UCAS

PATTERN RECOGNITION

Assignment 1

September 19, 2019

说　明

1、请选5道证明题和2道编程题

2、作业用中文撰写，提交时，务必注明题号，鼓励使用LaTeX撰写

3、编程题需提交代码，命名按照题号命名，并注明运行环境

4、文档按“学号_姓名.pdf”命名，文档文件和代码文件全部打包成“学号_姓名.zip”提交

5、上机题要在“.pdf”中简要说清思路和方法

6、本次作业截止时间为2019年10月2日，请在课程网站上及时提交

Question 1

Consider the following decision rule for a two-category one-dimensional problem: Decide ω_1 if $x > \theta$; otherwise decide ω_2.

(a) Show the probability of error for this rule is given by $P(error) = P(\omega_1)\int_{-\infty}^{\theta} p(x|\omega_1)dx + P(\omega_2)\int_{\theta}^{\infty} p(x|\omega_2)dx$

(b) By differentiating, show that a necessary condition to minimize $P(error)$ is that θ satisfy $p(\theta|\omega_1)P(\omega_1) = p(\theta|\omega_2)P(\omega_2)$

(c) Does this equation define θ uniquely?

(d) Give an example where a value of θ satisfying the equation actually maximizes the probability of error.

Question 2

In many pattern classification problems one has the option either to assign the pattern to one of c classes, or to *reject* it as being unrecognizable. If the cost for rejects is not too high, rejection may be a desirable action. Let

$$\lambda(\alpha_i|\omega_i) = \begin{cases} 0 & i = j \quad i, j = 1, ..., c \\ \lambda_r & i = c+1 \\ \lambda_s & otherwise \end{cases}$$

where λ_r is the loss incurred for choosing the $(c+1)$-th action, rejection, and λ_s is the loss incurred for making a substitution error. Show that the minimum risk is obtained if we decide ω_i if $P(\omega_i|x) \geq P(\omega_i|x)$ for all j and if $P(\omega_i|x) \geq 1 - \frac{\lambda_r}{\lambda_s}$, and reject otherwise. What happens if $\lambda_r = 0$? What happens if $\lambda_r > \lambda_s$?

Question 3

Consider the minimax criterion for a two-category classification problem.

(a) Explain why the overall Bayes risk must be concave down as a function of the prior $P(\omega_1)$

(b) Assume we have one-dimensional Gaussian distributions $p(x|\omega_1) \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, 2$,but completely unknown prior probabilities. Use the minimax criterion to find the optimal decision point x^* in terms of μ_i and σ_i under a zero-one risk.

(c) For the decision point x^* you found, what is the overall minimax risk?

(d) Assume $p(x|\omega) \sim \mathcal{N}(0, 1)$ and $p(x|\omega_2) \sim \mathcal{N}(0.5, 0.25)$, under a zero-one loss. Find x^* and the overall minimax loss.

Question 4

Let $p(x|\omega_i) \sim \mathcal{N}(\mu_i, \sigma^2)$ for a two-category one-dimensional problem with $p(\omega_1) = p(\omega_2) = \frac{1}{2}$

(a) Show that the minimum probability of error is given by

$$P_e = \frac{1}{\sqrt{2\pi}} \int_a^\infty e^{-\frac{\mu^2}{2}} d\mu$$

where $a = \frac{|\mu_1 - \mu_2|}{2\sigma}$

(b) Use the inequality

$$P_e = \frac{1}{\sqrt{2\pi}} \int_a^\infty e^{-\frac{\mu^2}{2}} d\mu \leq \frac{1}{\sqrt{2\pi}a} e^{-\frac{a^2}{2}}$$

to show that P_e goes to zero as $\frac{|\mu_1 - \mu_2|}{\sigma}$ goes to infinity.

Question 5

Starting from the definition of entropy $H(p(x)) = -\int p(x)\ln p(x)dx$, derive the general equation for the maximum-entropy distribution given constraints expressed in the general form

$$\int b_k(x)p(x)dx = a_k, \quad k = 1, 2, ..., q$$

as follows:

(a) Use Lagrange undetermined multipliers $\lambda_1, \lambda_2, ..., \lambda_q$ and derive the synthetic function:

$$H_s = -\int p(x)[\ln p(x) - \sum_{k=0}^{q} \lambda_k b_k(x)]dx - \sum_{k=0}^{q} \lambda_k a_k$$

State why we know $a_0 = 1$ and $b_0(x) = 1$ for all x.

(b) Take the derivative of H_s with respect to $p(x)$. Equate the integrand to zero, and thereby prove that the minimum-entropy distribution obeys

$$p(x) = exp[\sum_{k=0}^{q} \lambda_k b_k(x) - 1]$$

where the $q+1$ parameters are determined by the constraint equation.

Question 6

Suppose we have two normal distributions with the same covariances but different means: $\mathcal{N}(\mu_1, \Sigma)$ and N $\mathcal{N}(\mu_2, \Sigma)$. In terms of their prior probabilities $P(\omega_1)$ and $P(\omega_2)$, state the condition that the Bayes decision boundary not pass between the two means.

Computer Exercises

Several of the computer exercises will rely on the following data:

	ω_1			ω_2			ω_3		
sample	x_1	x_2	x_3	x_1	x_2	x_3	x_1	x_2	x_3
1	−5.01	−8.12	−3.68	−0.91	−0.18	−0.05	5.35	2.26	8.13
2	−5.43	−3.48	−3.54	1.30	−2.06	−3.53	5.12	3.22	−2.66
3	1.08	−5.52	1.66	−7.75	−4.54	−0.95	−1.34	−5.31	−9.87
4	0.86	−3.78	−4.11	−5.47	0.50	3.92	4.48	3.42	5.19
5	−2.67	0.63	7.39	6.14	5.72	−4.85	7.11	2.39	9.21
6	4.94	3.29	2.08	3.60	1.26	4.36	7.17	4.33	−0.98
7	−2.51	2.09	−2.59	5.37	−4.63	−3.65	5.75	3.97	6.65
8	−2.25	−2.13	−6.94	7.18	1.46	−6.66	0.77	0.27	2.41
9	5.56	2.86	−2.26	−7.39	1.17	6.30	0.90	−0.43	−8.71
10	1.03	−3.33	4.33	−7.50	−6.32	−0.31	3.52	−0.36	6.43

Computer Exercise 1

Illustrate the fact that the average of a large number of independent random variables will approximate a Gaussian by the following:

(a) Write a program to generate n random integers from a uniform distribution $U(x_l, x_u)$.

(b) Now write a routine to choose x_l and x_u randomly, in the range $-100 \leq x_l < x_u \leq 100$, and and n (the number of samples) randomly in the range $0 < n \leq 1000$.

(c) Generate and plot a histogram of the accumulation of 10^4 points sampled as just described.

(d) Calculate the mean and standard deviation of your histogram, and plot it

(e) Repeat the above for 10^5 and for 10^6. Discuss your results.

Computer Exercise 2

Explore how the empirical error:

(a) Write a procedure to generate sample points in d dimensions with a normal distribution having mean μ and covariance matrix Σ.

(b) Consider $p(\mathbf{x}|\omega_1) \sim N((1,0)^t, I)$ and $p(\mathbf{x}|\omega_2) \sim N((-1,0)^t, I)$, with $P(\omega_1) = P(\omega_2) = \frac{1}{2}$. By inspection, state the Bayes decision boundary.

(c). Generate $n = 100$ points (50 for ω_1 and 50 for ω_2) and calculate the empirical error.