

Applied AI (AI4007)

Sessional-II Exam

Date: 11-04-2025

Course Instructor

Ms. Rushda Muneer

Total Time (Hrs): 1

Total Marks: 30

Total Questions: 3

2505

Roll No

BSE-6A

Section

TKS

Student Signature

Do not write below this line

Attempt all the questions on the given answer sheets.

CLO 4: Summarize concepts and approaches in knowledge representation, planning, learning, robotics and other AI areas.

Q1: Answer the questions regarding the dataset shown below. Determine whether Andrew finds a particular type of food appealing based on the food's temperature, taste, and size. [10 marks]

Appealing	Temperature	Taste	Size
No	Hot	✓ Salty	Small
No	Cold	✓ Sweet	Large
No	Cold	✓ Sweet	Large
Yes	Cold	Sour	Small
Yes	H	Sour	Small
No	H	✓ Salty	Large
Yes	H	Sour	Large
Yes	Cold	✓ Sweet	Small
Yes	Cold	✓ Sweet	Small
No	H	✓ Salty	Large

- (a) What is the initial entropy of Appealing? Write down the formula and calculations. [2 marks]
- (b) Assume that Taste is chosen for the root of the decision tree. What is the information gain associated with this attribute? Write down the formula and calculations. [3 marks]
- (c) If Taste is selected as the first attribute for the root of the decision tree, what should be the selected attributes for the next level? Calculate the entropy, conditional entropy and information gain to reason your selection and draw the decision tree. [5 marks]

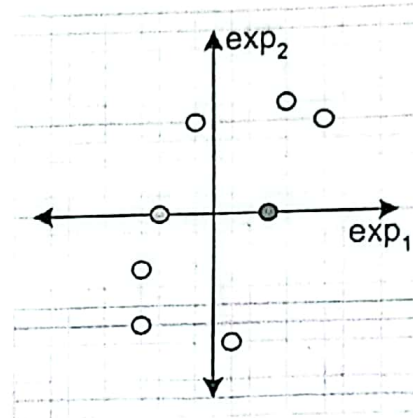
CLO 4: Summarize concepts and approaches in knowledge representation, planning, learning, robotics and other AI areas

Q2: Consider the following gene expression values from two microarray experiments for 8 genes. Negative expression in the table means that the gene had downregulated, while positive expression means that the gene as upregulated. A biologist is trying to find out whether these 8 genes can be

$$-\frac{4}{10} \left(\frac{2}{4} \log_2 \frac{2}{4} + \frac{2}{4} \log_2 \frac{2}{4} \right) - \frac{2}{4} \log_2 \frac{2}{4}$$

separated into two groups based on their behavior in the experimental conditions. In order to visualize the relationships, she has sketched a two-dimensional plot.

	exp_1	exp_2
$gene_1$	-4	-3
$gene_2$	6	5
$gene_3$	1	-7
$gene_4$	-4	-6
$gene_5$	4	6
$gene_6$	-1	5
$gene_7$	-3	0
$gene_8$	3	0

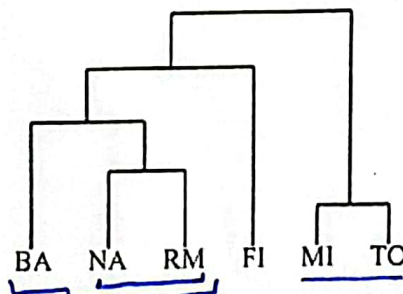


Use k-means clustering to divide these 8 genes into two clusters. Use gene 7 as the initial centroid of Cluster 1 and gene 8 as the initial centroid of Cluster 2. Use Euclidean distance for your calculations. Indicate the data points belonging to each cluster and give the coordinates of centroids at each iteration. Show first two iterations for determining the clusters. [10 marks]

CLO 4: Summarize concepts and approaches in knowledge representation, planning, learning, robotics and other AI areas

Q3: Give brief answers to the following questions.

- (a) Consider the following dendrogram. Using this dendrogram to create 3 clusters, what would the clusters be?



[2 marks]

- (b) Consider a dataset containing six one-dimensional points: {2, 4, 7, 8, 12, 14}. After three iterations of Hierarchical Agglomerative Clustering using Euclidean distance between points, we get the 3 clusters: $C_1 = \{2, 4\}$, $C_2 = \{7, 8\}$ and $C_3 = \{12, 14\}$. What is the distance between clusters C_1 and C_2 using Single Linkage and Complete Linkage? [4 marks]
- (c) What is a Confusion matrix and how is it used to calculate evaluation metrics for classification? Write down the evaluation metrics with their formula that can be computed via the confusion matrix. [4 marks]