

# Pattern Recognition

# Contents

- Probability
- Bayes Decision Theory
- Minimum Error Rate Classification

# Last Class

- Types of Learning

- Features

- Shape / boundary
- Texture Features
- Color/Intensity

Pattern Reg. 4 modules

(1) data acq ✓

(2) F.E.

(3) classifier design

(4) eval<sup>n</sup> & perf



Ex

→ chaincodes -



0 9 7 3 ... 6 6 7 .  
numerical values

$$(x^2 + y^2 = r^2)$$

- Normalized
- Starting point
- Rotation.

- We consider a Feature Vector for object description

$$\begin{bmatrix} \text{Temp} \\ \text{Hum} \\ \text{ppt} \end{bmatrix} \rightarrow \left\{ \begin{bmatrix} 30 \\ 25 \\ 10 \end{bmatrix} \begin{bmatrix} 15 \\ 15 \\ 5 \end{bmatrix} \dots \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} \right\} = \underline{\underline{3 \times n}}$$

# Decision Boundary (Linear)

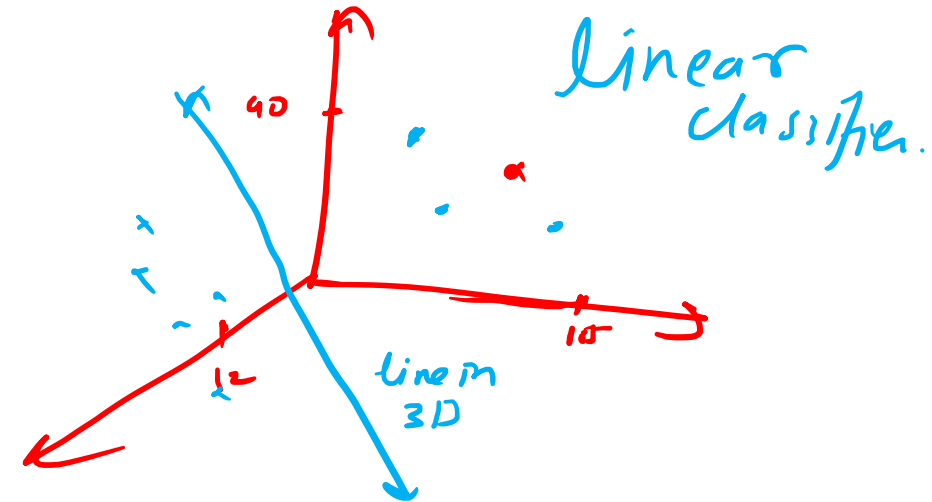
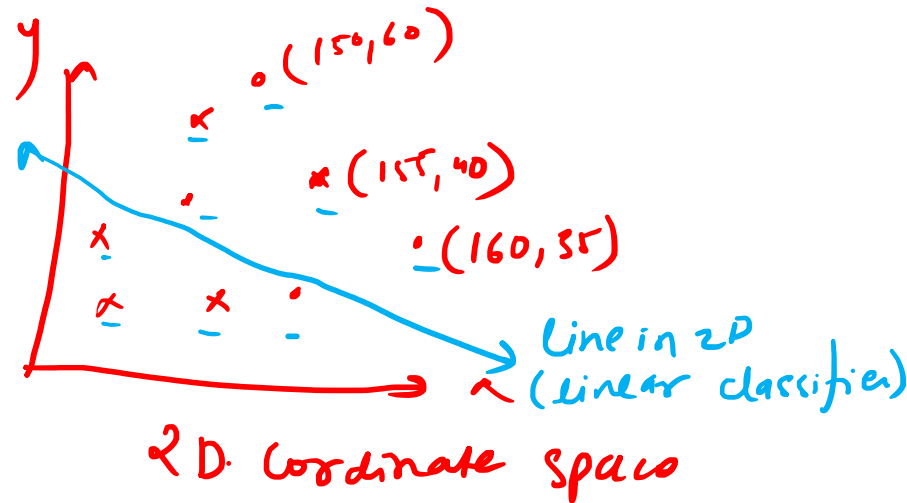
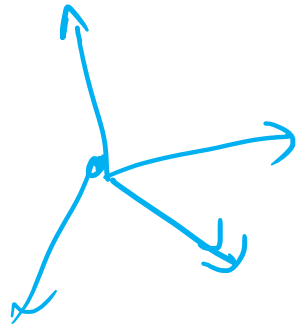
- FVs maps to points in feature space

$$FV_2 = \begin{bmatrix} 155 \\ 40 \end{bmatrix} \in \mathbb{R}^n$$

2 dim

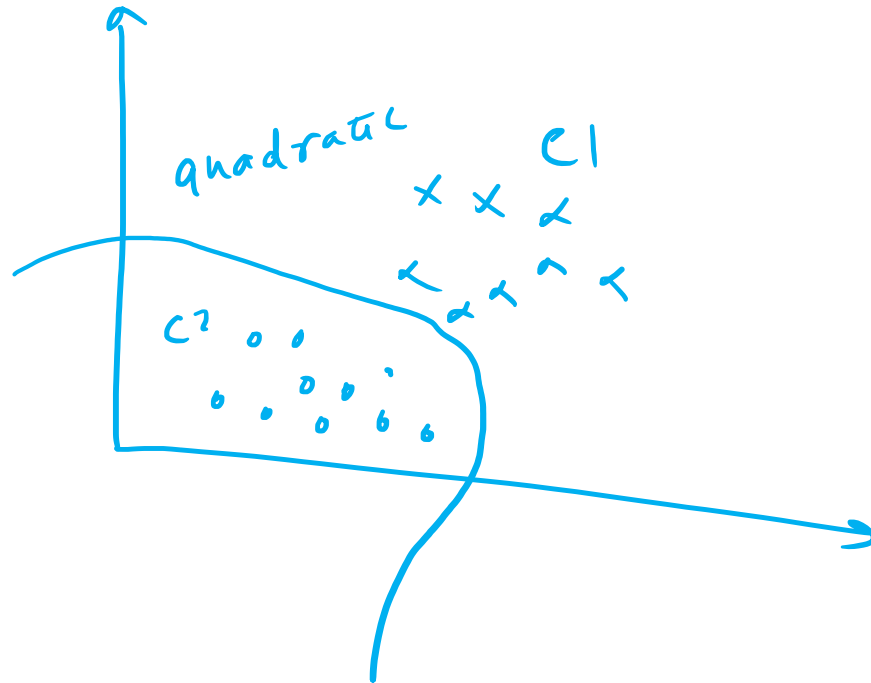
$$FV = \begin{bmatrix} 155 \\ 40 \\ 12 \end{bmatrix} \in \mathbb{R}^n$$

3D



# Decision Boundary (quadratic)

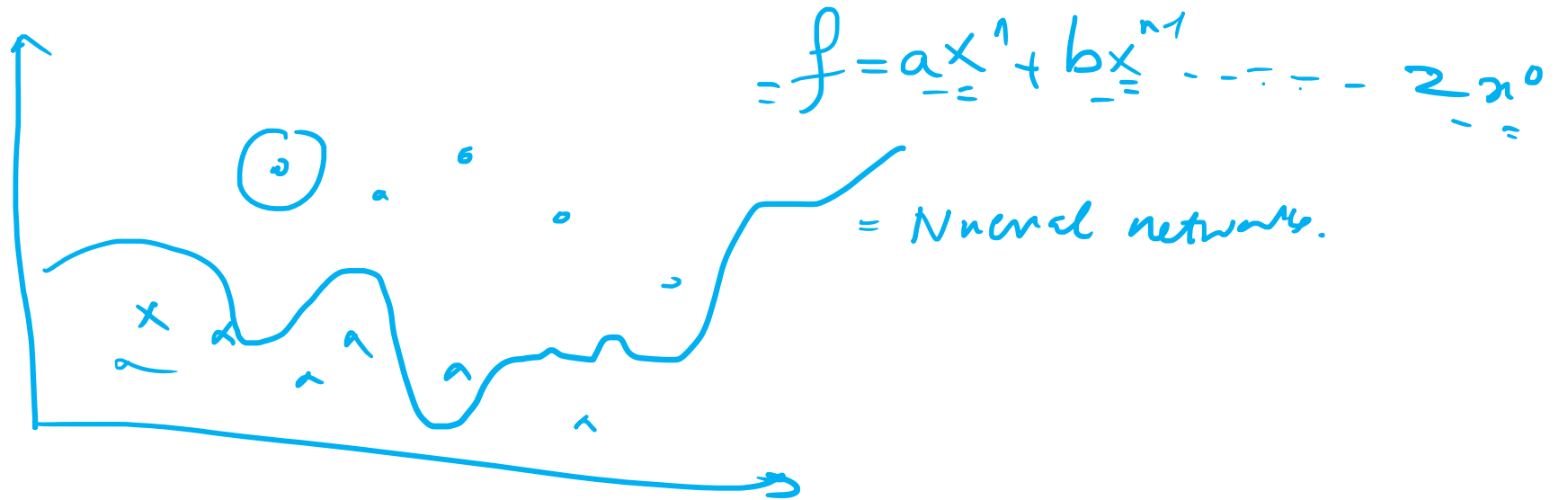
- FVs maps to points in feature space



# Decision Boundary (complicated)

- FVs maps to points in feature space

- Features
- feature vector
- feature space
- Decision boundary



# Bayesian decision theory —Duda et al.

- It makes the assumption that the decision problem is posed in probabilistic terms, and that all of the relevant probability values are known.

# Notations and probability

- Lets say we have two kinds (classes) of fishes-
  - Sea bass ( $\omega_1$ ) and salmon ( $\omega_2$ )
- We assume there is some prior probability (or simply prior)
- $P(\omega_1)$  -> next fish is a sea bass, or
- $P(\omega_2)$  -> next fish is a salmon.
- Suppose only two kinds , then  $P(\omega_1) + P(\omega_2) = 1$
- These prior probabilities reflect our prior knowledge of how likely we are to get a sea bass or salmon before the fish actually appears!



# Notations and probability

- And you are observing them coming on a conveyer belt:
- Only prior probabilities are known:
- i.e.,  $P(\omega_1)$ , and  $P(\omega_2)$



- What is the type of fish that will appear next without being allowed to see it??
  - If  $P(\omega_1) > P(\omega_2) \rightarrow \omega_1$ , otherwise
  - If  $P(\omega_1) < P(\omega_2) \rightarrow \omega_2$

# Notations and probability

- Is this a good way to predict?
- This rule makes sense if we are to judge just one fish, but if we are to judge many fish, using this rule repeatedly may seem a bit strange..

Class	Observation	Prior Probability
Sea bass	10	$P(\omega_1) = 10 \%$
Salmon	90	$P(\omega_2) = 90 \%$

- $P(\omega_2)$  is very large as compared to  $P(\omega_1)$ ....
- Therefore predicting  $P(\omega_2)$  will be right most of the times

# Notations and probability

- Is this a good way to predict?
- This rule makes sense if we are to judge just one fish, but if we are to judge many fish, using this rule repeatedly may seem a bit strange..

Class	Observation	Prior Probability
Sea bass	47	$P(\omega_1) = 47\%$
Salmon	53	$P(\omega_2) = 53\%$

- $P(\omega_2)$  is only bit large than  $P(\omega_1)$ ....
- Therefore predicting  $P(\omega_2)$  will be not be right most of the times, high chances of error!!!

# Notations and probability

- We usually do not have only this little information
- Lets say with each fish is associated a features with is weight
- For example, heavy / light
- we consider x to be a continuous random variable whose distribution depends on the state of nature (state represents class)

# Notations and probability

- So the probability for observing **heavy**, in this case, is approximately 42%. =  $20/47$ .

	Evidence	No. of Heavy	No. of Light	Total
$w_1$	Sea bass	20	27	47
$w_2$	Salmon			

- In the same way, I can also compute the probability for **light**, which is approximately 68%. =  $27/47$

# Notations and probability

(w<sub>2</sub>)

- So the probability for observing **heavy**, in this case, is approximately ..... 5/53


	Evidence	No. of Heavy	No. of Light	Total
w <sub>1</sub>	Sea bass	20	27	47 ✓
w <sub>2</sub>	Salmon	<u>5</u>	<u>48</u>	<u>53</u> ✓

(w<sub>2</sub>)

- So the probability for observing **light**, in this case, is approximately ..... 48/53.

# Notations and probability

- Let's say that the inspector is blind and can't differentiate the classes by seeing.
- So this is essentially the problem that we are facing. We can't access the class, but there is some true class that is somehow hidden.



Evidence	No. of Heavy	No. of Light	Total
<i>w1</i> Sea bass	20	27	47
<i>w2</i> Salmon	8	45	53

# Notations and probability

## *Joint Probability*

probability that you observe heavy and seabass at the same time

$$P(x = \text{heavy}, \omega_1) = 40\%$$

Evidence	No. of Heavy	No. of Light	Total
Sea bass	40	27	67
Salmon	3	30	33
	43	57	100



# Notations and probability

- We can of course analyze our fishes here.
- Here you see that **salmon** is the one that is biased towards **light weight**,
- and you can compute the probability for **heavy** given the class is **salmon**

Evidence	No. of Heavy	No. of Light	Total
Sea bass	40	27	67
Salmon	3	30	33
	43	57	100

# Notations and probability

- I start exploring the bucket of salmon ,
  - I will only have a 6% chance of producing heavy,
  - and I will have a 94% chance of producing light.

Evidence	No. of Heavy	No. of Light	Total
Sea bass	40	27	67
Salmon	3	30	33
	43	57	100

- Now, this is essentially the information that operator don't have.
- He get the observation heavy and light because the operator is blind.

# Notations and probability

Conditional Probability:

- $P(\omega_2=\text{salmon} \mid x=\text{heavy}) =$
- $P(\omega_2=\text{salmon} \mid x=\text{light}) =$

Evidence	No. of Heavy	No. of Light	Total
Sea bass	40	27	67
Salmon	3	30	33
	43	57	100

- $P(\omega_1=\text{seabass} \mid x=\text{heavy}) =$
- $P(\omega_1=\text{seabass} \mid x=\text{light}) =$

# Notations and probability

- And now I'm interested in figuring out whether it's the fish is salmon or seabass by looking at the feature

Evidence	No. of Heavy	No. of Light	Total
Sea bass	40	27	67
Salmon	2	30	33
	42	57	100

# Notations and probability

$$P(w_1) = 67/100$$

$$P(w_2) = 33/100$$

$$P(X=H|w_1) = 40/67$$

$$P(X=L|w_1) = 27/67$$

- If I get a light fish, then

- Chances of its being seabass is  $(27/57) = 47\%$

- Chance of being salmon is

$$P(w_1|X=H) \text{ and } P(w_2|X=H)$$

for given  $X \rightarrow$   
 $w_1?$  or  $w_2?$

Evidence	No. of Heavy	No. of Light	Total
→ Sea bass	40	27	67
Salmon	2	30	33
	42	57	100

# Notations and probability

- If I get a heavy fish, then
- Chances of its being seabass is  $(40/42) = 95.2\%$
- Chance of being salmon is  $(2/42) = 4.8\%$

Evidence	No. of Heavy	No. of Light	Total
Sea bass	40	27	67
Salmon	2	30	33
	42	57	100

- So observing heavy, in this case, is very strong evidence, that seabass is the underlying class
- And you can already see this is a very typical Pattern Classification problem

# Notations and probability

- Prior Probability
- Joint Probability
- Conditional Probability

$\epsilon$

# Notations and probability

- So we want to have some information about the the class from the observed evidence.
- And you already see that it's hard.
- But if we have distributions like this one, then you can get very interesting and very good evidence from our experiments.



# Bayes Rule

# Bayes Rule

$$P(x, \omega) = P(\omega) \cdot P(x|\omega)$$

$$= P(x) \cdot P(\omega|x)$$

# Bayes Rule

- We've seen that the joint probability density function here with  $x$  and  $\omega$  can be decomposed into the prior.
- So the probability of having certain fish times the class conditional probability density function.
- Obviously, the same joint pdf can be produced by using the probability of the evidence times the posterior probability.
- So we can express the same joint probability density function using those two decompositions.

# Bayes Rule

- Let prior probabilities  $P(\omega_j)$  and the conditional densities  $p(x | \omega_j)$  known.
- Weight of a fish and discover that its value is  $x$ .
- How does this measurement influence our attitude concerning the
- true state of nature — that is, the category of the fish?

# Bayes Rule

Rearranging these leads us to the answer to our question, which is called Bayes' formula:

# Bayes Rule

Rearranging these leads us to the answer to our question, which is called Bayes' formula:

$$P(\omega_j|x) = \frac{p(x|\omega_j)P(\omega_j)}{p(x)},$$

$$\text{where } p(x) = \sum_{j=1}^2 p(x|\omega_j)P(\omega_j).$$

Bayes' formula can be expressed informally in English by saying that

$$\textit{posterior} = \frac{\textit{likelihood} \times \textit{prior}}{\textit{evidence}}.$$

# Notations and probability

$$P(\omega_j|x) = \frac{p(x|\omega_j)P(\omega_j)}{p(x)},$$

- Bayes' formula shows that by observing the value of  $x$  we can convert the **prior probability  $P(\omega_j)$**  to the a **posteriori probability (or posterior) probability  $P(\omega_j | x)$**
- **By knowing  $p(x|\omega_j)$  the likelihood of  $\omega_j$  with respect to  $x$**

Notice that it is the product of the likelihood and the prior probability that is most important in determining the posterior probability;

the evidence factor,  $p(x)$ , can be viewed as merely a scale factor that guarantees that the posterior probabilities sum to one, as all good probabilities must.

# Bayes Rule

The variation of  $P(x|\omega_j)$  with  $x$  is illustrated in Fig. for the case  $P(\omega_1) = 2/3$  and  $P(\omega_2) = 1/3$ .

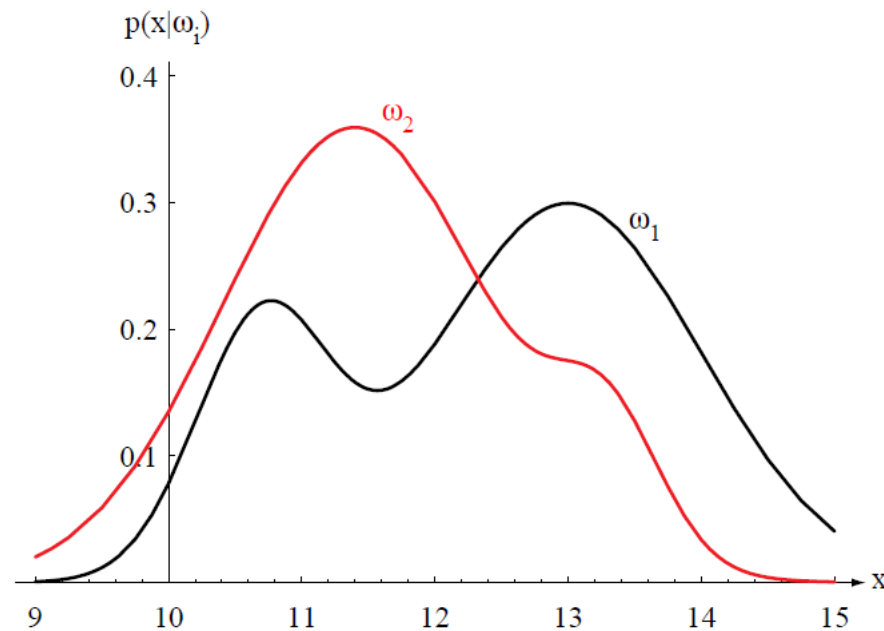
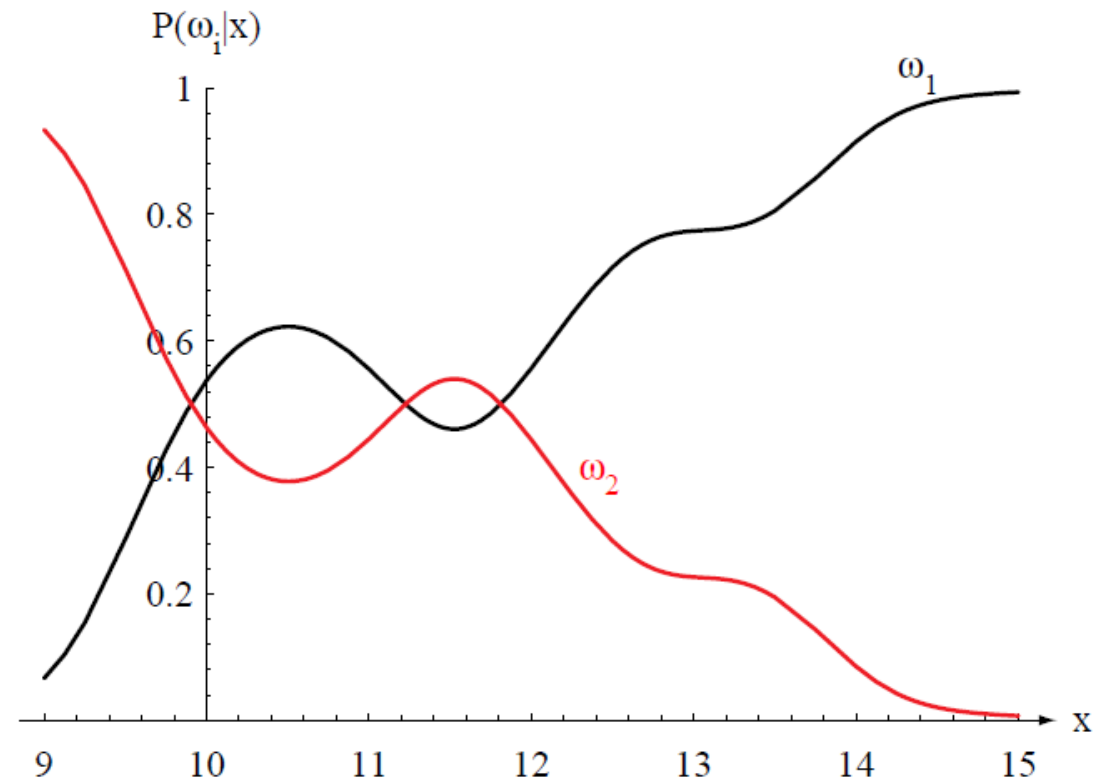


Figure 2.1: Hypothetical class-conditional probability density functions show the probability density of measuring a particular feature value  $x$  given the pattern is in category  $\omega_i$ . If  $x$  represents the weight of a fish, the two curves might describe the difference in length of populations of two types of fish. Density functions are normalized, and thus the area under each curve is 1.0.



# Bayes Rule



**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. Thus in this case, given that a pattern is measured to have feature value  $x = 14$ , the probability it is in category  $\omega_2$  is roughly 0.08, and that it is in  $\omega_1$  is 0.92. At every  $x$ , the posteriors sum to 1.0.

# Bayes Rule

- Given two classes
- If we have an observation x for which we estimate  $P(\omega_1 | x)$  ,  $P(\omega_2 | x)$ ,

# Bayes Rule

$$P(\omega_1|x) > P(\omega_2|x)$$

↳ classify to  $\omega_1$

- If we have an observation  $x$  for which  $P(\omega_1|x)$  is greater than  $P(\omega_2|x)$ , we would naturally be inclined to decide that the true state of nature is  $\omega_1$ .
- Similarly, if  $P(\omega_2|x)$  is greater than  $P(\omega_1|x)$ , we would be inclined to choose  $\omega_2$ .
- Probability of error?

$$P(\omega_2|x) > P(\omega_1|x) \rightarrow \omega_2$$

# Bayes Rule

$P(\omega_1|x) > P(\omega_2|x) \longrightarrow \omega_1$  (predicted)  
but the actual  $\rightarrow \omega_2$   
 $P(\text{error}|x) \rightarrow P(\omega_2|x)$

- If we have an observation  $x$  for which  $P(\omega_1|x)$  is greater than  $P(\omega_2|x)$ , we would naturally be inclined to decide that the true state of nature is  $\omega_1$ .
- Similarly, if  $P(\omega_2|x)$  is greater than  $P(\omega_1|x)$ , we would be inclined to choose  $\omega_2$ .
- Probability of error?

$$P(\text{error}|x) = \begin{cases} P(\omega_1|x) & \text{if we decide } \underline{\omega_2} \\ P(\omega_2|x) & \text{if we decide } \underline{\omega_1}. \end{cases}$$

# Bayes Rule

- Similarly, if  $P(\omega_2|x)$  is greater than  $P(\omega_1|x)$ , we would be inclined to choose  $\omega_2$ .
- Probability of error?

$$\underline{P(error|x)} = \begin{cases} P(\omega_1|x) & \text{if we decide } \omega_2 \\ P(\omega_2|x) & \text{if we decide } \omega_1. \end{cases}$$

- average probability of error is given by

# Bayes Rule

$$x = 11/L$$

$$P(x, \omega) = P(x) P(\omega | x)$$

- Similarly, if  $P(\omega_2 | x)$  is greater than  $P(\omega_1 | x)$ , we would be inclined to choose  $\omega_2$ .
- Probability of error?

$$P(\text{error} | x) = \begin{cases} P(\omega_1 | x) & \text{if we decide } \omega_2 \\ P(\omega_2 | x) & \text{if we decide } \omega_1. \end{cases}$$



- average probability of error is given by *over all possible values of x*  
 $x = -\infty$  to  $\infty$

$$\text{bd } \underline{P(\text{error})} = \int_{-\infty}^{\infty} \underbrace{P(\text{error}, x)}_{\text{joint}} dx = \int_{-\infty}^{\infty} \underbrace{P(\text{error} | x)}_{\checkmark} \underbrace{p(x)}_{\checkmark} dx$$

- and if for every x we insure that  $P(\text{error} | x)$  is as small as possible, then the integral must be as small as possible

Bayes Rule  $p(\omega_1|x) > p(\omega_2|x) \rightarrow \omega_1$

- Thus we have justified the following Bayes' decision rule for minimizing the probability of error:

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

- and under this rule error becomes

$$\underline{P(\text{error}|x)} = \min [P(\underline{\omega_1}|x), P(\underline{\omega_2}|x)].$$

- This form of the decision rule emphasizes the role of the posterior probabilities.

# Bayes Rule

$$P(\omega_1|x) = \frac{P(\omega_1) P(x|\omega_1)}{P(x)}$$

- We know how to express  $P(\omega_j|x)$

$$P(\omega_2|x) = \frac{P(\omega_2) P(x|\omega_2)}{P(x)}$$

- We know  $P(\text{error}|x) = \min [P(\omega_1|x), P(\omega_2|x)]$ .

$$P(\omega_1|x) > P(\omega_2|x)$$

- note that the evidence,  $p(x)$ , becomes unimportant as far as making a decision is concerned. Its just a scale factor

$$P(\omega_1) P(x|\omega_1) > P(\omega_2) P(x|\omega_2)$$

- Also  $P(\omega_1|x) + P(\omega_2|x) = 1$ .

- We obtain the following completely equivalent decision rule



# Bayes Rule

- We know how to express  $P(\omega_j | x)$
- We know  $P(error|x) = \min [P(\omega_1|x), P(\omega_2|x)]$ .
- note that the evidence,  $p(x)$ , becomes unimportant as far as making a decision is concerned. Its just a scale factor
- Also  $P(\omega_1/x) + P(\omega_2/x) = 1$ .
- We obtain the following completely equivalent decision rule:

Decide  $\omega_1$  if  $p(x|\omega_1)P(\omega_1)$  >  $p(x|\omega_2)P(\omega_2)$ ; otherwise decide  $\omega_2$ .

# Bayes rule

$$\chi = 10 f_{\omega} = H/L.$$

- Some additional insight can be obtained by considering a few special cases.
- If for some x we have  $p(x|\omega_1) = p(x|\omega_2)$ , then that particular observation gives us no information about the state of nature.
- On the other hand, if  $P(\omega_1) = P(\omega_2)$ , then the states of nature are equally probable;
- In general, both of these factors are important in making a decision, and the Bayes decision rule combines them to achieve the minimum probability of error.

# Bayesian Decision Theory – Continuous Features

We shall now formalize the ideas just considered, and generalize them in four ways:

- ① by allowing the use of more than one feature  $X = [x_1, x_2, \dots, x_d]$
- ② by allowing more than two states of nature  $\omega_1, \omega_2, \dots, \omega_c$
- ✓ by allowing actions other than merely deciding the state of nature  $\alpha_1, \alpha_2, \dots, \alpha_n$
- ✓ by introducing a loss function more general than the probability of error.  $J(\alpha_i | \omega_j)$

# Bayesian Decision Theory

2 classes  $\rightarrow \omega_1 \& \omega_2$

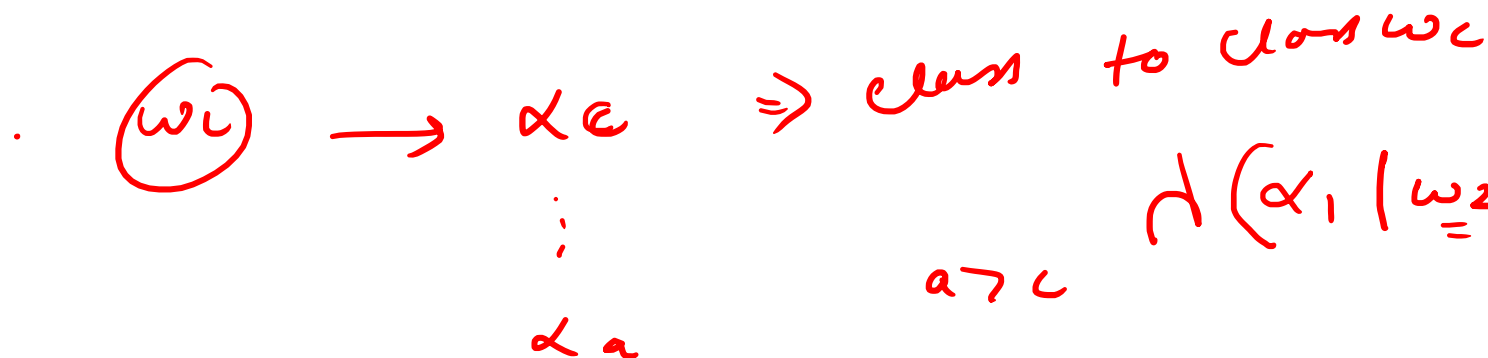
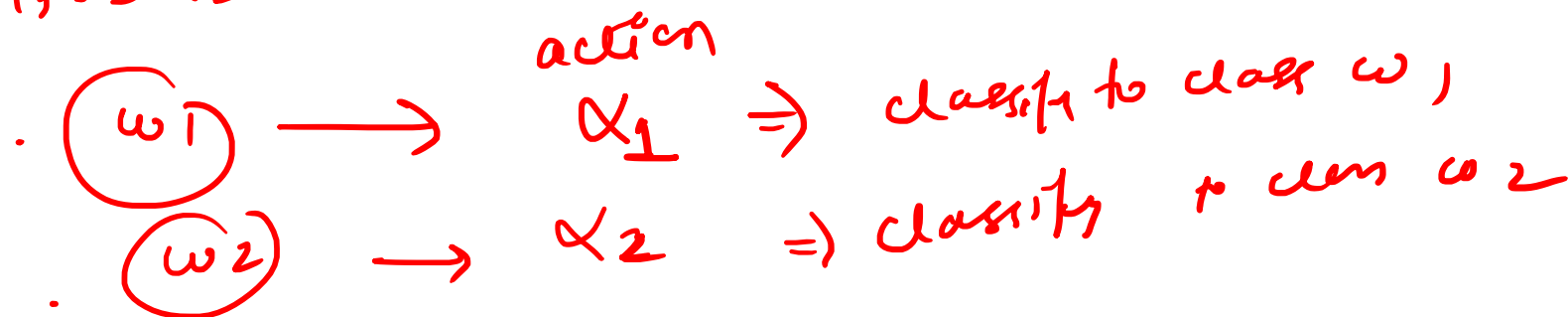
$\omega_1, \omega_2, \dots, \omega_c$

① • Let  $\omega_1, \dots, \omega_c$  be the finite set of  $c$  states of nature (“categories”)

classes

② • Let  $\alpha_1, \dots, \alpha_a$  be the finite set of a possible actions.

$\alpha_1, \dots, \alpha_2, \dots, \alpha_a$



$$P(\alpha_1 | \omega_2)$$

given true class is  $\omega_2$  but error class  $\omega_1$

# Bayesian Decision Theory

- Let  $\omega_1, \dots, \omega_c$  be the finite set of  $c$  states of nature (“categories”)
- Let  $\alpha_1, \dots, \alpha_a$  be the finite set of  $a$  possible actions.
- The loss function  $\lambda(\alpha_i | \omega_j)$  describes the loss incurred for taking action  $\alpha_i$  when the state of nature is  $\omega_j$ .

$$\lambda(\alpha_i | \omega_j)$$

# Bayesian Decision Theory

- ✓ Let  $\omega_1, \dots, \omega_c$  be the finite set of  $c$  states of nature (“categories”)
- ✓ Let  $\alpha_1, \dots, \alpha_a$  be the finite set of  $a$  possible actions.
- The loss function  $\lambda(\alpha_i | \omega_j)$  describes the loss incurred for taking action  $\alpha_i$  when the state of nature is  $\omega_j$ .
- Let the feature vector  $x$  be a  $d$ -component vector-valued random variable,
- and let  $p(x | \omega_j)$  be the state conditional probability density function for  $x$  — the probability density function for  $x$  conditioned on  $\omega_j$  being the true state of nature.

# Bayesian Decision Theory

- As before,  $P(\omega_j)$  describes the prior probability that nature is in state  $\omega_j$ .
- Then the posterior probability  $P(\omega_j | x)$  can be computed from  $p(x | \omega_j)$  by Bayes' formula:

$$P(\omega_j | x)$$

# Bayesian Decision Theory

- As before,  $P(\omega_j)$  describes the prior probability that nature is in state  $\omega_j$ .
- Then the posterior probability  $P(\omega_j | x)$  can be computed from  $p(x | \omega_j)$  by Bayes' formula:

$$P(\omega_j | x) = \frac{p(x | \omega_j) P(\omega_j)}{p(x)},$$

where the evidence is now

$$p(x) = \sum_{j=1}^c p(x | \omega_j) P(\omega_j).$$



# Bayesian Decision Theory

# Bayesian Decision Theory

# Bayesian Decision Theory

# Bayesian Decision Theory

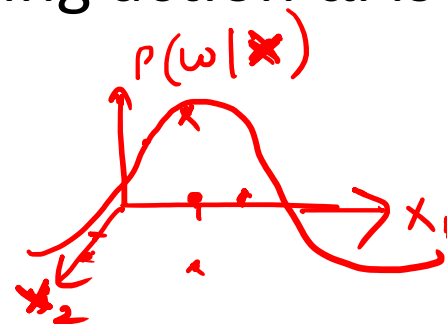
- Suppose that we observe a particular  $x$  and that we contemplate taking action  $\alpha_i$ .
- loss • If the true state of nature is  $\omega_j$ , by definition we will incur the loss  $\lambda(\alpha_i | \omega_j)$ .
- Since  $P(\omega_j | x)$  is the probability that the true state of nature is  $\omega_j$ , the expected loss associated with taking action  $\alpha_i$  is merely

$$R(\alpha_i | x)$$

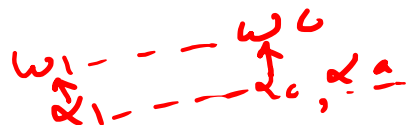
# Bayesian Decision Theory

	$x_1$	$x_2$
$\omega_1$	H/L	T/S
$\omega_2$	H/L	T/S

- Suppose that we observe a particular  $x$  and that we contemplate taking action  $\alpha_i$ .
- If the true state of nature is  $\omega_j$ , by definition we will incur the loss  $\lambda(\alpha_i | \omega_j)$ .
- Since  $P(\omega_j | x)$  is the probability that the true state of nature is  $\omega_j$ , the expected loss associated with taking action  $\alpha_i$  is merely



# Bayesian Decision Theory



error  $x = \int \dots$

- Suppose that we observe a particular  $x$  and that we contemplate taking action  $\alpha_i$ .

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

*Risk*

- If the true state of nature is  $\omega_j$ , by definition we will incur the loss  $\lambda(\alpha_i | \omega_j)$ .

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

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

- Since  $P(\omega_j | x)$  is the probability that the true state of nature is  $\omega_j$ , the expected loss associated with taking action  $\alpha_i$  is merely

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

# Bayesian Decision Theory

## RISK

- In decision-theoretic terminology, an expected loss is called a risk, and  $R(\alpha_i | x)$  is called the conditional risk.
- Whenever we encounter a particular observation  $x$ , we can minimize our expected loss by selecting the action that minimizes the conditional risk.
- We shall now show that this Bayes decision procedure actually provides the optimal performance on an overall risk

# Bayesian Decision Theory

- Our problem : find a decision rule against  $P(\omega_j)$  that minimizes the overall risk.
- A general decision rule is a function  $\alpha(x)$  that tells us which rule action to take for every possible observation.
- Every  $x$  the decision function  $\alpha(x)$  assumes one of the  $a$  values  $\alpha_1, \dots, \alpha_a$ .
- The overall risk  $R$  is the expected loss associated with a given decision rule.
- Since  $R(\alpha_i | x)$  is the conditional risk associated with action  $\alpha_i$ , and since the decision rule specifies the action, the overall risk is given by



# Bayesian Decision Theory

- Our problem : find a decision rule against  $P(\omega_j)$  that minimizes the overall risk.
- A general decision rule is a function  $\alpha(x)$  that tells us which rule action to take for every possible observation.
- Every  $x$  the decision function  $\alpha(x)$  assumes one of the  $a$  values  $\alpha_1, \dots, \alpha_a$ .
- The overall risk  $R$  is the expected loss associated with a given decision rule.
- Since  $R(\alpha_i | x)$  is the conditional risk associated with action  $\alpha_i$ , and since the decision rule specifies the action, the overall risk is given by

$$\underline{R} = \int \underline{R(\alpha(x) | x)} \underline{p(x)} dx,$$

# Bayesian Decision Theory

- Bayes decision rule: To minimize the overall risk, compute the conditional risk

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

→  $R(\alpha_i | \mathbf{x})$  =  $\sum R(\alpha_i | \omega_j) P(\omega_j | \mathbf{x})$

$$\boxed{\alpha_i \rightarrow \omega_i}$$

- for  $i = 1, \dots, a$  and select the action  $\alpha_i$  for which  $R(\alpha_i | \mathbf{x})$  is minimum.
- The resulting minimum overall risk is called the Bayes risk.

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

# Bayesian Decision Theory

## Two-Category Classification Problem

$$\begin{aligned}\alpha_1 &\rightarrow \omega_1 \\ \alpha_2 &\rightarrow \omega_2\end{aligned}$$

- Here action  $\alpha_1$  corresponds to deciding that the true state of nature is  $\omega_1$ , and action  $\alpha_2$  corresponds to deciding that it is  $\omega_2$ .

$$d_{ij} = \lambda(\alpha_i / \omega_j)$$

- For notational simplicity, let  $\lambda_{ij}$  =  $\lambda(\alpha_i / \omega_j)$  be the loss incurred for deciding  $\omega_i$  when the true state of nature is  $\omega_j$ .
- conditional risk given is

# Bayesian Decision Theory

## Two-Category Classification Problem

- Here action  $\alpha_1$  corresponds to deciding that the true state of nature is  $\omega_1$ , and action  $\alpha_2$  corresponds to deciding that it is  $\omega_2$ .
- For notational simplicity, let  $\lambda_{ij} = \lambda(\alpha_i/\omega_j)$  be the loss incurred for deciding  $\omega_i$  when the true state of nature is  $\omega_j$ .

- conditional risk given is 
$$\underline{R(\alpha_i|\mathbf{x})} = \sum_{j=1}^c \lambda(\alpha_i|\omega_j)P(\omega_j|\mathbf{x})$$

$$R(\alpha_1|\mathbf{x}) =$$

$$R(\alpha_2|\mathbf{x}) =$$

# Bayesian Decision Theory

## Two-Category Classification Problem

- Here action  $\alpha_1$  corresponds to deciding that the true state of nature is  $\omega_1$ , and action  $\alpha_2$  corresponds to deciding that it is  $\omega_2$ .
- For notational simplicity, let  $\lambda_{ij} = \lambda(\alpha_i | \omega_j)$  be the loss incurred for deciding  $\omega_i$  when the true state of nature is  $\omega_j$ .

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

- conditional risk given is

$$\begin{aligned}
 -R(\alpha_1 | \mathbf{X}) &= \lambda_{11} P(\omega_1 | \mathbf{X}) + \lambda_{12} P(\omega_2 | \mathbf{X}) \quad \text{and} \\
 R(\alpha_2 | \mathbf{X}) &= \lambda_{21} P(\omega_1 | \mathbf{X}) + \lambda_{22} P(\omega_2 | \mathbf{X}).
 \end{aligned}$$

Handwritten annotations in red:
 

- For  $\lambda_{11}$ :  $(\alpha_1 | \omega_1)$  and  $d_{11}$
- For  $\lambda_{12}$ :  $(\alpha_1 | \omega_2)$
- For  $\lambda_{21}$ :  $(\alpha_2 | \omega_1)$  and  $d_{21}$
- For  $\lambda_{22}$ :  $(\alpha_2 | \omega_2)$

# Bayesian Decision Theory

- There are a variety of ways of expressing the minimum-risk decision rule, each having its own minor advantages.
- The fundamental rule is to decide  $\omega_1$  if  $R(\alpha_1 | x) < R(\alpha_2 | x)$ .
- In terms of the posterior probabilities, we decide  $\omega_1$  if

$$\checkmark \quad \underline{(\lambda_{21} - \lambda_{11})P(\omega_1 | x)} > \underline{(\lambda_{12} - \lambda_{22})P(\omega_2 | x)}.$$

$P(\omega_1) P(x | \omega_1)$

# Bayes Decision Theory

- By employing Bayes' formula, we can replace the posterior probabilities by the prior probabilities and the conditional densities.
- This results in the equivalent rule, to decide  $\omega_1$  if

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

# Bayesian Decision Theory

Another alternative, which follows at once under the reasonable assumption that  $\lambda_{21} > \lambda_{11}$ , is to decide  $\omega_1$  if

likelihood ratio

$$\left[ \frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)} > \frac{\lambda_{12} - \lambda_{22} \frac{P(\omega_2)}{P(\omega_1)}}{\lambda_{21} - \lambda_{11}} \right] \quad (17)$$

$p(\mathbf{x}|\omega_1) =$

- We can consider  $p(\mathbf{x}|\omega_j)$  a function of  $\omega_j$  (i.e., the likelihood function),
- then likelihood ratio is  $p(\mathbf{x}|\omega_1)/p(\mathbf{x}|\omega_2)$ .
- Thus the Bayes decision rule can be interpreted ratio as calling for deciding  $\omega_1$  if the likelihood ratio exceeds a threshold value that is independent of the observation  $\mathbf{x}$ .



# Bayes Decision Theory

- The likelihood ratio

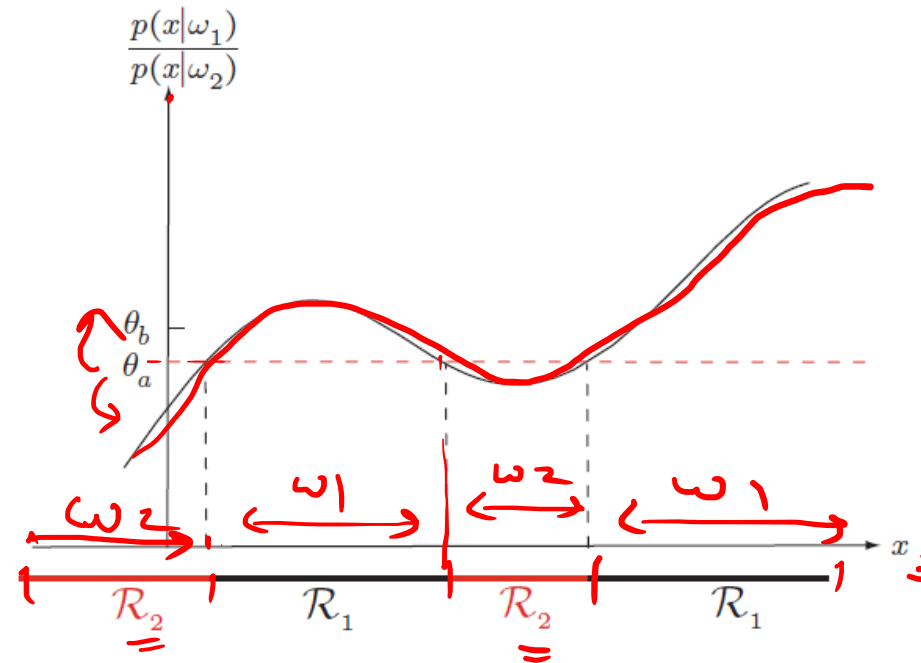


Figure 2.3: The likelihood ratio  $p(x|\omega_1)/p(x|\omega_2)$  for the distributions shown in Fig. 2.1. If we employ a zero-one or classification loss, our decision boundaries are determined by the threshold  $\theta_a$ . If our loss function penalizes miscategorizing  $\omega_2$  as  $\omega_1$  patterns more than the converse, (i.e.,  $\lambda_{12} > \lambda_{21}$ ), we get the larger threshold  $\theta_b$ , and hence  $\mathcal{R}_1$  becomes smaller.

# Minimum-Error-Rate Classification

# Minimum-Error-Rate Classification

- Given 'c' classes  $\omega_1, \omega_2, \dots, \omega_c$
- Action  $\alpha_i$  is usually interpreted as the decision that the true state of nature is  $\omega_i$ . (predicted class).  $\alpha_i \rightarrow \omega_i$
- If action  $\alpha_i$  is taken and the true state of nature is  $\omega_j$ , then
  - the decision is correct if  $i = j$ , and in error if  $i \neq j$ .
- Then so-called symmetrical or zero-one loss function can be defined as:

# Minimum-Error-Rate Classification

- Given 'c' classes
- Action  $\alpha_i$  is usually interpreted as the decision that the true state of nature is  $\omega_i$ . (predicted class).
- If action  $\alpha_i$  is taken and the true state of nature is  $\omega_j$ , then
  - the decision is correct **if  $i = j$ , and in error if  $i \neq j$ .**
- Then so-called symmetrical or zero-one loss function can be defined as:

$$\underline{R(\alpha_i|\mathbf{x})} = \sum_{j=1}^c \underline{\lambda(\alpha_i|\omega_j)} \underline{P(\omega_j|\mathbf{x})} \quad \checkmark$$

# Minimum-Error-Rate Classification

- Given 'c' classes
- Action  $\alpha_i$  is usually interpreted as the decision that the true state of nature is  $\omega_i$ . (predicted class).
- If action  $\alpha_i$  is taken and the true state of nature is  $\omega_j$ , then
  - the decision is correct if  $i = j$ , and in error if  $i \neq j$ .
- Then so-called symmetrical or zero-one loss function can be defined as:

$$\begin{aligned}
 R(\alpha_i | \mathbf{x}) &= \sum_{j=1}^c \lambda(\alpha_i | \omega_j) P(\omega_j | \mathbf{x}) \\
 &= \sum_{j \neq i} P(\omega_j | \mathbf{x}) \\
 &= 1 - P(\omega_i | \mathbf{x})
 \end{aligned}$$

$$\begin{aligned}
 \lambda(\alpha_i | \omega_i) &\rightarrow 0 \\
 \lambda(\alpha_i | \omega_j) &\rightarrow 1
 \end{aligned}$$

# Minimum-Error-Rate Classification

- Given 'c' classes
- Action  $\alpha_i$  is usually interpreted as the decision that the true state of nature is  $\omega_i$ . (predicted class).
- If action  $\alpha_i$  is taken and the true state of nature is  $\omega_j$ , then
  - the decision is correct if  $i = j$ , and in error if  $i \neq j$ .
- Then so-called symmetrical or zero-one loss function can be defined as:

$$\begin{aligned} \underline{R(\alpha_i | \mathbf{x})} &= \sum_{j=1}^c \lambda(\alpha_i | \omega_j) P(\omega_j | \mathbf{x}) \\ &= \sum_{j \neq i} P(\omega_j | \mathbf{x}) = 1 - P(\omega_i | \mathbf{x}) \\ &= 1 - \boxed{P(\omega_i | \mathbf{x})} \end{aligned}$$

# Minimum-Error-Rate Classification

$$P(\omega_i | x) \max_i$$

Bayes decision rule : minimize risk calls for selecting the action that minimizes the conditional risk.

- Thus, to minimize the average probability of error, we should select the  $i$  that maximizes the posterior probability  $P(\omega_i | x)$ .

In other words, for minimum error rate

$$\text{Decide } \omega_i \text{ if } P(\omega_i | \underline{x}) > P(\omega_j | \underline{x}) \quad \text{for all } j \neq i.$$

$P(\omega_j | x)$