\omega_1) = 2/3$$
$$P(\omega_2) = 1/3$$

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