姓名:_____

学号:_____

院系:_____

___级 ___班

大 连 理 工 大 学

课 程 名 称：机器学习与模式识别　　试　卷：A

授课院(系)：信息与通信工程学院　考试日期：2020 年 08 月 19 日　试卷共 3 页

	1	2	3	4	5	总分
标准分	31	12	22	20	15	100
得　分						

注意：
1. 考试题目为英文，答题请用中文或者英文（不可混合）；
2. 请在答题纸上作答，并标清题号。

1. Please use 1-3 sentences to answer each of the following questions.

(1) (6pts) List three major sources of classification error.

(2) (5pts) List one drawback for using *k*-nearest neighbor method for density estimation.

(3) (5pts) Both principle component analysis and linear discriminant analysis are considered as feature reduction algorithms. What is the fundamental difference between them?

(4) (5pts) What is the curse of dimensionality?

(5) (5pts) Use one sentence to describe No Free Lunch Theorem.

(6) (5pts) Use one sentence to describe Ugly Duckling Theorem.

2. In the following algorithms, which can be considered as unsupervised approaches? Which are supervised ones? Please circle your answers. (3 pts each, 12 pts in total)

Adaptive boosting (unsupervised supervised)
k-Means clustering (unsupervised supervised)
Support vector machine (unsupervised supervised)
Principle component analysis (unsupervised supervised)

3. *k*-Means Clustering Algorithm (22 pts).

Supporse you are given 11 samples in a 2D plane:

(1,1) (1,5) (2,1) (3,1) (5,2) (3,4) (4,4) (2,5) (1,7) (2,8) (4,6)

(1) (16 pts) Use *k*-Means algorithm to divide the given samples into 2 clusters. Please demonstrate **the first iteration** of the implementation procedure and provide the resulting centroids after the first iteration.

Initial centroids: (1,1) for cluster 1; (1,6) for cluster 2.

(2) (6 pts) List two weaknesses of *k*-Means algorithm

4. Evaluating the Classifiers (20 pts).

(1) (6pts) What is the major difference between Jackknife algorithm and the simple cross-validation algorithm?

(2) (6pts) What is the advantage of Bootstrap algorithm comparing to Jackknife algorithm?

(3) (8pts) Receiver Operating Characteristic (ROC) curve can also be used to evaluate classifiers. In a given pattern recognition application, suppose we have tested two classifiers, and obtained their ROC curves as shown in Fig.1. In this application, which classifier performs better? Why?

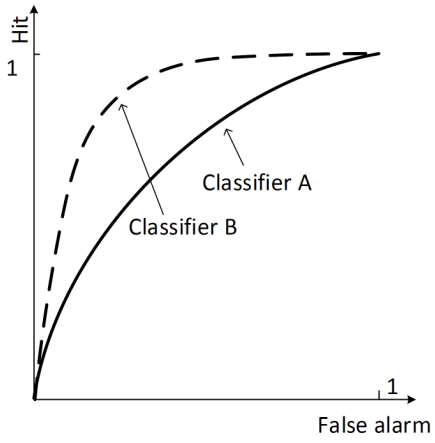

Figure 1: ROC curves of two classifiers.

5. Bayesian Decision Theory (15 pts).

Consider a one-dimensional two-category classification problem. The class-conditional densities of these two categories, $\omega_1$and ω_2, are given by

$$p(x|\omega_1) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{(x-2)^2}{2}\right\}, x \in (-\infty, +\infty)$$
$$p(x|\omega_2) = \begin{cases} 1/4 & 0 \le x \le 4 \\ 0 & \text{otherwise} \end{cases}.$$

(1) (5pts) Suppose the two categories have equal priors. Show the decision boundaries and decision regions using maximizing-a-posterior decision rule.

(2) (5pts) Suppose the two categories have equal priors and the loss function is given by

$$\lambda(\omega_1|\omega_2) = 10, \quad \lambda(\omega_2|\omega_1) = 20, \quad \lambda(\omega_1|\omega_1) = 0, \quad \lambda(\omega_2|\omega_2) = 0,$$

where $\lambda(\omega_i|\omega_j)$ denotes the loss for deciding ω_i when the state of nature is ω_j. Show the decision boundaries and decision regions using minimizing-the-risk decision rule.

(3) (5pts) Under the assumption of equal priors in both categories, we still use minimizing-the-risk decision rule, but change the loss function to

$$\lambda(\omega_1|\omega_2) = \alpha, \quad \lambda(\omega_2|\omega_1) = \beta, \quad \lambda(\omega_1|\omega_1) = 0, \quad \lambda(\omega_2|\omega_2) = 0,$$

where α and β are two positive constants, and $\alpha > \beta$. Will the decision region of category 1 grow bigger or smaller compared to that in (2)? Why?