# Pattern Recognition

# Contents

- Probability
- Bayes Decision Theory
- Minimum Error Rate Classification

# Last Class

- Types of Learning

- Features
  - Shape / boundary
  - Texture Features
  - Color/Intensity

- We consider a Feature Vector for object description

# Decision Boundary (Linear)

- FVs maps to points in feature space

# Decision Boundary (quadratic)

- FVs maps to points in feature space

# Decision Boundary (complicated)

- FVs maps to points in feature space

# Bayesian decision theory

- 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(ω1) -> next fish is a sea bass, or
- P(ω2) -> next fish is a salmon.

- Suppose only two kinds , then P(ω1) + P(ω2) =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 $P(\omega 1) > P(\omega 2)$ - > $\omega 1$, otherwise
  - If $P(\omega 1) < P(\omega 2)$ - > $\omega 2$

# 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 | $P(\omega1)$= 10 % |
| Salmon | 90 | $P(\omega2)$= 90 % |

- $P(\omega2)$ is very large as compared to $P(\omega1)$….
- Therefore predicting $P(\omega2)$ 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%.

| Evidence | No. of Heavy | No. of Light | Total |
|---|---|---|---|
| Sea bass | 20 | 27 | 47 |
| Salmon | | | |

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

# Notations and probability

- So the probability for observing ***heavy***, in this case, is approximately ……..

| Evidence | No. of Heavy | No. of Light | Total |
|---|---|---|---|
| Sea bass | 20 | 27 | 47 |
| Salmon | 5 | 48 | 53 |

- So the probability for observing ***light***, in this case, is approximately ……..

# 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 |
|---|---|---|---|
| Sea bass | 20 | 27 | 47 |
| Salmon | 8 | 45 | 53 |

# Notations and probability

***Joint Probability***

probability that you observe heavy and seabass at the same time

**P(x= heavy, ω1 ) = 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(ω2=salmon| x=heavy ) =**
- **P(ω2=salmon|x=light ) =**

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

- **P(ω1=seabass| x=heavy ) =**
- **P(ω1=seabass| x=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

- If I get a light fish, then
- Chances of its being seabass is (27/57) =47%
- Chance of being salmon is

| 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 ( /42) = %
- Chance of being salmon is

| 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

# 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 **ω** 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

- We note first that the (joint) probability density of finding a pattern that is in category $\omega_j$ and has feature value $x$ can be written two ways:

$$p(\omega_j, x) =$$

# 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:

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

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

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

# 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(ωj)** to the a **posteriori probability (or posterior) probability P(ωj |x)**

- **By knowing p(x|ωj) the likelihood of ωj 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 $P(\omega 2) = 1/3$.

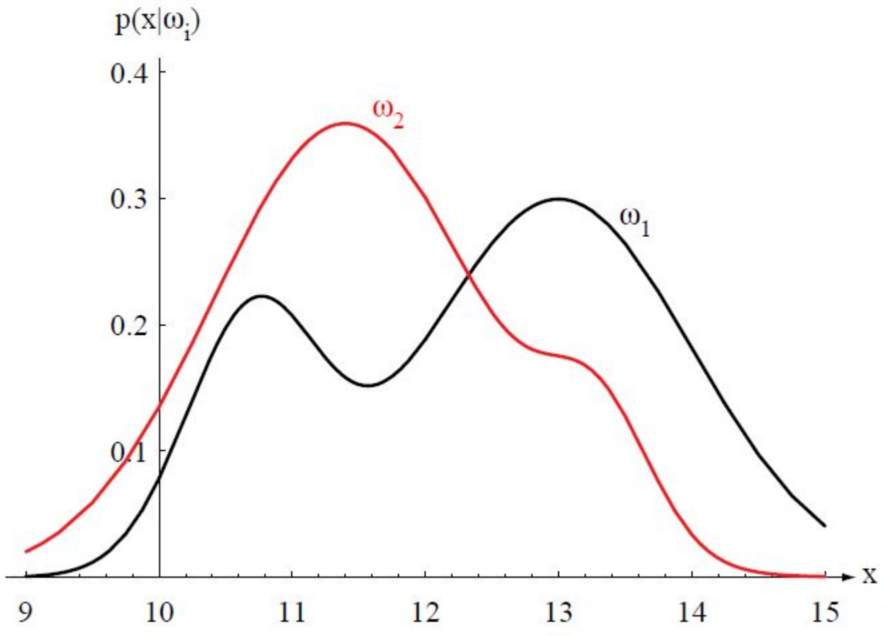
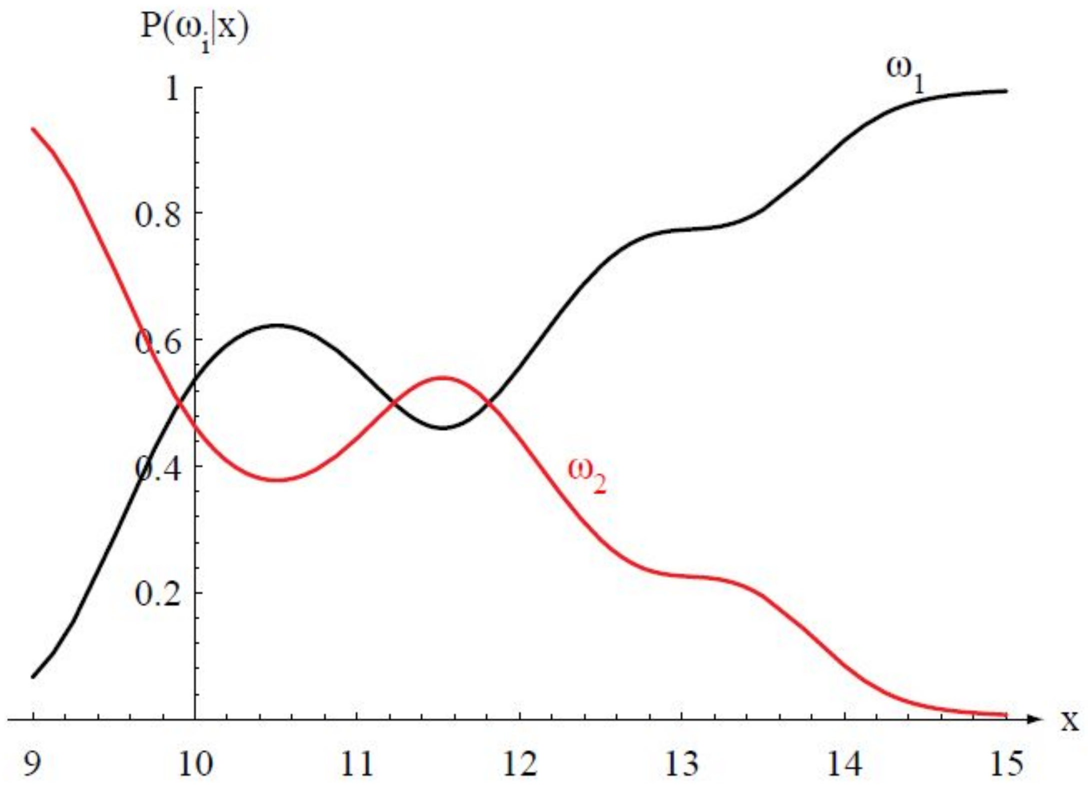



Figure 2.1: Hypothetical class-conditional probability density functions show the probability density of measuring a particular feature value $x$ given the pattern is in category $\omega_i$. If $x$ represents the weight of a fish, the two curves might describe the difference in length of populations 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

- 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?

# Bayes Rule

- If we have an observation x for which P(ω1|x) is greater than P(ω2|x), we would naturally be inclined to decide that the true state of nature is ω1.
- Similarly, if P(ω2|x) is greater than P(ω1|x), we would be inclined to choose ω2.
- Probability of error?

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

# Bayes Rule

- Similarly, if P(ω2|x) is greater than P(ω1|x), we would be inclined to choose ω2.
- Probabilitv of error?

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

- average probability of error is given by

# Bayes Rule

- Similarly, if P(ω2|x) is greater than P(ω1|x), we would be inclined to choose ω2.
- Probabilitv of error?

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

- average probability of error is given by

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

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

# Bayes Rule

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

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

- and under this rule error becomes

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

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

# Bayes Rule

- We know how to express P(ωj|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

# Bayes Rule

- We know how to express P(ωj|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:

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

# Bayes rule

- 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:

- by allowing the use of more than one feature
- by allowing more than two states of nature
- by allowing actions other than merely deciding the state of nature
- by introducing a loss function more general than the probability of error.

# Bayesian Decision Theory

- Let ω1, ..., ωc be the finite set of c states of nature (“categories”)
- Let  α1, ..., αa be the finite set of a possible actions.

# Bayesian Decision Theory

- Let $\omega_1, \ldots, \omega_c$ be the finite set of c states of nature ("categories")
- Let $\alpha_1, \ldots, \alpha_a$ 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 $\omega_j$ .

# Bayesian Decision Theory

- Let $\omega_1, \ldots, \omega_c$ be the finite set of c states of nature (“categories”)
- Let $\alpha_1, \ldots, \alpha_a$ 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 $P(\omega j \mid x)$ can be computed from $p(x|\omega j)$ by Bayes' formula:

# 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:

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

where the evidence is now

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

# 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 $\alpha_i$.

- 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

$$R(\alpha_i|\mathrm{x})$$

# Bayesian Decision Theory

- 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

- 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

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

# 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 $\alpha 1, ..., \alpha a$.
- 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 $\alpha 1, ..., \alpha a$.
- 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

$$R = \int R(\alpha(\mathrm{x})|\mathrm{x})p(\mathrm{x})\ d\mathrm{x},$$

# Bayesian Decision Theory

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

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

- for i = 1,...,a and select the action αi for which R(αi|x) is minimum.
- The resulting minimum overall risk is called the Bayes risk.

# 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

# 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

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

$$\begin{aligned} R(\alpha_1|\mathbf{x}) &= \\ R(\alpha_2|\mathbf{x}) &= \end{aligned}$$

# 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 $\omega j$ .

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

- conditional risk given is

$$\begin{aligned} R(\alpha_1|\mathrm{x}) &= \lambda_{11}P(\omega_1|\mathrm{x}) + \lambda_{12}P(\omega_2|\mathrm{x}) \qquad \text{and} \\ R(\alpha_2|\mathrm{x}) &= \lambda_{21}P(\omega_1|\mathrm{x}) + \lambda_{22}P(\omega_2|\mathrm{x}). \end{aligned}$$

# 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

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

# 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 ω1 if

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

# Bayesian Decision Theory

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

$$\frac{p(\mathrm{x}|\omega_1)}{p(\mathrm{x}|\omega_2)} > \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \frac{P(\omega_2)}{P(\omega_1)}. \tag{17}$$

- 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 **x**.

# Bayes Decision Theory

- The likelihood ratio

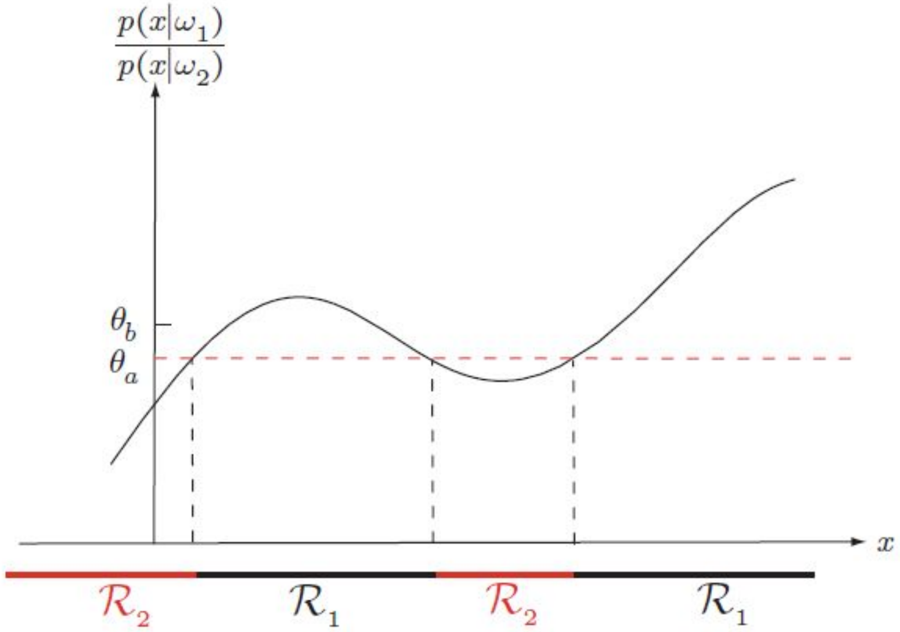



Figure 2.3: The likelihood ratio $p(x|\omega_1)/p(x|\omega_2)$ for the distributions shown in Fig. 2.1. If we employ a zero-one or classification loss, our decision boundaries are determined by the threshold $\theta_a$. If our loss function penalizes miscategorizing $\omega_2$ as $\omega_1$ patterns 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
- 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:

# 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 != j.**
- Then so-called symmetrical or zero-one loss function can be defined as:

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

# Minimum-Error-Rate Classification

- Given ‘c’ classes
- Action αi is usually interpreted as the decision that the true state of nature is ωi. (predicted class).
- If action αi is taken and the true state of nature is ωj , then
  - the decision is correct **if i = j, and** in **error if i != j.**
- Then so-called symmetrical or zero-one loss function can be defined as:

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

# Minimum-Error-Rate Classification

- Given ‘c’ classes
- Action αi is usually interpreted as the decision that the true state of nature is ωi. (predicted class).
- If action αi is taken and the true state of nature is ωj , then
  - the decision is correct **if i = j, and** in **error if i != j.**
- Then so-called symmetrical or zero-one loss function can be defined as:

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

# Minimum-Error-Rate Classification

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(ωi|x).

In other words, for minimum error rate

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