

Data Mining (Mineria de Dades)

Lluís A. Belanche
belanche@lsi.upc.edu

Soft Computing Research Group

Dept. de Llenguatges i Sistemes Informàtics (Software department)

Universitat Politècnica de Catalunya

2011-2012

Data Mining (Part II)

Contents

1. Association Rules
2. Bayesian decision theory
3. Bayesian classifiers: LDA/QDA
4. Bayesian classifiers: Naïve Bayes
5. Bayesian classifiers: Nearest Neighbours
6. Neural Networks: Multilayer Perceptrons
7. Support Vector Machines
8. Ensemble Learning

Bayesian decision theory (I)

Introduction: Bayes' formula

Thomas Bayes: XVIII-century priest. His works on the celebrated formula were found upon his death

Discrete random variables Let A a discrete r.v. with pmf P_A . We use the shorthand notation $P(a)$ to mean $P_A(A = a)$. Similarly we write $P(b|a)$ to mean $P_{B|A}(B = b|A = a)$, etc, where

$$P(b|a) = \frac{P(b, a)}{P(a)}, \quad P(a) > 0$$

(**prior**, **joint** and **conditional** probabilities)

Bayesian decision theory (I)

Introduction: Bayes' formula

Let $\{a_1, \dots, a_n\}, \{b_1, \dots, b_m\}$ the sets of possible values that A, B can take.
Then, for any $a \in \{a_1, \dots, a_n\}$:

$$P(a) = \sum_{i=1}^m P(a, b_i) = \sum_{i=1}^m P(a|b_i)P(b_i)$$

Since $P(a, b) = P(b, a)$, it follows that, for any a_k, b_j :

$$P(b_j|a_k) = \frac{P(a_k|b_j)P(b_j)}{\sum_{i=1}^m P(a_k|b_i)P(b_i)}, \text{ with } \sum_{j=1}^m P(b_j|a_k) = 1$$

(**posterior** probabilities)

Bayesian decision theory (I)

Introduction: Bayes' formula

Example 1 *We have two urns: in urn U_1 there are A_1 apples and O_1 oranges; in urn U_2 there are A_2 apples and O_2 oranges. We choose one urn at random and blindly pick up a fruit, that turns out to be an apple. What is the probability that the chosen urn was U_1 ?*

$$P(U_1|apple) = \frac{P(apple|U_1)P(U_1)}{P(apple|U_1)P(U_1) + P(apple|U_2)P(U_2)} =$$

$$\frac{\frac{A_1}{A_1+O_1} \cdot \frac{1}{2}}{\frac{A_1}{A_1+O_1} \cdot \frac{1}{2} + \frac{A_2}{A_2+O_2} \cdot \frac{1}{2}} = \frac{A_1(A_2 + O_2)}{A_1(A_2 + O_2) + A_2(A_1 + O_1)}$$

Bayesian decision theory (I)

Introduction: Bayes' formula

Exercise 1 *In a certain city 85 % of the taxis are painted blue and the rest are painted green. There is an accident in which one taxi is involved. An eyewitness says that the taxi was green. This eyewitness is 80 % reliable. What is the probability that the taxi was indeed green?*

Bayesian decision theory (I)

Introduction: Bayes' formula

Continuous random variables Let X, Y two continuous r.v. with pdf p_X, p_Y and joint density p_{XY} . We use the shorthand notation $p(x)$ to mean $p_X(X = x)$, etc.

$$p(x) = \int_{\mathbb{R}} p(x, y) dy; \quad p(y) = \int_{\mathbb{R}} p(x, y) dx$$

Therefore:

$$p(y|x) = \frac{p(x|y)p(y)}{\int_{\mathbb{R}} p(x|y)p(y) dy}, \text{ with } \int_{\mathbb{R}} p(y|x) dy = 1$$

Bayesian decision theory (I)

Introduction: Bayes' formula

Mixed random variables Suppose X is a continuous r.v. and Y is a discrete r.v. with values in $\{y_1, \dots, y_m\}$.

In this case, $p(\cdot|y_i)$ is a continuous r.v. and $P(\cdot|x)$ is a discrete r.v. Moreover,

$$P(y_j|x) = \frac{p(x|y_j)P(y_j)}{\sum_{i=1}^m p(x|y_i)P(y_i)}, \text{ with } \sum_{j=1}^m P(y_j|x) = 1$$

Bayesian decision theory (I)

Decision rules

We are interested in determining the class or category of objects of nature according to Ω , a discrete r.v. with values $\{\omega_1, \omega_2\}$ that represent the two possible classes.

The prior probabilities are $P(\omega_1), P(\omega_2)$. How should we classify objects?

rule 1: if $P(\omega_1) > P(\omega_2)$ then class of object is ω_1 else is ω_2

This rule classifies all objects into the same class; therefore it makes errors!

$$P_{rule1}(error) = \min\{P(\omega_1), P(\omega_2)\}$$

useful only if $P(\omega_1) \ll P(\omega_2)$ or $P(\omega_1) \gg P(\omega_2)$.

Bayesian decision theory (I)

Decision rules

Suppose now that X is a discrete r.v. that measures a *feature* of the objects with values $\{x_1, \dots, x_n\}$. In this setting, $P(\omega_i|x) = \frac{P(x|\omega_i)P(\omega_i)}{p(x)}$ is the *posterior* probability that an object with measured feature x belongs to class $P(\omega_i), i = 1, 2$.

We have $P(x) = P(x|\omega_1)P(\omega_1) + P(x|\omega_2)P(\omega_2)$, the *unconditional* distribution of x .

Upon observing an *evidence* x , the Bayes formula converts *prior* class probabilities $P(\omega_i)$ into posterior probabilities $P(\omega_i|x)$. How should we classify objects now?

rule 2: if $P(\omega_1|x) > P(\omega_2|x)$ then class of object is ω_1 else class is ω_2 .

$$P_{rule2}(error) = \sum_{i=1}^n \min\{P(\omega_1|x_i), P(\omega_2|x_i)\}P(x_i)$$

(this rule is known as the **Bayes rule** or the **Bayes classifier**)

Bayesian decision theory (I)

Decision rules

Lemma. For all $a, b, c, d \in \mathbb{R}$, $\min(a, b) + \min(c, d) \leq \min(a + c, b + d)$

Proposition 1 $P_{rule2}(error) \leq P_{rule1}(error)$

Proof.

$$\begin{aligned} & \sum_{i=1}^n \min\{P(\omega_1|x_i), P(\omega_2|x_i)\}P(x_i) \\ &= \sum_{i=1}^n \min\{P(x_i)P(\omega_1|x_i), P(x_i)P(\omega_2|x_i)\} && \text{(Bayes formula)} \\ &= \sum_{i=1}^n \min\{P(x_i|\omega_1)P(\omega_1), P(x_i|\omega_2)P(\omega_2)\} && \text{(iterated lemma)} \\ &\leq \min\{\sum_{i=1}^n P(x_i|\omega_1)P(\omega_1), \sum_{i=1}^n P(x_i|\omega_2)P(\omega_2)\} \\ &= \min\{P(\omega_1) \sum_{i=1}^n P(x_i|\omega_1), P(\omega_2) \sum_{i=1}^n P(x_i|\omega_2)\} \\ &= \min\{P(\omega_1), P(\omega_2)\} \end{aligned}$$

The probabilities of error are equal only if $P(x_i|\omega_1) = P(x_i|\omega_2)$ for all i .

Bayesian decision theory (I)

Example 2 We have a conveyor belt carrying two classes of pills, suitable for two different illnesses. These pills go in two colors {yellow, white}.

$$P(\omega_1) = \frac{1}{3}, P(\omega_2) = \frac{2}{3}$$

$$P(\text{yellow}|\omega_1) = \frac{1}{5}, P(\text{white}|\omega_1) = \frac{4}{5};$$

$$P(\text{yellow}|\omega_2) = \frac{2}{3}, P(\text{white}|\omega_2) = \frac{1}{3}$$

Therefore,

$$P(\text{yellow}) = P(\omega_1)P(\text{yellow}|\omega_1) + P(\omega_2)P(\text{yellow}|\omega_2) = \frac{1}{3} \cdot \frac{1}{5} + \frac{2}{3} \cdot \frac{2}{3} = \frac{23}{45}$$

$$P(\text{white}) = P(\omega_1)P(\text{white}|\omega_1) + P(\omega_2)P(\text{white}|\omega_2) = \frac{1}{3} \cdot \frac{4}{5} + \frac{2}{3} \cdot \frac{1}{3} = \frac{22}{45}$$

$$P(\omega_1|\text{yellow}) = \frac{P(\text{yellow}|\omega_1)P(\omega_1)}{P(\text{yellow})} = \left(\frac{1}{5} \cdot \frac{1}{3}\right) / \frac{23}{45} = \frac{3}{23}; \quad P(\omega_2|\text{yellow}) = 1 - P(\omega_1|\text{yellow}) = \frac{20}{23}$$

$$P(\omega_1|\text{white}) = \frac{P(\text{white}|\omega_1)P(\omega_1)}{P(\text{white})} = \left(\frac{4}{5} \cdot \frac{1}{3}\right) / \frac{22}{45} = \frac{6}{11}; \quad P(\omega_2|\text{white}) = 1 - P(\omega_1|\text{white}) = \frac{5}{11}$$

$$\text{Therefore } P(\text{error}) = \frac{23}{45} \cdot \frac{3}{23} + \frac{22}{45} \cdot \frac{5}{11} = \frac{13}{45} < \frac{1}{3} = \min\{\frac{1}{3}, \frac{2}{3}\}$$

Bayesian decision theory (I)

Continuous variables

The next step is to consider a r.v. X with pdf $p(x)$ that measures a *continuous* feature of the objects. Let $\mathcal{P}$ be the support of p , i.e. $\mathcal{P} = \{x \in \mathbb{R} | p(x) > 0\}$.

In this setting, $p(x|\omega_i), i = 1, 2$ are the conditional densities of x for every class.

Proposition 2 $P_{rule2}(error) \leq P_{rule1}(error)$

Proof.

$$\begin{aligned} & \int_{\mathcal{P}} \min\{P(\omega_1|x), P(\omega_2|x)\} p(x) dx \\ &= \int_{\mathcal{P}} \min\{p(x)P(\omega_1|x), p(x)P(\omega_2|x)\} dx && \text{(Bayes formula)} \\ &= \int_{\mathcal{P}} \min\{p(x|\omega_1)P(\omega_1), p(x|\omega_2)P(\omega_2)\} dx && \text{(standard result for integrals)} \\ &\leq \min\{\int_{\mathcal{P}} p(x|\omega_1)P(\omega_1) dx, \int_{\mathcal{P}} p(x|\omega_2)P(\omega_2) dx\} \\ &= \min\{P(\omega_1) \int_{\mathcal{P}} p(x|\omega_1) dx, P(\omega_2) \int_{\mathcal{P}} p(x|\omega_2) dx\} \\ &= \min\{P(\omega_1), P(\omega_2)\} \end{aligned}$$

The probabilities of error are equal only if $p(\cdot|\omega_1) = p(\cdot|\omega_2)$.

Bayesian decision theory (I)

Example 3 We have a conveyor belt carrying two classes of pills, suitable for two different illnesses. This time the pills go in two colors, shaded in $[0, 2]$, with probabilities:

$$P(\omega_1) = \frac{1}{3}, P(\omega_2) = \frac{2}{3}, p(x|\omega_1) = \frac{2-x}{2}, p(x|\omega_2) = \frac{x}{2}.$$

Then

$$p(x) = P(\omega_1)p(x|\omega_1) + P(\omega_2)p(x|\omega_2) = \frac{1}{3} \cdot \frac{2-x}{2} + \frac{2}{3} \cdot \frac{x}{2} = \frac{2+x}{6}$$

$$P(\omega_1|x) = \frac{p(x|\omega_1)P(\omega_1)}{p(x)} = \left(\frac{2-x}{2} \cdot \frac{1}{3}\right) / \frac{2+x}{6} = \frac{2-x}{2+x}; \quad P(\omega_2|x) = 1 - P(\omega_1|x) = 1 - \frac{2-x}{2+x} = \frac{2x}{2+x}$$

$$P(\text{error}) = \int_0^{2/3} P(\omega_2|x)p(x) dx + \int_{2/3}^2 P(\omega_1|x)p(x) dx \quad \left(\frac{2}{3} \text{ is the solution of } \frac{2-x}{2+x} = \frac{2x}{2+x}\right)$$

$$= P(\omega_2) \int_0^{2/3} p(x|\omega_2) dx + P(\omega_1) \int_{2/3}^2 p(x|\omega_1) dx = \frac{2}{9} < \frac{1}{3} = \min\{\frac{1}{3}, \frac{2}{3}\}$$

Bayesian decision theory (I)

The Bayes classifier

Exercise 2 Consider again the setting of Example 1 (the 'color' discrete variable). Extend the possible values to $\{\text{yellow}, \text{white}, \text{red}\}$, with $P(\omega_1) = \frac{1}{3}$, $P(\omega_2) = \frac{2}{3}$.

Assume the conditional probabilities given by:

	yellow	white	red
ω_1	$\frac{1}{5}$	$\frac{3}{5}$	$\frac{1}{5}$
ω_2	$\frac{2}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

Construct the Bayes classifier for this case.

Bayesian decision theory (I)

The Bayes classifier

The Bayes classifier can be extended in two ways:

1. Consider a vector $X = (X_1, \dots, X_n)^t$ of continuous r.v. with pdf $p(\mathbf{x}) = p(x_1, \dots, x_n)$ that measures n continuous features of the objects.
2. Consider a finite number of classes Ω , a discrete r.v. with values $\omega_1, \dots, \omega_c$, that represent the possible classes.

Therefore we have conditional probabilities $p(\mathbf{x}|\omega_i), 1 \leq i \leq c$.

Bayesian decision theory (I)

The Bayes classifier

The Bayes classifier can also have a *rejection class* (illustrated here for two classes):

if $P(\omega_1|\mathbf{x}) > P(\omega_2|\mathbf{x})$ then class of object is ω_1

else if $P(\omega_1|\mathbf{x}) < P(\omega_2|\mathbf{x})$ then class of object is ω_2

else do not classify

Therefore for every feature vector $\mathbf{x}$ we take one of three possible *actions*.

Bayesian decision theory (I)

The Bayes classifier

Consider a finite set of actions $A = \{a_1, \dots, a_m\}$. For each $a_i \in A$, denote by $l(a_i|\omega_j)$ the loss for choosing a_i when x is known to be in ω_j .

(note this is a simplified setting in which the loss does not depend on x)

Example 4 Let $m = c + 1$ and let a_i stand for “classify x into class ω_i ” for $1 \leq i \leq c$; let a_{c+1} stand for “do not classify x ”. A possible set of losses is:

$$l(a_i|\omega_j) = 1; \quad 1 \leq i, j \leq c, i \neq j$$

$$l(a_i|\omega_i) = 0; \quad 1 \leq i \leq c$$

$$l(a_{c+1}|\omega_j) = \frac{1}{2}; \quad 1 \leq j \leq c$$

which suggests that a decision not to classify is less costly than a misclassification

Bayesian decision theory (I)

The notion of risk

For a given feature vector $\mathbf{x}$, define the *conditional risk* of an action as:

$$r(a_i|\mathbf{x}) = \sum_{j=1}^c l(a_i|\omega_j)P(\omega_j|\mathbf{x})$$

A *decision rule* is any function $a : \mathcal{P} \in \mathbb{R}^n \rightarrow A$ that assigns an action $a(\mathbf{x})$ to every $\mathbf{x}$ s.t. $p(\mathbf{x}) > 0$. Define the *total risk* of a decision rule as:

$$R(a) = \int_{\mathcal{P}} r(a(\mathbf{x})|\mathbf{x})p(\mathbf{x}) d\mathbf{x}$$

Bayesian decision theory (I)

The notion of risk

We are interested in the decision rule that minimizes the total risk. Consider the rule

$$\hat{a}(x) = \arg \min_{1 \leq j \leq m} r(a_j | x)$$

(you may recognize it as the Bayes rule!)

Given that this rule minimizes the argument of the integral for every possible x , it follows that the Bayes rule has the lowest possible risk.

The value of $R(\hat{a})$ is called the *Bayes risk*.

Bayesian decision theory (I)

The notion of risk

Example 5 Yet again the conveyor belt carrying two classes of pills that go in two shaded colors (yellow, white), i.e. a scalar feature $x \in [0, 2]$, with probabilities:

$$P(\omega_1) = \frac{2}{3}, P(\omega_2) = \frac{1}{3}, p(x|\omega_1) = \frac{2-x}{2}, p(x|\omega_2) = \frac{1}{2}.$$

and three possible actions: a_1 – classify as ω_1 , a_2 – classify as ω_2 , and a_3 – do not classify.

Let the loss matrix $l_{ij} \equiv l(a_i|\omega_j)$ be:

	ω_1	ω_2
a_1	0	1
a_2	1	0
a_3	$\frac{1}{4}$	$\frac{1}{4}$

Compute the optimal decision rule and its associated risk.

Bayesian decision theory (I)

The notion of risk

We have

$$p(x) = \frac{2}{3} \cdot \frac{2-x}{2} + \frac{1}{3} \cdot \frac{1}{2} = \frac{5-2x}{6}$$

$$P(\omega_1|x) = \frac{2}{3} \cdot \frac{2-x}{2} / \frac{5-2x}{6} = \frac{4-2x}{5-2x}; \quad P(\omega_2|x) = 1 - P(\omega_1|x) = 1 - \frac{4-2x}{5-2x} = \frac{1}{5-2x}$$

The conditional risks are:

$$r_1(x) \equiv r(a_1|x) = 0 \cdot P(\omega_1|x) + 1 \cdot P(\omega_2|x) = \frac{1}{5-2x}$$

$$r_2(x) \equiv r(a_2|x) = 1 \cdot P(\omega_1|x) + 0 \cdot P(\omega_2|x) = \frac{4-2x}{5-2x}$$

$$r_3(x) \equiv r(a_3|x) = \frac{1}{4} \cdot P(\omega_1|x) + \frac{1}{4} \cdot P(\omega_2|x) = \frac{1}{4}$$

Now the Bayes rule chooses for each x the action with minimum conditional risk.

Bayesian decision theory (I)

The notion of risk

Therefore:

$0 \leq x \leq \frac{1}{2} \Rightarrow \text{take action } a_1 \Rightarrow \text{choose } \omega_1$

$\frac{1}{2} \leq x \leq \frac{11}{16} \Rightarrow \text{take action } a_3 \Rightarrow \text{do not classify}$

$\frac{11}{16} \leq x \leq 2 \Rightarrow \text{take action } a_2 \Rightarrow \text{choose } \omega_2$

Questions:

- What is the probability that an object is not classified?
- What is the total risk?

Bayesian decision theory (I)

The notion of risk

The probability that an object is not classified is:

$$\int_{\frac{1}{2}}^{\frac{11}{16}} \frac{5 - 2x}{6} dx = \frac{59}{108}$$

The total risk is:

$$\int_0^{\frac{1}{2}} r_1(x)p(x) dx + \int_{\frac{1}{2}}^{\frac{11}{16}} r_3(x)p(x) dx + \int_{\frac{11}{16}}^2 r_2(x)p(x) dx = \frac{1}{12} + \frac{4}{27} + \frac{1}{216} = \frac{1377}{5832}$$

Bayesian decision theory (I)

The likelihood-ratio test (LRT)

Consider the simple two-class case: a_1 — classify as ω_1 , a_2 — classify as ω_2 .

Given a feature vector $\mathbf{x}$, we take action a_1 when $r(a_1|\mathbf{x}) < r(a_2|\mathbf{x})$:

$$l_{11}P(\omega_1|\mathbf{x}) + l_{12}P(\omega_2|\mathbf{x}) < l_{21}P(\omega_1|\mathbf{x}) + l_{22}P(\omega_2|\mathbf{x})$$

For $\mathbf{x} \in \mathcal{P}$, applying Bayes' formula and grouping terms:

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

Assuming (rather naturally) that $l_{21} > l_{11}$ and that $l_{12} > l_{22}$, then:

$$\Lambda(\mathbf{x}) = \frac{p(\mathbf{x}|\omega_1)}{p(\mathbf{x}|\omega_2)}; \quad \lambda = \frac{(l_{12} - l_{22})P(\omega_2)}{(l_{21} - l_{11})P(\omega_1)}$$

The test $\Lambda(\mathbf{x}) > \lambda$ (choose a_1) or $\Lambda(\mathbf{x}) < \lambda$ (a_2) is called the *likelihood-ratio test*.

Bayesian decision theory (I)

The likelihood-ratio test

Exercise 3 Consider a classification problem with two features $x_1, x_2 \in [0, 1]$ and two classes for which $P(\omega_1) = \frac{1}{3}, P(\omega_2) = \frac{2}{3}, p(\mathbf{x}|\omega_1) = 4x_1x_2, p(\mathbf{x}|\omega_2) = x_1 + x_2$.

Given two possible actions: a_1 — classify as ω_1 , a_2 — classify as ω_2 , with loss matrix:

	ω_1	ω_2
a_1	$\frac{1}{4}$	$\frac{1}{2}$
a_2	1	$\frac{1}{8}$

- Compute the LRT and represent it graphically
- Find the optimal decision rule and its associated risk

Bayesian decision theory (I)

0/1 losses

In many applications the 0/1 loss is used (usually in absence of more precise information):

$$l_{ij} = \begin{cases} 0 & \text{if } i = j \\ 1 & \text{if } i \neq j \end{cases}$$

Consider c classes and actions a_i — classify $\mathbf{x}$ into ω_i . Then:

$$r(a_i|\mathbf{x}) = \sum_{j=1}^c l_{ij}P(\omega_j|\mathbf{x}) = \sum_{j=1, i \neq j}^c P(\omega_j|\mathbf{x}) = 1 - P(\omega_i|\mathbf{x})$$

Bayesian decision theory (I)

Discriminant functions

Consider again the previous setting of c actions (classes) and a feature vector $\mathbf{x} \in \mathcal{P}$.

Discriminant functions of the form $g_k : \mathcal{P} \rightarrow \mathbb{R}$ are a useful conceptual tool to build an abstract classifier.

Given $\mathbf{x}$, it is assigned to class ω_i when $g_i(\mathbf{x})$ is the highest among the values $g_1(\mathbf{x}), \dots, g_c(\mathbf{x})$.

Examples:

- $g_k(\mathbf{x}) = P(\omega_k|\mathbf{x})$
- $g_k(\mathbf{x}) = P(\omega_k)p(\mathbf{x}|\omega_k)$
- $g_k(\mathbf{x}) = -r(a_k|\mathbf{x})$

If g_k is a discriminant function, then so is $h \circ g_k$, for any strictly monotonic function h .

Bayesian decision theory (I)

Discriminant functions

For two classes, we can use a single discriminant function (called a *dichotomizer*), instead of two.

Let $g(\mathbf{x}) = g_1(\mathbf{x}) - g_2(\mathbf{x})$; then assign $\mathbf{x}$ to class ω_1 if $g(\mathbf{x}) > 0$ and to class ω_2 if $g(\mathbf{x}) < 0$.

For instance, $g(\mathbf{x}) = p(\mathbf{x}|\omega_1)P(\omega_1) - p(\mathbf{x}|\omega_2)P(\omega_2)$

Exercise 4 *Show that this is equivalent to the corresponding LRT.*

Bayesian decision theory (I)

Exercise 5 Consider a classification problem with two features $x_1, x_2 \in [0, 1]$ and two classes for which $P(\omega_1) = \frac{1}{3}, P(\omega_2) = \frac{2}{3}, p(\mathbf{x}|\omega_1) = x_1 + x_2, p(\mathbf{x}|\omega_2) = 2x_1$.

and three possible actions: a_1 — classify as ω_1 , a_2 — classify as ω_2 , and a_3 — do not classify, with loss matrix:

	ω_1	ω_2
a_1	$\frac{1}{3}$	1
a_2	1	$\frac{1}{2}$
a_3	α	α

- Compute the LRT as a function of α
- Find the optimal decision rule as a function of α
- What is the smallest α for which action a_3 is never taken?
- What is the Bayes risk for this value of α ?

Bayesian decision theory (I)

Exercise 6 Consider a classification problem with one feature $x > 0$ and two classes for which $P(\omega_1) = \frac{3}{5}, P(\omega_2) = \frac{2}{5}, p(x|\omega_1) = e^{-x}, p(x|\omega_2) = \begin{cases} 1/2 & \text{if } x \in (0, 1] \\ e^{-2(x-1)}/2 & \text{if } x \geq 1 \end{cases}$

and two possible actions: a_1 — classify as ω_1 , a_2 — classify as ω_2 , with loss matrix:

	ω_1	ω_2
a_1	$\frac{1}{4}$	1
a_2	1	$\frac{1}{4}$

- Compute the LRT
- Find the optimal decision rule and the Bayes risk
- How often is an object classified into class ω_1 ?

Bayesian decision theory (I)

Worked Exercise (I)

The company *Bayesian Public Relations LTD* has been charged to organize an open-air fashion parade on Saturday night; today is Thursday.

Given x the current weather forecast, the weatherman says rain for Saturday night, with probability p . A decision has to be made today, because of the necessary preparations.

The company reasons:

1. If the event goes on and it finally does not rain, the company's reputation will be highly increased. Cost $l_{11} < 0$.
2. If the event goes on and it rains, it will be catastrophic. Cancellation after start, money back, reschedule with participants, damage to sponsors, etc. Cost $l_{12} \gg 0$.
3. If the event does not go on and it does not rain, it will also be bad (may be not so as the previous case). Criticisms, reschedule with participants, damage to sponsors, etc. Cost $l_{21} > 0$.
4. If the event does not go on and it rains, the company's reputation will be increased. However, the event has not taken place. Cost $l_{22} = 0$.

Bayesian decision theory (I)

Worked Exercise (II)

Let ω_1, ω_2 the classes “it will not rain” and “it will rain” (on Saturday night). We know $P(\omega_1|x) = 1 - p$ and $P(\omega_2|x) = p$. Let a_1 – “go on with it!”, a_2 – “better cancel it!”.

Then the event should go on if:

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

Therefore

$$\frac{1 - p}{p} > \frac{l_{12}}{l_{21} - l_{11}}$$

Putting $l_{12} = 2, l_{21} = 3, l_{11} = -1$, we obtain that the event should go on only if $p < \frac{2}{3}$.