

**The University of Nottingham Ningbo China**

SCHOOL OF COMPUTER SCIENCE

A LEVEL 3 MODULE, AUTUMN SEMESTER 2020-2021

**MACHINE LEARNING**

Time allowed: 2 hours

---

*Candidates may complete the front cover of their answer book and sign their desk card but must NOT write anything else until the start of the examination period is announced*

***Answer All Questions***

*Only silent, self contained calculators with a Single-Line Display or Dual-Line Display are permitted in this examination.*

*Dictionaries are not allowed with one exception. Those whose first language is not English may use a standard translation dictionary to translate between that language and English provided that neither language is the subject of this examination. Subject specific translation dictionaries are not permitted.*

*No electronic devices capable of storing and retrieving text, including electronic dictionaries, may be used.*

***DO NOT turn examination paper over until instructed to do so***

**Question 1:**

- (a) What is PCA? Briefly describe the goals of PCA? Is PCA a supervised learning technique or an unsupervised learning technique? Compare the similarity and difference between feature selection and PCA.

[8 marks]

- (b) Referring to decision tree, briefly explain: How to pick the root node? How to use entropy to associate information with probability?

[12 marks]

**Question 2:****[22 marks]**

- (a) We will use a single Perceptron classifier to classify 2D real number input into two classes (+1 and -1). The training data are listed as follows.

Input	Output
(2,2)	1
(4,-2)	-1
(-6,-2)	-1
(-6,2)	1

Train the Perceptron with online mode using the training data above from the first row to the last row, with the initialization of weights  $w=[1 \ 0 \ 0]^T$  and  $\eta=0.5$ . For each iteration, calculate the current output of Perceptron ( $o$ ) and the updated weights ( $w$ ). Note that you need to show intermediate steps of your calculation. An answer without the intermediate steps will not receive any marks.

[16 marks]

- (b) Perceptron and ADLINE both use a decision plane to separate two classes. Compare and contrast the two training techniques.

[6 marks]

**Question 3:****[30 marks]**

In a task to analyze the water quality of lakes, it is known that lakes come from one of two classes, natural lake  $C_1$  or artificial lake  $C_2$ . Each water quality sample of a lake  $Q$  has two features,  $Q = (q_1, q_2)$ . An experiment has collected 6 samples and their feature values and classification are shown in the following table. Answer the following questions by showing details of the working process. An answer without the working process will not receive any marks.

$Q = (q_1, q_2)$		Class
$q_1$	$q_2$	C
2	2	$C_1$
7	7	$C_2$
2	4	$C_1$

4	3	C <sub>1</sub>
5	4	C <sub>2</sub>
6	5	C <sub>2</sub>
5	2.5	=?

- (a) Based on the data with known class, design a KNN classifier with K=3 by the Euclidian distance to classify the unknown sample Q=(5, 2.5) in the last row of the table.

[10 marks]

- (b) Suppose the data in the above table are collected in a way that the class labels are unknown, i.e., we do not know C in the last column. Cluster the data into two clusters using K-Means algorithm with Euclidian distance and set the initial cluster centres as  $W_1=(2, 2)$  and  $W_2=(4, 3)$ .

[20 mark]

#### Question 4:

[28 marks]

- (a) Please list the 6 steps of Cross Industry Standard Process of Data Mining (CRISP-DM).

[6 marks]

- (b) Both classification and regression are machine learning techniques for prediction from collected data. Compare and contrast the two techniques.

[4 marks]

- (c) In Bayesian learning, one often adopt m-estimate of probability, whose formula is given below.

$$\frac{N_c + mP}{N + mP}$$

What  $N_c$ ,  $N$ ,  $m$ ,  $P$  stands for? How to choose  $P$ ? What is the advantage of adopting m-estimate?

[8 marks]

- (d) We often use kernel functions together with SVM for better classification performance. Briefly describe **what is a kernel function**, what problem it tackles and why it can tackle the problem.

[6 marks]

- (e) Recurrent Neural Network (RNN) is often used when predicting a sequence of outputs from a sequence of input data. Briefly describe the idea of RNN. You may **draw a simple diagram** to help illustrate the idea.

[4 marks]