



MASENO UNIVERSITY
UNIVERSITY EXAMINATIONS 2021/2022

**FOURTH YEAR SECOND SEMESTER EXAMINATIONS FOR THE
DEGREE OF BACHELOR OF SCIENCE IN COMPUTER SCIENCE**

MAIN CAMPUS

CCS 418: ADVANCED DATABASES

Date: 28th October, 2021

Time: 3.30 - 6.30pm

INSTRUCTIONS:

- DO NOT write anywhere on this question paper
- Answer question ONE and any other TWO questions

MASENO UNIVERSITY

ISO 9001:2008 CERTIFIED



QUESTION ONE

- a) With the aid of a diagram, describe two types of transaction scheduling (4 marks)
- b) Differentiate between object database and object relational database (4 marks)
- c) Discuss any three types of knowledge to be mined in data mining environment (6 marks)
- d) Using suitable examples justify the need for adoption of multimedia databases in computer science environment (4 marks)
- e) Describe the architecture of a linked multi media database system (4 marks)
- f) Explain any four characteristics of a data ware house (8 marks)
 - both online & off line
 - stores historical data

QUESTION TWO

- a) Explain any three transparency issues in a distributed database environment (6 marks)
- b)
 - i. Define reconstructiveness (1 mark)
 - ii. Explain the significance of reconstructiveness in databases (3 marks)
- c) DDBMS architectures are generally developed depending on three parameters; Distribution, autonomy and Heterogeneity. Explain these 3 concepts on Peer-to-peer architectures (6 marks)
 - Peer-to-peer*
 - Heterogeneity*
- d) Differentiate between replication and fragmentation. (4 marks)
 - replication*
 - fragmentation*

whole db
is copied to
other dbs

a db is divided
into different portions/fragments.
and distributed to
different sites.

reassembling

QUESTION THREE

- a. Using a sample schema and relevant code, Explain the difference between Vertical fragmentation and horizontal fragmentation. (7 marks)
- b. Outline the main tiers/levels of interaction found in typical client/server architecture. (3 marks)
- c. Discuss the main differences that exists between a 2-tier and a 3-tier client/server architectures. (4 marks)
- d. Explain any four ways to characterize a multimedia database management system within database systems (6 marks)

QUESTION FOUR

- a. State what you understand by an object-oriented database system. (2 marks)
- b. Name any three limitations experienced by the relational database model and state how the object-oriented database model has managed to overcome each of the limitations. (6 marks)
into integrity, loss of data independence.
- c. Briefly explain how the object-oriented model ensures data security. (4 marks)
- d. By use of an example, state what a schema is, in as far object-oriented database systems is concerned. (2 marks)
- e. Discuss any three disadvantages of object-oriented database systems. (6 marks)

QUESTION FIVE

Given the following schemas :

EMPLOYEE (Empld, Name, City, Department, Salary)

PROJECT (PId, City, Department, Status)

WORKS (Empld, Pid, Hours)

- a) Write a query that displays records of John Doe and shows the following columns; Pid, status and name. (5 marks)

- b) Translate your query in a) above to relational algebra. Show your process. (6 marks)
- c) Draw a corresponding query tree for the relational algebra in (b) above (5 marks)
- d) Explain hash-join join operator computation approach (4 marks)



MASENO UNIVERSITY
UNIVERSITY EXAMINATIONS 2021/2022

**FOURTH YEAR SECOND SEMESTER EXAMINATIONS FOR THE
DEGREE OF BACHELOR OF SCIENCE IN INFORMATION
TECHNOLOGY**

MAIN CAMPUS

**ABS 424: ENTREPRENEURSHIP AND SMALL BUSINESS
MANAGEMENT**

Date: 25th October, 2021

Time: 3.30 – 6.30pm

INSTRUCTIONS:

- DO NOT write anywhere on this question paper
- Answer question ONE and any other THREE questions



Question One

Read the case below and answer the questions that follow:

The 'New Style' Tailor's shop at Miss Alice Simiyu

Alice Simiyu (Not hear real name) wants to open a tailor's shop. Sewing has been her hobby for many years. She taught herself how to make patterns, her hand - stitching is almost as perfect as machine stitching and for the last year she has been using her sister-in-law's sewing machine to sew curtains, cushion covers, pillow cases and other simple items for her family and friends and also one or two dresses for herself.

After graduating from Maseno University with a Diploma in Food Technology, Alice could not decide what to do next. One of her former lecturers asked her to work in his well-established shop, selling furniture and she did so far a few months but did not like it much and the salary was not very good either. Some of her friends had paid her for the clothes she had made for them. She knew that there was no other tailor in her rich estate and there were many people who could be her customers.

She estimated that if she could make a minimum of one dress a day she could earn more money than by selling furniture in her lecturer's shop. She was quite sure that her sister-in-law would let her use the sewing machine until she could afford to buy her own sewing machine.

So she decided to open a tailoring shop. The name was easy to find: "New Style". Now she could start her business. A friend suggested that she could have some leaflets printed with the name of her business, her address and a price list and distribute them at her instructor's shop and other places.

- a) What method of business idea generation do you think Alice has employed and why do you think so? (2marks)
- b) Do you think Alice has made the right decision? Justify your answer (8marks)
- c) Which entrepreneurial characteristics, if any does Alice display and which would you say are the right ones for the business she has gone into? Give convincing reasons for your answer (10marks)
- d) Explain the backward and forward linkages this business may create in the economy (5marks)

importance of business

9066647

Question Two

Discuss attributes that show you are cut for entrepreneurship and not employment (15marks)

Question Three

Are entrepreneurs made or born? Discuss using relevant case studies (15marks)

Question Four

How relevant are personal factors like age, gender and education in determining the level of creativity and innovation at individual level (15marks)

Question Five

Discuss the challenges entrepreneurs face as they go through the various stages of business growth (15marks)

Question Six

Using relevant examples, explain the factors that determine the success of a Franchise business (15marks)

Or

Many family business collapses immediately after the death of the founder. Explain the challenges associated with managing family business. What interventions can you suggest to address these challenges? (15marks)

- discuss business plans with the family
- good advisors
- include family members in the business
- Proper communication ~~clerk~~
- Proper finance budgeting
- Disperse

$$\begin{array}{r} & 53 \\ 27 & \times 3 \\ \hline 81 & \\ & 159 \\ \hline & 239 \end{array}$$



MASENO UNIVERSITY
UNIVERSITY EXAMINATIONS 2021/2022

**FOURTH YEAR SECOND SEMESTER EXAMINATIONS FOR THE
DEGREE OF BACHELOR OF SCIENCE IN COMPUTER SCIENCE,
BACHELOR OF SCIENCE IN MATHEMATICS AND COMPUTER
SCIENCE**

MAIN CAMPUS

CCS 422: ADVANCED COMPILER CONSTRUCTION

Date: 22nd October, 2021

Time: 7.30 - 10.30 am

INSTRUCTIONS:

- DO NOT write anywhere on this question paper
- Answer ALL questions in SECTION A and any other TWO from SECTION B
- Write your registration number on all sheets of the answer book used.
- Use a NEW PAGE FOR EVERY QUESTION attempted, and indicate number on the space provided on the page of the answer sheet.

MASENO UNIVERSITY

ISO 9001:2008 CERTIFIED



SECTION A: ANSWER ALL QUESTIONS

Question one (30 marks)

- a) Distinguish between machine independent and machine dependent optimizations. (4 marks)
- a) With the aid of a relevant diagram, explain the responsibilities of a compiler's frontend and the functions of its components. (8 marks)
- b) Use examples to explain the following loop optimization methods. (6 marks)
 - i) Loop-invariant code motion.
 - ii) Strength reduction.
 - iii) Induction variable elimination.
- c) State and briefly explain five properties to be considered when choosing an intermediate language. (5 marks)
- d) Given an expression $d := (a - b) + (a - c) + (a - c)$ and two registers R0 and R1.
Represent the expression in a three-address code then generate a resultant target code. (4 marks)
(3 marks)
- e) Compute the cost of the following target code.

*MOV b, R0
ADD c, R0
MOV R0, a
MOV a, R0
ADD e, R0
MOV R0, d*

S A S E N D O

SECTION B: ANSWER ANY TWO QUESTIONS

Question Five (20 marks)

- a) Explain briefly how the following transformations can be applied on a basic block. (10 marks)
- i) Common subexpression elimination.
 - ii) Dead-code elimination.
 - iii) Renaming of temporary variables.
 - iv) Interchange of two independent adjacent statements.
 - v) Algebraic Transformation.
- b) Differentiate between a flow graph and a directed acyclic graph. (2 marks)
- c) Given the three-address code representation below:
- 1) prod := 0
 - 2) i := 1
 - 3) t1 := 4 * i
 - 4) t2 := a[t1]
 - 5) t3 := 4 * i
 - 6) t4 := b[t3]
 - 7) t5 := t2 * t4
 - 8) t6 := prod + t5
 - 9) prod := t6
 - 10) t7 := i + 1
 - 11) i := t7
 - 12) if <= 20 goto (3)
- i) Identify the basic blocks then draw an equivalent flow graph for the above code. (4 marks)
 - ii) Draw a directed acyclic graph for the second basic block identified above. (4 marks)

Question Two (20 marks)

- a) Describe any four languages that can be used to represent intermediate code using a typical representation example in each case. (10 marks)
- b) Discuss briefly any five generic issues in the design of code generators. (10 marks)

Question Three (20 marks)

- a) Discuss five categories of optimizations using appropriate examples. (10 marks)
- b) Discuss peephole optimization using examples of program transformations that display its characteristics. (10 marks)

Question Four (20 Marks)

- a) Explain graph coloring by Simplification as a process of global allocation of register during target code generation. (6 marks)
- b) What are the rules to find "leader" in basic block? (3 marks)
- c) Specify four advantages of directed acyclic graphs. (3 marks)
- d) Construct a directed acyclic graph and optimal target code for the expression

$$X = ((a+b)/(b-c)-(a+b)*(b-c)+f).$$
 (8 marks)

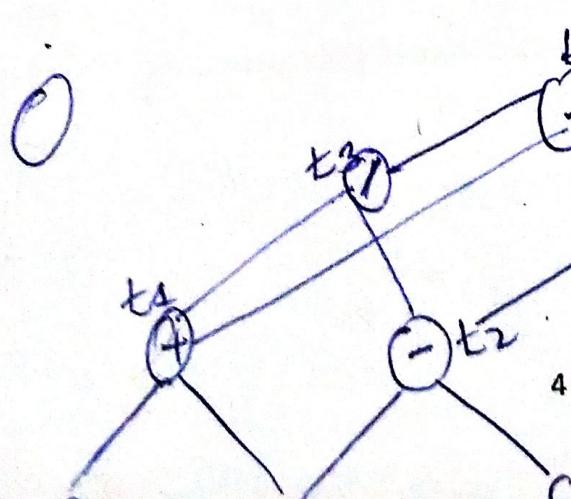
$$\begin{aligned}
 t_1 &:= a + b \\
 t_2 &= b - c \\
 t_3 &= t_1 + t_2 \\
 t_4 &= t_3 - t_1 \\
 t_5 &= t_4 * t_2 \\
 t_6 &= t_5 + f
 \end{aligned}$$

Mov g, R₀
 ADD b, R₀
 Mov

Mov. b, R₀
 ADD a, R₀
 MOV R₀, R₁
 SUB C, R₁

DIV R₀, R₁ R₁, R₁
 SUB R₀, R₁ R₀, R₁
 MUL R₁, R₀, R₀, R₁

ADD f, R₁
 MOV R₁, X





MASENO UNIVERSITY
UNIVERSITY EXAMINATIONS 2021/2022

**FOURTH YEAR SECOND SEMESTER EXAMINATIONS FOR THE
DEGREE OF BACHELOR OF SCIENCE IN COMPUTER SCIENCE,
BACHELOR OF SCIENCE IN MATHEMATICS AND COMPUTER
SCIENCE AND BACHELOR OF SCIENCE IN COMPUTER
TECHNOLOGY**

MAIN CAMPUS

**CCS 404/CCT 407: SOCIAL, LEGAL AND PROFESSIONAL ISSUES
IN COMPUTING**

Date: 22nd October, 2021

Time: 7.30 - 10.30 am

INSTRUCTIONS:

- DO NOT write anywhere on this question paper
- Answer question ONE (COMPULSORY) and any other TWO questions
- Start a new question on a new page

MASENO UNIVERSITY

ISO 9001:2008 CERTIFIED



SECTION A: COMPULSORY

Question One (30 Marks)

(a) Analyse the case study below using the doing ethics techniques and answer the questions below.

You are a computer programmer working for a small business that provides specialized financial services to local, mostly small businesses. You have been working for company X for about six months. Recently, X has been occupied with re-engineering the inventory system of a local hardware chain, ABC hardware. The objective is to enable ABC to keep better track of their inventory, to be more responsive to changes in customer demand and to adopt a "just in time" strategy to reduce inventory.

Your supervisor calls you in his office. "Do you know of any existing software products to help ABC keep better track of its inventory?" you mention a particular product that you have worked with in another job and point out that ABC could use it without any modifications. The only drawback, you point out, is that this software is somewhat expensive. Your supervisor leans back on his chair and says "that's no problem. We have that software. Why don't you just install it on ABC's computers?"

You diplomatically indicate that this would violate the licencing agreement X has with the developers of the software. "Do it anyway", your supervisor says. "Nobody's going to find out, and ABC is a very important client. We need to do all we can to make them happy."

Reference: The case of the troubled computer Programmer, William J. Frey

- | | |
|---|-----------|
| i. Who is affected? | [2 Marks] |
| ii. What are the ethical issues and implications? | [3 marks] |
| iii. What can be done about it? | [3 Marks] |
| iv. What are the options? | [3 Marks] |
| v. Which option is best- and why? | [3 Marks] |

(b) Explain the exclusive rights accorded to the owners of a copy right. [4 Marks]

(c) Discuss the purpose of professional bodies and using examples show how they can contribute to professional practice in organisations. [6 Marks]

(d) Discuss the categories of Private and Personal information in the context of Professional Issues in IT. [6 Marks]

Privacy
Access

Page 2 of 3
Accuracy

SECTION B: ANSWER ANY TWO QUESTIONS

Question Two (20 marks)

- a) Consider the case study in section A, and explain how each of the ethical theories (deontological and utilitarian) can be applied to argue that your supervisor's action is unethical. [5 Marks]
- b) Outline the benefits of adopting codes of conduct in the field of Computing and information Technology. [5 Marks]
- c) The Internet allows for free exchange of knowledge, a chaotic electronic freeway that now girdles the Earth. Discuss the ethical and legal considerations of Internet use. [10 Marks]

Question Three (20 marks)

- (a) What are some of the arguments used for or against the use of technology to monitor employees? Explain with the help of examples. [10 Marks]
- (b) At a recent computing conference, a delegate was heard commenting that, "The future credibility of our industry will only be assured once all computing practitioners adhere to policies and standards set by their professional body, after all, practitioners in professions such as Medicine and the Law need a license to practice". Do you agree or disagree with this statement? Give FOUR reasons to support your answer. [10 Marks]

Question Four (20 marks)

- (a) Do computer professionals have a presumed, or *prima facie*, obligation of loyalty to their employers? Explain. Why is it important to consider the ethical impact of any systems development? [10 Marks]
- (b) Why is freedom of expression not an absolute right? How is technology used to abuse this right? [10 Marks]

Question Five (20 marks)

- (a) What is informational privacy? Why are certain aspects and uses of internet search engines controversial from a privacy perspective? [10 Marks]
- (b) Explain how technology is an impediment to intellectual property protection. [10 Marks]



MASENO UNIVERSITY

UNIVERSITY EXAMINATIONS 2020/2021

**FOURTH YEAR SECOND SEMESTER EXAMINATIONS FOR THE
DEGREE OF BACHELOR OF SCIENCE IN COMPUTER SCIENCE
AND BACHELOR OF SCIENCE IN COMPUTER TECHNOLOGY**

MAIN CAMPUS

CCS 412/CCT 418: NATURAL LANGUAGE PROCESSING

Date: 18th October, 2021

Time: 3.00 – 6.00 pm

INSTRUCTIONS:

- DO NOT write anywhere on this question paper
- Answer question ONE and any other TWO questions.
- Uses of mobile phones are strictly forbidden



Question One Compulsory (30 Marks)

- A. Give example real-world applications of Natural Language Processing that you are familiar with (6 marks)
- B. Natural Language Processing is a non-trivial task. Give 2 challenges that make NLP difficult. (4 marks)
- C. Give example popular libraries that can be used for NLP (2 marks)
- D. Differentiate between Syntax and Semantics (4 marks)
- E. Differentiate between lemmatization and Stemming (4 marks)
- F. Explain the concept of Tokenization (2 marks)
- G. Explain the concept of Parsing as used in NLP (2 marks)
- H. What are N-grams? Give the correct bi-gram and tri-gram representations of the sentence "Udacity is a great source to learn Natural Language Processing Online." (6 marks)

Question two (20 marks)

- A. Briefly explain the meaning of stopwords and give examples (4 marks)
- A. Which two techniques can be used for the purpose of keyword normalization? (4 marks)
- B. Bag-of-words is a common document representation technique. Explain four limitations of using this technique. (6 marks)
- C. When creating a vector representation of a document, list 5 tricks that can be employed in order to reduce the dimensionality of the feature vector? (6 marks)
 - Stopword removal
 - correct spelling
 - ignore case
 - ignore punctuation

Question three (20 marks)

- A. Explain the concept of Text classification (2 marks)
- B. What are some common metrics used to test an NLP model e.g. a text classifier (6 marks)
- C. Differentiate between Term Frequency and Inverse Document Frequency. (6 marks)
- D. In a corpus of N documents, one document is randomly picked. The document contains a total of T terms and the term "data" appears K times. What is the correct value for the product of TF (term frequency) and IDF (inverse-document-frequency), if the term "data" appears in approximately one-third of the total documents? (6 marks)

Question four (20 marks)

- A. What is Named Entity Recognition and where can this technique be applied? (4 marks)
- B. Explain the concept of word sense disambiguation as used in NLP using an example (4 marks)
- B. Using an example, explain relation extraction as used in NLP (4 marks)
- C. Explain the NLP techniques that form part of the pipeline that can be used to complete the task of information extraction (8 marks)

0.4472

Question five (20 marks)

A. Differentiate between semantic similarity and lexical similarity

(4 marks)

B. You have a corpus that is made up of the below 4 documents:

Doc1: The sky is blue.

Doc2: The sun is bright today.

Doc3: The sun in the sky is bright.

Doc4: The shining sun, the bright sun.

i. Remove stop-words and list the vocabulary of the corpus

(2 marks)

ii. Represent each document using the bag-of-words document representation.

(4 marks)

iii. Use Jaccard similarity to determine the document that is most similar to Doc3

(5 marks)

iv. Use cosine similarity to determine the document that is most similar to Doc3.

(5 marks)



MASENO UNIVERSITY
UNIVERSITY EXAMINATIONS 2021/2022

**FOURTH YEAR SECOND SEMESTER EXAMINATION FOR THE DEGREE
OF BACHELOR OF SCIENCE IN COMPUTER SCIENCE, BACHELOR OF
SCIENCE COMPUTER TECHNOLOGY AND BACHELOR OF SCIENCE IN
MATHEMATICS AND COMPUTER SCIENCE**

MAIN CAMPUS

CCS 414/CCT 420: PATTERN RECOGNITION

Date: 1st November, 2021

Time: 3.30 – 6.30pm

INSTRUCTIONS:

- Do not write anywhere on this Question paper
- Answer Question ONE and any other TWO.

Question ONE(30mks)

- a) According to the Bayesian decision theory, a posterior probability, $P(w_i|x)$ is computed by the equation

$$P(w_i|x) = P(x|w_i)P(w_i)/P(x)$$

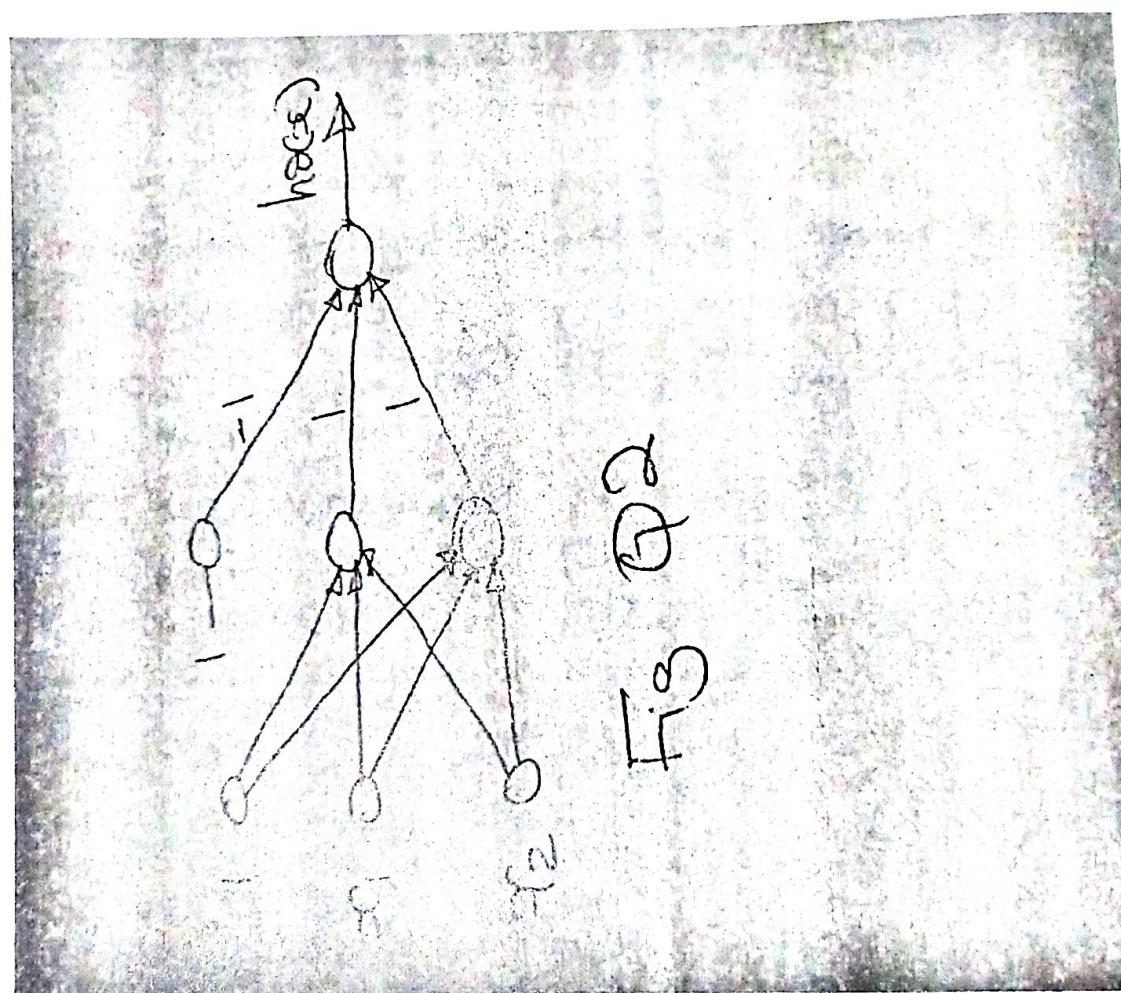
Briefly explain the meaning of the following probabilities in the equation

- i) $P(x|w_i)$ (2mks)
- ii) $P(w_i)$ (2mks)
- iii) $P(x)$ (2mks)

- b) Suppose you train a logistic classifier $h\theta(x)=g(\theta_0+\theta_1x_1+\theta_2x_2)$. Suppose $\theta_0=6, \theta_1=0, \theta_2=-1$. Sketch the decision boundary found by your classifier? (6mks)
- c) Which of the following statements are true? Check all that apply. Briefly explain your answer
- i) The sigmoid function $g(z)=1/(1+e^{-z})$ is never greater than one (>1) (2mks)
 - ii) For logistic regression, sometimes gradient descent will converge to a local minimum (and fail to find the global minimum). (3mks)
 - iii) Linear regression always works well for classification if you classify by using a threshold on the prediction made by linear regression. (3mks)
- d) i) Sketch the diagram of a typical pattern recognition system (3mks)
ii) Briefly describe the function of each of the stages in (d)(i) (7mks)

Question TWO (20 mks)

- a) Sketch the diagram of a biological neuron and explain its operation (4mks)
- b) Consider the multilayer perceptron of Fig Q2. Compute the output for the following activation functions:
 - i) Sigmoid (8mks)
 - ii) Unit step function (8mks)



Question THREE(20mks)

a) Briefly explain the principle of K-means clustering(5mks)

b) Consider the following eight points:

$P_1(2,2), P_2(1,14), P_3(10,7), P_4(1,11), P_5(3,4), P_6(11,8), P_7(4,3), P_8(12,9)$

Taking P_1, P_4 , and P_7 as initial centroids, apply K-means clustering algorithm($k=3$)

to calculate successive positions of the centroids after:

- i) One iteration (4mks) $\rightarrow \begin{matrix} 2,2 \\ 1,14 \\ 10,7 \\ 11,8 \end{matrix}$
- ii) Two iterations (4mks) $\rightarrow \begin{matrix} 2,2 \\ 1,11 \\ 2,96,2 \\ 1,72,11 \\ 4,6,3 \end{matrix}$
- iii) Three iterations (4mks) $\rightarrow \begin{matrix} 2,2 \\ 1,11 \\ 2,96,2 \\ 1,72,11 \\ 4,6,3 \\ 3,48,2 \\ 4,8,11 \\ 6,28,3 \end{matrix}$

c) Are there any changes to the positions of the centroids after the third iteration? Explain (3mks)

Question FOUR(20mks)

- a) Describe the Nearest Neighbour (NN) classification algorithm (6mks)
 b) Consider the following training data for NN classification:

x1	x2	classification(y)
7	7	0
7	4	0
3	4	1
1	4	1

Use the k-NN algorithm to correctly classify the entry $x_1 = 3, x_2 = 7$.

Show clearly every step of your computation (14 mks)

Training
 for $i = 1, \dots, n$ in the dimensions using dataset $D = \{d_i\}^n$
 storing algorithm $(x^{(i)}, f_x^{(i)})$

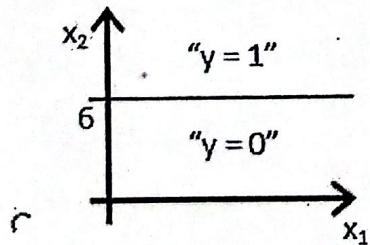
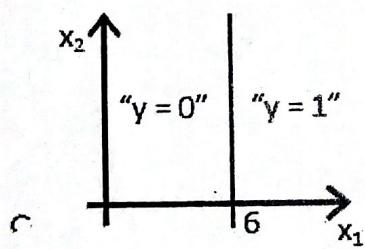
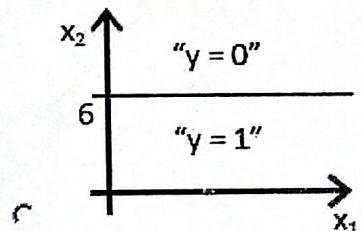
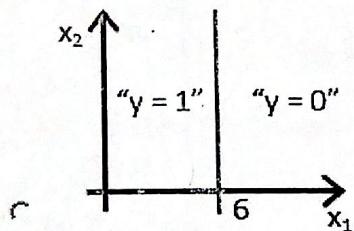
Prediction
 closest-point := none
 closest-distance := ∞

for $i = 1, \dots, n$
 current-distance := $d(x^{(i)}, x^{(v)})$
 if current-distance > current-distance
 then closest-distance := current-distance
 closest-point := $x^{(i)}$
 prediction $h(x^{(v)})$ is the target value of the closest point

Question FIVE(20mks)

- a) Suppose you train a logistic classifier $h\theta(x)=g(\theta_0+\theta_1x_1+\theta_2x_2)$.
Suppose $\theta_0=6, \theta_1=0, \theta_2=-1$. Which of the following figures represents the decision boundary found by your classifier?

Explain your answer (10mks)



- b) Part a) above refers to a two-class classifier. Explain how to achieve multiclass classification (10 mks)

Prediction algorithm
 (lossless-pair): none
 (lossy-pair): ∞
 $x^{(n)} = x^{(1)}, \dots, x^{(n)}$
 $f(x^{(1)}, \dots, x^{(n)}) = d(x^{(1)}, x^{(2)}, \dots, x^{(n)})$
 $\text{current-difference} > \text{lossy-difference}$
 $\text{current-difference} = \text{current-difference}$
 $x^{(i)} = \text{lossy-pair} = \text{current-difference}$
 $(\text{lossy-difference})^*$

A Training algorithm
for $i = 1, \dots, n$ in the dimensions using dataset $D = \{D\} = n$
(x_i) $f_{(i)}(x_i)$ training algorithm