

Question 1: [25 Marks]

Mr. Smith has a smart home system that regulates the heating and cooling according to the temperature of the house. The system uses fuzzy logic to decide whether to turn on the heater, air conditioner or do nothing. The temperature is classified into three categories: Cold, Comfortable and Hot. The action taken are classified into three categories: Heating, No Action and Cooling.

- Cold: $0 \leq T \leq 15$ Degree Celsius
- Comfortable: $15 < T \leq 25$ Degree Celsius
- Hot: $T > 25$ Degree Celsius

1.1 Explain why the fuzzy set as opposed to crisp set is more suitable to describe the classification temperature for Smith's smart home system. **[4]**

1.2 Consider the following conditions for Smith's smart home system.

- **Cold:** Membership is 1 for temperatures up to 10°C, and linearly decreases to 0 at 15°C.
- **Comfortable:** Membership increases linearly from 0 to 1 between 15°C and 20°C, remains 1 from 20°C to 25°C, and then linearly decreases to 0 between 25°C and 30°C.
- **Hot:** Membership is 0 for temperatures up to 25°C, increases linearly from 0 to 1 between 25°C and 30°C, and remains 1 for temperatures above 30°C.

Define the membership functions for the fuzzy sets of temperature classification for Mr. Smith's smart home system. Use the above conditions to calculate the slope for each linear segment and provide the membership functions in piecewise linear form. **[15]**

1.3 Determine the membership values of the temperature at 18 degrees Celsius in the fuzzy sets Cold, Comfortable, and Hot. **[6]**

Question 2 [25 Marks]

You are hosting a dinner party and need to arrange the seating for 10 guests around a rectangular table. The objective is to optimize the seating arrangement according to the following criteria.

- Certain pairs of guests prefer to sit next to each other.

- Certain pairs of guests prefer not to sit next to each other.
- Balance the seating so that the conversation can flow naturally across the table.

2.1 Outline the steps you would take to apply a GA to solve the seating arrangement problem. Describe each step briefly from initialization to termination. **[12]**

2.2 Describe how you would represent the seating arrangement as a chromosome in a Genetic Algorithm. Include details on encoding the order of guests and how you would ensure the representation is valid. **[6]**

2.3 Imagine you are training a deep neural network to classify images of animals into categories such as cats, dogs, and birds.

2.3.1 Explain the difference between local minima and global minima in the context of training the deep neural network for image classification. **[4]**

2.3.2 Explain the impact of overfitting in the context of local minima **[3]**

Question 3 [25 Marks]

3.1 Describe the main characteristics of supervised, unsupervised and semi supervised machine learning method. **[6]**

3.2 You are tasked with developing a machine learning model for a set of real-world problems as in Table 1. **In your answer booklet**, complete the table by identifying the most suitable machine learning method (supervise, unsupervised or semi-supervised) and provide the reasoning for your selection. **[9]**

Table 1: Learning problems

No	Learning problem	Suitable ML method (1 Mark)	Provide Justification (1 Mark)
3.2.1	A news aggregator website wants to organize its articles into topics without having predefined categories.		
3.2.2	A bank wants to detect fraudulent transactions. They have a large dataset of transactions, but only a small		

	subset is labelled as fraudulent or not		
3.2.3	An e-commerce company wants to segment its customers into distinct groups based on their purchasing behaviour to target marketing efforts more effectively		

3.3 You are developing a machine learning model to classify customer reviews into 'positive' or 'negative' sentiments for an e-commerce website. You have split your data into training, evaluation, and testing sets.

3.3.1 Describe how you would use these sets during the development process and explain how each set helps in assessing and improving the model's performance. **[6]**

3.3.2 Provide a brief explanation of how you would address potential issues of overfitting or underfitting using these datasets. **[4]**

Question 4 [25 Marks]

4.1 K-Means Clustering is a widely used algorithm in unsupervised machine learning for partitioning a dataset into distinct clusters. Selection of number of clusters (k) is a critical step in K-Means algorithm.

4.1.1 How does the choice of k affect the quality and interpretability of clusters in K-Means clustering? **[4]**

4.1.2 Describe the iterative process of K-Means clustering, including the assignment and update steps. How does this process lead to convergence? **[3]**

4.2 Consider a dataset with customer features: Annual Income, Age, Number of Purchases per Year, Average Purchase Value, Browsing Time per Visit, and Customer Loyalty Score. The code snippet provided in Figure 1 is applied to get the optimal cluster count for this dataset and Figure 2 illustrates the output.

4.2.1 What does WCSS stand for, and why is it an important metric in K-Means clustering? **[4]**

4.2.2 Based on the Elbow plot, what is the optimal cluster count and what visual clues are considered to identify the optimal cluster count from the Elbow plot. [3]

```
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
import numpy as np

# Example dataset
data = np.array([
    [60000, 25, 15, 200, 30, 8],
    [80000, 35, 20, 300, 45, 9],
    [50000, 30, 10, 150, 20, 7],
    [90000, 40, 25, 350, 50, 10],
    [70000, 45, 18, 250, 40, 8],
    [85000, 50, 22, 400, 55, 9],
    [55000, 20, 12, 100, 25, 6],
    [65000, 30, 16, 200, 35, 7]
])

# Normalize the data
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
data_scaled = scaler.fit_transform(data)

# Using the elbow method to determine the optimal number of clusters
wcss = []
for k in range(1, 9):
    kmeans = KMeans(n_clusters=k, init='k-means++', max_iter=300, n_init=10, random_state=42)
    kmeans.fit(data_scaled)
    wcss.append(kmeans.inertia_)

# Plotting the graph
plt.plot(range(1, 9), wcss)
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.show()
```

Figure 1: Sample Python code

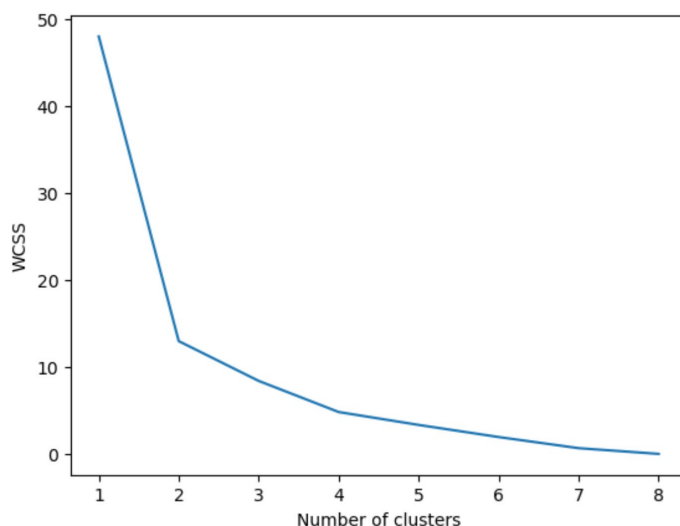


Figure 2: Output

4.3 Your university is developing an Artificial Neural Network (ANN)-based system to automate the assessment of student assignments and exams. The system is designed to classify

assignments into categories such as "Excellent," "Good," "Average," and "Needs Improvement" based on various features, including textual content, adherence to guidelines, and overall presentation. The aim is to provide consistent grading, personalized feedback, and actionable insights to both students and faculty.

The system will analyse students' submissions, compare them with a set of predefined criteria, and generate grades and feedback. The university is also concerned about the ethical implications and potential biases of using such a system.

4.3.1 Identify the type of classification problem this ANN model is addressing. Justify your answer by explaining the characteristics of the problem. **[3]**

4.3.2 Which activation function is most appropriate for the output layer of the ANN? Justify your choice by discussing how the selected function handles the classification task and ensures accurate results. **[4]**

4.3.3 What privacy and data protection measures should be implemented for the AI system used in student assessment? **[4]**

Diwedd/End