

**THE UNIVERSITY OF HONG KONG**  
**FACULTY OF ENGINEERING**  
**DEPARTMENT OF COMPUTER SCIENCE**

**COMP7404 Computational Intelligence and Machine Learning**

**FINAL EXAMINATION**

**Date: 18 December 2018**

**Time: 6:30 pm – 8:30 pm**

- Write your university no. on every page.
- Answer all questions in this exam.
- Put your answers in the space provided in this exam paper only.
- Round your answers to 4 decimal places when necessary.
- Only approved calculators as announced by the Examinations Secretary can be used in this examination. It is candidates' responsibility to ensure that their calculator operates satisfactorily, and candidates must record the name and type of the calculator used on the front page of the examination script.

**Acknowledgment**

I have not engaged and will not engage in any activities that dishonestly improve my results or dishonestly improve/hurt the results of others in this exam.

Your University No.: \_\_\_\_\_

Date: \_\_\_\_\_

Question	Weight (in %)	Your Mark (in %)
1	12	
2	20	
3	24	
4	12	
5	23	
6	9	
<b>Total</b>	<b>100</b>	

### 1. Linear Regression

Shun Fat Consulting Company is considering renting an office in the Kowloon business districts. The table below lists the office sizes ( $x$ , in thousand square feet) and monthly rental prices ( $y$ , in thousand HK dollar) of the offices being considered. Based on the provided data, you are required to use the linear regression algorithm to develop a learning model to predict the rental price given an office size.

Office Size ( $x$ , in thousand square feet)	Rental Price ( $y$ , in thousand HK dollar)
8	5
4	2
15	9
12	7

(a) Assuming  $\theta_0=0$ , what is the computational cost of a linear regression model with  $\theta_1=0.1$ ? (show your steps in the space below and put your final answer in the box shown on right) (3 marks)

Your Answer

(b) Assuming  $\theta_0=0$ , what is the computational cost of a linear regression model with  $\theta_1=0.3$ ? (show your steps in the space below and put your final answer in the box shown on right) (3 marks)

Your Answer

(c) Assuming  $\theta_0=0$ , what is the computational cost of a linear regression model with  $\theta_1=0.6$ ? (show your steps in the space below and put your final answer in the box shown on right) (3 marks)

Your Answer

(d) Assuming that  $\theta_0=0$  and that you consider only these  $\theta_1$  values: 0.1, 0.3, 0.6, which  $\theta_1$  should be used to build the linear regression model? Why? (3 marks)

Your Answer

## 2. Decision Tree Classification

To determine whether a client belongs to high-risk category, the Granny Insurance Company collected a small number of client cases showing that age and insurance policy type can determine whether the client should be classified in the “high-risk” group. The following table lists the data instances with their classifications. Based on the provided data, you are required to use the ID3 algorithm to develop a decision tree for classifying whether a client is in the “High\_Risk” group.

Client ID	Age	Policy Type	High_Risk
1	Middle	A	Yes
2	Old	A	No
3	Old	B	Yes
4	Young	B	No
5	Young	A	Yes
6	Middle	B	No
7	Young	B	No

(a) What is the entropy at the root node of the decision tree? (show your steps of computation in the space below and put your final answer in the box shown on right) (3 marks)

Your Answer

(b) What is the information gain of using the “Age” feature to classify whether a client is in the “High Risk” group? (show your steps of computation in the space below and put your final answer in the box shown on right) (3 marks)

Your Answer

(c) What is the information gain of using the “Policy Type” feature to classify whether a client is in the “High Risk” group? (show your steps of computation in the space below and put your final answer in the box shown on right) (3 marks)

Your Answer

(d) Based on the ID3 algorithm for classifying the class of “High\_Risk” group, which feature should be used to split the data instances at the root node? Explain your choice quantitatively (ignore the feature “Client ID” in when choosing a feature to split data instances) (3 marks).

Your Answer

(e) If “Client ID” is used in classifying the “High Risk” group, what is the information gain? (3 marks)

Your Answer

(f) Compare the information gain considered in (e) with the information gain of the features considered in (b) and (c) above, should “Client ID” be used? Why? (3 marks)

(g) Define “overfitting” in terms of the errors when evaluating hypotheses using different datasets. (2 marks)



### 3. Logistic Regression

Causeway Bay Book Store recently issued its credit card. The table below provides data about the credit card applicants' income and age, and the actual decisions on their applications (1=approve; 0=deny). Based on the provided data, you are required to use the logistic regression algorithm to recommend decisions for new applications. Assuming that the logistic regression uses a sigmoid function  $h_{\theta}(x) = g(z)$  as explained in the course lecture, and that the algorithm predicts  $y=1$  if  $h_{\theta}(x) \geq 0.5$  and predicts  $y=0$  if  $h_{\theta}(x) < 0.5$ . Also,  $z = -40 + x_1 + x_2$ .

Applicant	Income ( $x_1$ , in thousand HKD)	Age Beyond 18 ( $x_2$ )	Actual Decision ( $y$ ) (1=approve; 0=deny)
1	20	30	1
2	25	40	1
3	17	12	0
4	18	10	0
5	20	8	0
6	18	34	1
7	39	51	1
8	12	23	0

(a) What is the predicted decision for Applicant 8? Show your computation in the space below and put your final answer in the box shown on right (3 marks).

Your Answer

(b) What is the predicted decision for a new applicant with  $x_1 = 16$  and  $x_2 = 20$ ? Show your computation in the space below and put your final answer in the box shown on right (3 marks).

Your Answer

(c) Suppose Causeway Bay Bookstore now considers three different classes in the decision for credit card application: deny, on-hold, approve. Specify below an "one-vs.-all algorithm" to use logistic regression for multi-class classification and for prediction on new input. Clearly fill in the following blanks to illustrate the steps of the algorithm (use the same notation introduced in the course lecture) (4 marks).

Training: For each class being considered:

Output:

Prediction: On a new input  $x$ :

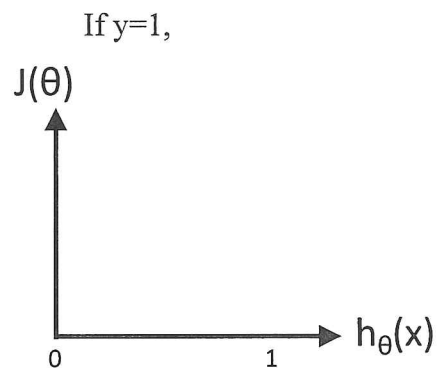
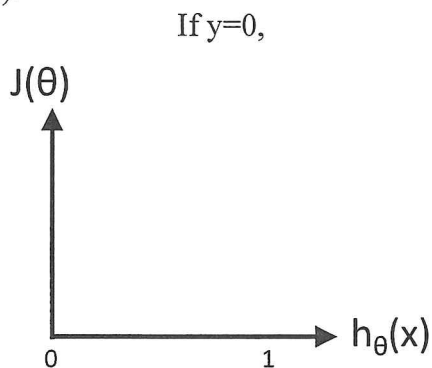
(d) If linear regression is used instead of logistic regression for this task, what are the problems? (6 marks)

Problem 1: \_\_\_\_\_

Problem 2: \_\_\_\_\_

Problem 3: \_\_\_\_\_

(e) For each of the two cases:  $y=0$  and  $y=1$ , sketch below the curve to show the relationship between cost of prediction error ( $J(\theta)$ ) and the prediction ( $h_\theta(x)$ ) of logistic regression algorithm (4 marks).



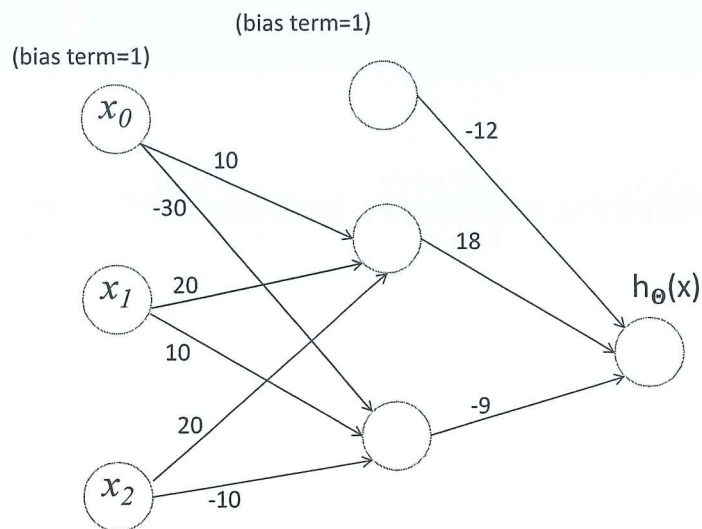
<p><u>If <math>y=0</math>,</u></p> <p><math>J(\theta) = -\log(1-h_\theta(x))</math></p> <p><math>J(\theta)</math> approaches _____ when <math>h_\theta(x)</math> approaches 0.</p> <p><math>J(\theta)</math> approaches _____ when <math>h_\theta(x)</math> approaches 1.</p>	<p><u>If <math>y=1</math>,</u></p> <p><math>J(\theta) = -\log(h_\theta(x))</math></p> <p><math>J(\theta)</math> approaches _____ when <math>h_\theta(x)</math> approaches 0.</p> <p><math>J(\theta)</math> approaches _____ when <math>h_\theta(x)</math> approaches 1.</p>
---	---

(f) Using notation introduced in the course lecture, write in the space below the cost function for applying gradient descent to the logistic regression algorithm (4 marks).

$J(\theta) =$

#### 4. Artificial Neural network

Wan Chai Financial Consulting uses the following artificial neural network to learn credit debt default instances of their customers. The two features  $x_1$  and  $x_2$  are credit score and family income (in thousand HK dollar) respectively. The predicted class is either default ( $y=1$ ) or not default ( $y=0$ ). The learned weights are shown on arrows in the following diagram.



(a) Label the empty nodes and sets of arrows in the diagram above with  $a$  and  $\Theta$  (having appropriate subscripts or superscripts consistent with course lecture) (2 marks).

(b) How many hidden layer(s) does this neural network contain? Write your answer in the box on right (2 marks).

Your Answer

(c) Based on the same activation function and notation in the course lecture, what is the predicted class for an input vector (300, 50)? Assume that  $x_0 = 1$  and that the algorithm predicts  $y=1$  if  $h_{\theta}(x) \geq 0.5$  and predicts  $y=0$  if  $h_{\theta}(x) < 0.5$ . Show your steps and computation in the space below and put your final answer in the box shown on right (8 marks).

Your Answer
-------------



**5. Clustering**

Mei Lai Cosmetics Company is trying to measure how their customers are different from each other. Each customer is represented by three attributes: age, income (in thousands dollar), and weight (in pound). Using the following methods, compute the distance between two customers: Mary (20, 70, 120) and Janna (24, 80, 110) (use the numbers shown here to compute the distance). Show your computation in the space below and put your final answer in the box shown on right.

(a) Euclidean distance (3 marks)

Your Answer

(b) Manhattan distance (3 marks)

Your Answer

(c) Minowski distance, using  $q = 4$  (3 marks)

Your Answer

(d) The purchase history of four customers (Alice, Betty, Candy, and Deborah) is shown in the table below ("Yes" refers to a prior purchase of the product; "No" refers to no prior purchase of the product). Assuming the purchase history is an asymmetric binary feature (with "Yes" being more valuable than "No"). Compute the asymmetric dissimilarity,  $D$ , between each of the following pairs of customers. Show your computation in the space below and put your final answer in the box shown on right (6 marks).

Customer	Product W	Product X	Product Y	Product Z
Alice	No	Yes	Yes	Yes
Betty	No	Yes	Yes	No
Candy	No	No	No	Yes
Deborah	Yes	No	Yes	Yes

i.  $D(\text{Alice}, \text{Betty}) =$

ii.  $D(\text{Betty}, \text{Candy}) =$

iii.  $D(\text{Candy}, \text{Deborah}) =$

(e) Mei Lai Cosmetics Company has a number of branches, each represented by the X and Y coordinates on a map: M1 (20, 100), M2 (20, 50), M3 (15, 10), M4 (5, 8), M5 (70, 50), M6 (10, 20), M7 (40, 90). The company is trying to cluster the branches and has chosen M1, M4, and M6 as the initial centroids of Clusters C1, C2 and C3 respectively. You are required to use the K-means algorithm to perform the first iteration computation. Upon completing the first iteration computation, you are then required to compute three new centroids and identify the members (branches) assigned to each cluster. Show your steps and computation in the space below and put your final answers in the table shown on the bottom of the next page (8 marks).

## Your Answers

Cluster	Coordinates of New Centroid		Cluster Members (circle the correct answer(s))				
C1	X =	Y =	<input checked="" type="checkbox"/> M1	<input type="checkbox"/> M2	<input type="checkbox"/> M3	<input type="checkbox"/> M5	<input type="checkbox"/> M7
C2	X =	Y =	<input checked="" type="checkbox"/> M4	<input type="checkbox"/> M2	<input type="checkbox"/> M3	<input type="checkbox"/> M5	<input type="checkbox"/> M7
C3	X =	Y =	<input checked="" type="checkbox"/> M6	<input type="checkbox"/> M2	<input type="checkbox"/> M3	<input type="checkbox"/> M5	<input type="checkbox"/> M7

**6. Evaluating machine learning models**

You developed different learning models to predict customer responses to email marketing. The following are the models' performances:

Model	Accuracy	Precision	Recall	Specificity
1	0.6	0.8	0.57	0.67
2	0.8	0.9	0.4	0.5
3	0.5	0.7	0.6	0.6

For each of the following parts, write the model number and reason on the lines provided.

(a) The model that performed the best in predicting correctly among the customers for whom the model has predicted that they did respond to marketing emails is (3 marks):

Model \_\_\_\_\_ Reason: \_\_\_\_\_

(b) The model that performed the best in predicting correctly among the customers who have actually responded to marketing emails is (3 marks):

Model \_\_\_\_\_ Reason: \_\_\_\_\_

(c) The model that performed the best in predicting correctly among the customers who actually have not responded to marketing emails is (3 marks):

Model \_\_\_\_\_ Reason: \_\_\_\_\_

END OF PAPER

*Please be reminded to write your university number on every page.*

This page is intentionally left blank. You may use it to draft your answers.