西交利物浦大学
Xi'an Jiaotong-Liverpool University

Pattern Recognition

Lecture 04(a). Bayesian Decision Theory(cont.)

Dr. Shanshan ZHAO

School of AI and Advanced Computing
Xi'an Jiaotong-Liverpool University

Academic Year 2023-2024

Table of Contents

1. Recap
   - recap Bayesian Decision theory
   - recap Minimizing the misclassification rate
2. Minimizing the Average Risk(Expected loss)
3. Exercise

1. Recap
   - recap Bayesian Decision theory
   - recap Minimizing the misclassification rate
2. Minimizing the Average Risk(Expected loss)
3. Exercise

Bayesian Decision theory

Design classifiers to recommend decisions that minimize some total expected "risk".

- fundamental statistical approach to the problem of pattern classification
- ideal case, optimal classifier
- compare with all other classifiers

Decision Rule Using Posteriors

Recap: Using Bayes' rule, the posterior probability of category ω_j given measurement x is given by: Probabilities *Bayes rule*.

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

The *Bayes classification rule* can be stated as

- Decide ω_1 if $P(\omega_1|x) > P(\omega_2|x)$; or
- Decide ω_1 if $p(x|\omega_1)P(\omega_1) > p(x|\omega_2)P(\omega_2)$;

  Decision making relies on both the priors and the likelihoods and Bayesian Decision Rule combines them to achieve the minimum probability of error.

Error Probability

For the two class situation, we have

$$P(error|x) = \begin{cases} P(\omega_1|x) & \text{if we decide } \omega_2 \\ P(\omega_2|x) & \text{if we decide } \omega_1 \end{cases} \tag{1}$$

We can minimize the probability of error by :

$$\text{Decide} \quad \omega_1 \quad \text{if} \quad P(\omega_1|x) > P(\omega_2|x); \quad \text{otherwise decide} \quad \omega_2 \tag{2}$$

$$P(error|x) = \min[P(\omega_1|x), P(\omega_2|x)] \tag{3}$$

And , this minimizes the average probability of error too:

$$P(error) = \int_{-\infty}^{\infty} P(error|x)p(x)dx \tag{4}$$

Because the integral will be minimized when we can ensure each $P(error|x)$is as small as possible.

1. Recap
   - recap Bayesian Decision theory
   - recap Minimizing the misclassification rate
2. Minimizing the Average Risk(Expected loss)
3. Exercise

Minimizing the misclassification rate

Goal: To make as a few misclassifications as possible.
A mistake occurs when an input vector belongs to class ω_1 is assigned to class ω_2 or vice versa.

$$
\begin{aligned}
P(error) &= P(x \in \mathscr{R}_2, \omega_1) + P(x \in \mathscr{R}_1, \omega_2) \\
&= \int_{\mathscr{R}_2} p(x, \omega_1) dx + \int_{\mathscr{R}_1} p(x, \omega_2) dx \\
&= \int_{\mathscr{R}_2} p(x|\omega_1) P(\omega_1) dx + \int_{\mathscr{R}_1} p(x|\omega_2) P(\omega_2) dx
\end{aligned}
$$

Question: Is it true that the probability of misclassification is minimised by assigning each point to the class with the maximum posterior probability?

Minimizing the misclassification rate

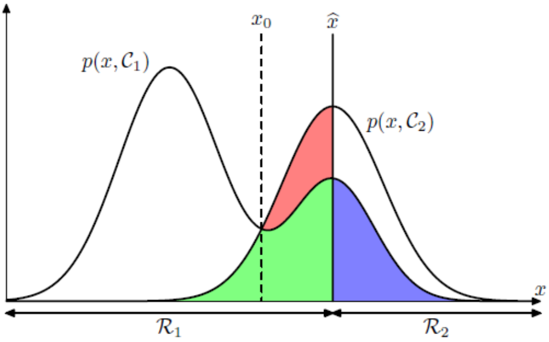

Figure 1: Schematic illustration of the joint probabilities $p(x, C_k)$ for each of two classes plotted against x, together with the decision boundary $x = \hat{x}$. *(from Bishop's PRML, page 40)*

Minimizing the misclassification rate

It is possible to extend this justification for a decision rule based on **maximum posterior probability**.
Therefore, we consider the probability for a pattern being correctly classified $P(correct)$.

Exercises:

- 1. $P(correct) = ?$
- 2. Prove that the maximum posterior probability decision rule is equivalent to minimising the probability of misclassification.

1.

$$\begin{aligned} P(correct) &= P(x \in \mathscr{R}_1, \omega_1) + P(x \in \mathscr{R}_2, \omega_2) \\ &= \int_{\mathscr{R}_1} p(x, \omega_1)dx + \int_{\mathscr{R}_2} p(x, \omega_2)dx \\ &= \int_{\mathscr{R}_1} p(x|\omega_1)P(\omega_1)dx + \int_{\mathscr{R}_2} p(x|\omega_2)P(\omega_2)dx \end{aligned}$$

2.

$$\begin{aligned} P(error) &= P(x \in \mathscr{R}_1, \omega_2) + P(x \in \mathscr{R}_2, \omega_1) \\ &= \int_{\mathscr{R}_1} p(x, \omega_2)dx + \int_{\mathscr{R}_2} p(x, \omega_1)dx \\ &= \int_{\mathscr{R}_1} p(x|\omega_2)P(\omega_2)dx + \int_{\mathscr{R}_2} p(x|\omega_1)P(\omega_1)dx \\ &= P(\omega_2)\int_{\mathscr{R}_1} p(x|\omega_2)dx + P(\omega_1)\int_{\mathscr{R}_2} p(x|\omega_1)dx \\ &= P(\omega_2)[1 - \int_{\mathscr{R}_2} p(x|\omega_2)dx] + P(\omega_1)[1 - \int_{\mathscr{R}_1} p(x|\omega_1)dx] \\ &= P(\omega_2) + P(\omega_1) - [\int_{\mathscr{R}_2} p(x|\omega_2)P(\omega_2)dx + \int_{\mathscr{R}_1} p(x|\omega_1)P(\omega_1)dx] \\ &= 1 - P(correct) \end{aligned}$$

Minimizing the Average Risk(Expected loss)

The classification error probability is not always the best criterion to be adopted for minimization. $\Longrightarrow$ assign a penalty term to weigh each error.
For the sake of generality, this *risk* or *loss* associated with $\omega_k(k = 1, 2, ..., M)$ is defined as

$$L_k = \sum_{i=1}^{M} \lambda_{ki} \int_{\mathscr{R}_k} p(x, \omega_i) dx \tag{5}$$

where, λ_{ki} is the loss incurred when a value of x, whose true class is ω_i, but we assign it to class ω_k. $\mathscr{R}_1$ denotes that region in feature space where the classifier decides ω_1 and likewise for $\mathscr{R}_2$ and ω_2, etc.

Minimizing the Average Risk(Expected loss)

The risk is (**Note that we use r_k to denote the term within the integral**):

$$L = \sum_{k=1}^{M} L_k = \sum_{k=1}^{M} \int_{\mathscr{R}_k} \left(\sum_{i=1}^{M} \lambda_{ki} p(x, \omega_i) \right) dx = \sum_{k=1}^{M} \int_{\mathscr{R}_k} r_k dx$$

Minimizing the risk is achieved if each of the integrals is minimized, which is equivalent to selecting partitioning regions so that $x \in \mathscr{R}_a \quad \text{if} \quad r_a \equiv \sum_{i=1}^{M} \lambda_{ai} p(x, \omega_i) \quad < \quad r_b \equiv \sum_{i=1}^{M} \lambda_{bi} p(x, \omega_i)$ $\forall b \neq a$

Minimizing the Average Risk(Expected loss)

The two-class case

$$\begin{aligned} r_1 &= \lambda_{11} p(x, \omega_1) + \lambda_{12} p(x, \omega_2) \\ r_2 &= \lambda_{21} p(x, \omega_1) + \lambda_{22} p(x, \omega_2) \end{aligned} \tag{6}$$

We assign x to ω_1 if $r_1 < r_2$, that is,

$$\begin{aligned} (\lambda_{12} - \lambda_{22}) p(x, \omega_2) &< (\lambda_{21} - \lambda_{11}) p(x, \omega_1) \\ (\lambda_{12} - \lambda_{22}) p(x|\omega_2) P(\omega_2) &< (\lambda_{21} - \lambda_{11}) p(x|\omega_1) P(\omega_1) \end{aligned} \tag{7}$$

The decision rule now becomes,

$$x \in \omega_1 \quad \textit{if} \quad r_{12} \equiv \frac{p(x|\omega_1)}{p(x|\omega_2)} > \frac{P(\omega_2)}{P(\omega_1)} \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \tag{8}$$

(It is natural to assume that $\lambda_{ij} > \lambda_{ii}$.)

Minimizing the Average Risk(Expected loss)

The decision rule now becomes,

$$x \in \omega_1 \quad if \quad r_{12} \equiv \frac{p(x|\omega_1)}{p(x|\omega_2)} > \frac{P(\omega_2)}{P(\omega_1)} \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}} \tag{9}$$

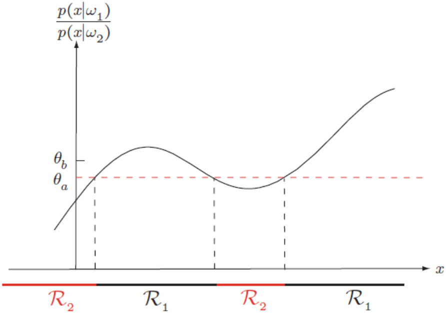

Figure 2: The likelihood ratio $p(x|\omega_1)/p(x|\omega_2)$

Exercise

In a two-class problem with a single feature x the pdfs are Gaussians with variance $\sigma^2 = 1/2$ for both classes and mean values 0 and 1, respectively, that is,

$$\begin{aligned} p(x|\omega_1) &= \frac{1}{\sqrt{\pi}} exp(-x^2) \\ p(x|\omega_2) &= \frac{1}{\sqrt{\pi}} exp(-(x-1)^2) \end{aligned} \tag{10}$$

If $P(\omega_1) = P(\omega_2) = 1/2$, compute the threshold value x_0

- (i) for minimum error probability
- (ii) for minimum risk if the loss matrix is

$$L = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{11} \end{bmatrix} = \begin{bmatrix} 0 & 0.5 \\ 1 & 0 \end{bmatrix} \tag{11}$$

Solution

(i) Let $p(x|\omega_1)P(\omega_1) = p(x|\omega_2)P(\omega_2)$,
we have

$$\frac{1}{\sqrt{\pi}}exp(-x^2) = \frac{1}{\sqrt{\pi}}exp(-(x-1)^2)$$

by solving the equation, we get

$$x_0 = \frac{1}{2}$$

Solution

(ii) Based on equation (9), we let

$$\frac{p(x|\omega_1)}{p(x|\omega_2)} = \frac{P(\omega_2)}{P(\omega_1)} \frac{\lambda_{12} - \lambda_{22}}{\lambda_{21} - \lambda_{11}}$$

to find the threshold value.
By substituting $P(\omega_2), P(\omega_1), \lambda$ into the equation above, we have

$$\frac{exp(-x^2)}{exp(-(x-1)^2)} = 0.5$$

$$2exp(-x^2) = exp(-(x-1)^2)$$

Take a logarithm

$$\ln 2 - x^2 = -(x-1)^2$$

$$x_0 = \frac{\ln 2 + 1}{2}$$

Thank You !

$\mathscr{Q}$ & $\mathscr{A}$