



Sri Lanka Institute of Information Technology

B.Sc. Honours Degree in Information Technology

Final Examination
Year 4, Semester 1 (2024)

IT4060 – Machine Learning

Duration: 2 Hours

May 2024

Instructions to Candidates:

- ◆ This paper has 03 structured questions and 15 MCQ questions.
- ◆ Answer all questions in the spaces provided.
- ◆ The total mark for the paper is 100.
- ◆ Non-programmable calculators are allowed. No other electronic devices capable of storing and retrieving text, including mobile phones are not allowed.

Question 1

[20 marks]

a. A healthcare organization uses a machine learning model to predict patient risks for chronic illnesses. The model, however, was trained on a dataset predominantly consisting of urban populations.

i. Which challenge is this scenario most likely to encounter (Note that the answer is one of 'Bias', 'Data privacy' and 'Lack of Interpretability').

ii. Briefly explain (using 1-2 sentences) the potential impact of that challenge on the model's accuracy?

iii. Name one technique that can be used to reduce the impact of this challenge.

(3 marks)

b. A retail company develops a machine learning model to forecast sales based on historical sales data. The model is trained on data from the past five years, which include periods of economic stability.

i. If the economic conditions significantly change, such as during a recession, what challenge is the model most likely to face (Note that the answer is one of 'data privacy', 'lack of generalizability' or 'lack of interpretability')?

ii. Briefly justify (using 1-2 sentences) your answer to part i.

iii. Name one technique that can be used to reduce the impact of this challenge.

(3 marks)

c. A financial institution uses a complex machine learning model to decide whether to approve loans for applicants. Although the model achieves high accuracy, loan officers and customers often question its decisions due to the lack of transparency.

i. What challenge does this scenario highlight, and why is it important to address (Note that the challenge is one of "lack of data privacy", "lack of Generalizability" or "Lack of Interpretability").

ii. Briefly justify your answer to part i. using 1-2 sentences.

iii. Name one technique that can be used to address this challenge.

(3 marks)

d. A tech company is developing an image recognition system to categorize various urban landscapes such as parks, residential areas, and commercial zones from satellite images.

i. If the development team is planning to use a traditional (non Deep Learning) algorithm such as SVM to enhance the accuracy and to limit the size of the dataset necessary, state whether manual feature engineering is necessary or not for training this model.

ii. Briefly justify your answer using 1-2 sentences.

(2 marks)

e. A manufacturing company needs to allocate its limited resources optimally among various production lines to maximize output. The resources include raw materials, machine time, and labor. The company decides to use Linear Programming (LP) to solve this problem.

i. Name **two** conditions under which LP an appropriate optimization technique for this scenario.

ii. What are the expected **three** main components of the LP model?

(3 marks)

f. A technology company is developing a genetic algorithm to optimize the layout of components on a circuit board to minimize the circuit path length. The effectiveness of their genetic algorithm largely depends on the fitness function used.

i. Briefly describe (using 1 sentence) a fitness function that can be used for this purpose.

ii. Briefly justify your answer (using 1-2 sentences) to part i.

(3 marks)

g. A data science team at a healthcare analytics company is using linear regression to predict patient readmission rates based on various clinical and demographic factors. The team is interested in both preventing overfitting and identifying which factors are most influential in predicting readmissions. They consider using one of L1 (Lasso Regression) or L2 (Ridge Regression) regularization techniques that modify the loss function by adding a penalty term.

i. Which regularization technique should they use if they aim to enhance interpretability by performing feature selection?

ii. Briefly justify your answer using 1-2 sentences.

(3 marks)

Question 2

[25 marks]

a. A financial analyst is using gradient descent to train a linear regression model aimed at predicting stock prices based on economic indicators. Throughout the training process, the analyst notices that the training error not only fails to decrease but occasionally spikes dramatically with each iteration.

i. What could be the likely cause of this erratic training error?

ii. What adjustment should the analyst consider making to the training process?

(2 marks)

b. A financial analyst is developing a linear regression model to predict future stock prices from a range of economic indicators. During the model training phase using gradient descent, the analyst notices that while the error on the training set decreases to a very low value, the error on the validation set starts to increase after several iterations.

i. What is the most likely cause of this observed discrepancy between training and validation performance?

ii. Name two strategies that could be used to address it?

(3 marks)

c. In training a machine learning model to predict real estate prices, a data scientist uses features such as square footage, number of bedrooms, age of the property, and location. The model shows poor performance, particularly because the gradient descent algorithm converges very slowly and the training process is dominated by the square footage feature due to its much larger numerical range compared to other features like the number of bedrooms.

i. What is the name of the most effective method to address this issue.

ii. Briefly explain (using 1-2 sentences) how it addresses this issue?

(3 marks)

d. A technology company is developing a logistic regression model to predict whether users will click on personalized ads based on their browsing history. The model utilizes a sigmoid function to output probabilities that determine the likelihood of a user clicking on an ad.

What is the characteristic feature of the sigmoid function that makes it suitable for doing such predictions? (2 marks)

e. A marketing firm uses logistic regression to analyze and classify online shopping behaviors into three categories: Electronics, Clothing, and Home Goods. The firm has developed three logistic regression models, each predicting the likelihood of a customer's preference for one of these categories over the others. When analyzing a new customer's browsing data, the following probabilities are output by each model:

- Electronics model: 0.70
- Clothing model: 0.20
- Home Goods model: 0.10

(4 marks)

i. Based on these probabilities, into which category is this customer most likely to be classified, and why?

ii. What is the name given to this method of using multiple models to classify an unknown sample?

iii. Why does logistic regression require such an approach in this kind of classification problems?

f. A telecommunications company uses an SVM to classify customer satisfaction based on features such as service usage, billing amount, and customer support interactions. The data contains some noise due to discrepancies in how satisfaction is reported. The company employs slack variables in a soft margin SVM to handle this noisy data.

i. Why are slack variables critical in this SVM application, and how do they affect the classification of customer satisfaction?

ii. Briefly explain (using 1 sentence) how the parameter C in the loss function is related to the Slack variables, in SVM?

iii. Should C be increased or decreased to reduce overfitting?

(4 marks)

g. An online news portal uses an SVM to categorize news articles into topics such as politics, sports, and technology. Given the complexity and non-linear separability of text data, the portal implements kernel methods in their SVM.

i. How do kernel methods enhance the SVM's accuracy in classifying text data?

ii. Name one kernel function that can be used.

(3 marks)

h. The following dataset depicts the cardiovascular risk of patients based on 3 input features.

Patient	Age Category	SBP Category	Cholesterol Category	Risk Class
1	Young	Normal	Moderate	Low Risk
2	Young	High	Low	High Risk
3	Young	Normal	Low	Low Risk
4	Middle-aged	Very High	High	High Risk
5	Senior	High	Very High	High Risk
6	Middle-aged	Normal	High	Low Risk
7	Middle-aged	Low	Moderate	Low Risk
8	Middle-aged	Very High	High	High Risk
9	Senior	Very High	Moderate	High Risk
10	Senior	Very High	High	High Risk

i. Write an expression to calculate the entropy of the dataset (**note. you may write \log to represent \log base 2**).

ii. Write an expression to calculate the entropy of the young peoples' subset if the dataset is split based on age (**note. you may write \log to represent \log base 2**). Note that the entropy should be **normalized** by the proportion of the young peoples' dataset.

iii. Suppose that the information gain of age is 0.7, SBP category is 0.6, and cholesterol category is 1.3. What is the attribute that should be selected to split the dataset first?

(4 marks)

Question 3

[25 marks]

a. Imagine you are working on a project to predict customer churn for a telecom company using a Random Forest model. After training the model, you notice that it is overfitting the training data, leading to poor performance on the test data.

i. One method to address this overfitting is to change the ensemble method. Which ensemble method can be used to potentially reduce overfitting in this scenario? Briefly justify your answer (using 1-2 sentences).

ii. Another approach is to modify the number of trees in the Random Forest. How should the number of tree be altered to reduce overfitting? Briefly justify your answer (using 1-2 sentences).

(4 marks)

b. You are tasked with segmenting customers for a retail company to target marketing campaigns more effectively. The dataset includes features such as purchase history, demographics, and browsing behavior.

i. Name one main hyperparameter that needs to be set when applying K-means to solve this problem.

ii. What is the name of the heuristic based method that can be used to estimate that hyperparameter?

iii. Briefly explain (using 1-2 sentences) how the cost function is used in that method.

(4 marks)

c. You are working with a high-dimensional dataset from a gene expression study and need to compress the data before doing a predictive task. The dataset has thousands of features and you need to reduce the number of features.

i. Name one dimensionality reduction technique that can be used for this task.

ii. In the compression process '97% of variance is retained'. Briefly explain (using 1-2 sentences) the meaning of that claim.

(3 marks)

d. You are working on a project to analyze customer behavior for an online retail company. You have a dataset containing various features such as purchase history, browsing patterns, and demographic information. Your goal is to group similar customers together to tailor marketing strategies effectively.

You are interested in understanding the hierarchical relationships between different customer groups. You need a clustering algorithm that can provide a dendrogram to visualize these relationships and allow you to decide the number of clusters dynamically.

i. What is the most appropriate clustering algorithm that you can use in the above scenario?

ii. Name two distance functions that can be used along with that clustering algorithm.

(3 marks)

e. You are working on a project to analyze customer behavior for an online retail company. You have a dataset containing various features such as purchase history, browsing patterns, and demographic information. Your goal is to group similar customers together to tailor marketing strategies effectively.

The dataset contains a large number of customers with varying **densities** of data points, including some noise (**outliers**). You need a clustering algorithm that can handle these density variations and effectively identify the core clusters without being influenced by noise. Further, the clusters should be **non-overlapping**. Also, you need the algorithm itself to suggest the number of clusters, rather than you yourself defining that number.

i. What is the most appropriate clustering algorithm that can be used for this purpose?

ii. Name two hyper-parameters that are important in using this particular algorithm.

(3 marks)

f. A speech recognition system works by analyzing audio signals and converting them into phonemes, the smallest units of sound in a language. The system needs to handle the variability in pronunciation and noise in the audio input. A predictive model will help in recognizing the sequence of phonemes given the observed audio features.

i. Briefly explain (using 1-2 sentences) why the Hidden Markov Model (HMM) is suitable to do this conversion.

ii. Identify the hidden states and the observable states in the above scenario, if HMM is used to solve this problem.

(3 marks)

g. You are developing a navigation system for a delivery robot that operates in a warehouse with dynamically changing obstacles and partial visibility due to shelves and other objects blocking the view. The robot needs to learn the optimal path to deliver packages efficiently.

i. Why is Q-learning necessary for this scenario instead of policy iteration or value iteration?

ii. Identify what is meant by a **state** and the **action** in this scenario.

iii. Given the Q-update rule $Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \max_{a'} Q(s',a') - Q(s,a)]$, name and briefly explain the function of the term γ (gamma), within the context of this application.

(5 marks)

Machine Learning - IT4060

Started on	Friday, 24 May 2024, 1:23 PM
State	Finished
Completed on	Friday, 24 May 2024, 1:23 PM
Time taken	10 secs

Question 1

Not answered

Marked out of 2.00

What are the values of the feature x_2 after doing mean normalization?

Note: You may use 'range' as the denominator, for the normalization.

x_1	x_2
100	1000
200	2000
300	5000
200	2000

Select one:

- ☐ a. 0.125, 0.625, 0.125, -0.375
- ☐ b. None of the given answers are correct
- ☐ c. 0.375, 0.125, 0.625, -0.125
- ☐ d. -0.125, 0.625, -0.125, -0.375
- ☐ e. -0.375, -0.125, 0.625, -0.125

Question **2**

Not answered

Marked out of 2.00

Which of the following may NOT be useful features in classifying students that may graduate and not graduate, when they are being enrolled to a university degree? Suppose there is a id to uniquely identify each student. Note that there're multiple answers.

Select one or more:

- ☐ a. First name
- ☐ b. A level marks
- ☐ c. O level marks
- ☐ d. Height of the student
- ☐ e. School attended

Question **3**

Not answered

Marked out of 2.00

Which of the following are NOT appropriate reward signals for the given problems if Reinforcement learning is used? Note that there're multiple answers.

Select one or more:

- ☐ a. A model to train autonomous navigation to a robot - Time taken to travel to the goal
- ☐ b. A model to optimize the investment plan - The estimated cost of each investment
- ☐ c. A Reinforcement learning model to optimize the fuel economy of a hybrid car by optimizing the battery usage - Distance traveled per kilometer
- ☐ d. A model to optimize the investment plan - The expected profit of each investment
- ☐ e. A model to train autonomous navigation to a robot - The speed at which the robot is moving

Question 4

Not answered

Marked out of 2.00

	Buy	Not Buying
Free delivery	6	14
No Free delivery	10	16

Provided that free delivery is available on a particular day, what is the probability that the item will be purchased (**Note: round up the answer to one decimal place. e.g. 0.8**) ? You may use the following Bayesian rule in the calculation.

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

Answer:

Question **5**

Not answered

Marked out of 2.00

Which of the following are examples for anomaly detection problems? Note that there are multiple answers.

Select one or more:

- ☐ a. Grouping social media users based on their posts
- ☐ b. Detecting issues in a embedded system based on a collection of individual test cases
- ☐ c. Predicting the emotional status of a user based on facial expressions
- ☐ d. Detecting insurance fraud
- ☐ e. Recommending books in an online bookstore

Question 6

Not answered

Marked out of 2.00

Which of the following statements are NOT true regarding the Gaussian model for anomaly detection? Note that there are multiple answers.

Select one or more:

- ☐ a. If a particular attribute does not fit a Gaussian distribution, this approach can never be applied, even after doing a transformation to that attribute.
- ☐ b. Gaussian model is just one approach for anomaly detection and there exists multiple other algorithms that can be used for anomaly detection.
- ☐ c. Multivariate Gaussian distribution anomaly detection can be applied if the input attributes are not independent from each other
- ☐ d. If the input attributes are not independent of each other, this model or any variation of it can never be applied.
- ☐ e. If a particular input attribute does not fit a gaussian distribution, we can still consider that after doing some transformation to it to fit a Gaussian distribution.

Question **7**

Not answered

Marked out of 2.00

Which of the following are possible applications of recommender systems? Select multiple correct answers.

Select one or more:

- ☐ a. Suggesting personalized advertisements based on user browsing history.
- ☐ b. Automatic generation of financial reports for companies.
- ☐ c. Recommending movies or products based on user preferences and past behavior.
- ☐ d. Predicting changes in weather patterns over a decade.
- ☐ e. Optimizing traffic flow in real-time GPS navigation systems.

Question 8

Not answered

Marked out of 2.00

Which of the following are NOT true regarding recommender systems?

Select one or more:

- ☐ a. Content based filtering algorithm tries to estimate the input feature vectors of the items being recommended.
- ☐ b. Collaborative filtering algorithm only learns the users' attribute vectors
- ☐ c. Collaborative filtering algorithm estimates the users' attribute vectors and the input feature vectors of the items being recommended, continuously.
- ☐ d. Content based filtering algorithm tries to estimate the user attribute vectors based on the feature vectors of the items being recommended and previous ratings.
- ☐ e. In recommender systems the parameter vectors estimated are user specific, unlike in supervised learning.

Question 9

Not answered

Marked out of 2.00

Which of the following are NOT potential applications of Reinforcement learning? Note that there're multiple answers.

Select one or more:

- ☐ a. Customizing a question generation application for each user by learning from user responses
- ☐ b. Training an autonomous vehicle to drive based on experience
- ☐ c. Predicting whether a person has diabetes or not, based on historical medical data of patients
- ☐ d. Grouping the visitors to an ecommerce web site, based on their purchase history.
- ☐ e. Training a robot to navigate a previously unknown space

Question **10**

Not answered

Marked out of 2.00

A textbook manufacturing company, there are two types of books manufactured. The profit of manufacturing type A book is 10 Rs. The profit of manufacturing a type B book is 12 Rs. If we use linear programming, what could be the objective function that needs to be used, by varying the number of books manufactured of each type?

Select one:

- ☐ a. maximize $(10B + 12A)$
- ☐ b. None of the given answers are correct
- ☐ c. minimize $(10A + 12B)$
- ☐ d. maximize $(10A + 12B)$
- ☐ e. minimize $(10B + 12A)$

Question **11**

Not answered

Marked out of 2.00

Assume that an optimization algorithm is to be developed to maximize the fuel usage of a hybrid car engine, while maximizing battery life. Which of the following statements are NOT correct? Note that there're multiple answers.

Select one or more:

- ☐ a. The objective function should have the distance traveled per liter of fuel and the no. of battery discharges as its attributes.
- ☐ b. The cost of the battery should be a parameter in the objective function
- ☐ c. The objective function should only have the distance traveled per liter of fuel
- ☐ d. The size of the battery can be a constraint
- ☐ e. The cost of the battery can be a constraint

Question **12**

Not answered

Marked out of 2.00

Why might one choose to use a big data platform such as Apache Spark, for running machine learning algorithms? Note that there're multiple correct answers.

Select one or more:

- ☐ a. To improve the aesthetic presentation of data analysis results.
- ☐ b. To enhance the security of the data by implementing advanced encryption methods exclusive to big data platforms.
- ☐ c. To utilize distributed computing to process data faster and more efficiently.
- ☐ d. To handle large volumes of data efficiently, which is essential for training more accurate machine learning models.
- ☐ e. To reduce the costs associated with physical data storage by compressing data files.

Question **13**

Not answered

Marked out of 2.00

The true positive rate of a medical test is 95%. The true negative rate of a medical test is 97%. The probability that a selected person from the population has the disease is 0.01. Given a test is positive for a particular person, what is the probability that the person actually has the disease (**Note. round the answer to one decimal place, e.g. 0.7**) ? The Bayesian rule to be used is given below.

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

Answer:

Question **14**

Not answered

Marked out of 2.00

In a credit card fraud detection system, which of the following attributes may NOT be useful. Note that there are multiple answers.

Select one or more:

- ☐ a. Credit limit
- ☐ b. Number of transaction on the same day by the same card
- ☐ c. Transaction time
- ☐ d. CVC number (3 digit number at the back of the card)
- ☐ e. Number of letters in the card holders name

Question **15**

Not answered

Marked out of 2.00

Which of the following are examples of exploration in reinforcement learning? Note that there're multiple answers.

Select one or more:

- ☐ a. Navigate in a previously unknown direction in a robot navigation problem
- ☐ b. In a mineral exploration problem, try the locations that have been already mined
- ☐ c. Develop an investment plan based on previously known investment options and their returns
- ☐ d. Recommending dresses to an individual buyer, based on the ratings given by all the buyers.
- ☐ e. Try out a previously unknown new investment option in an investment plan

