

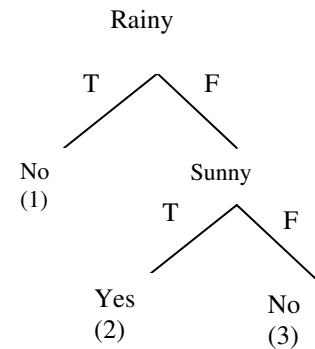


Full credit is 100. Start each question in a new page.
The exam is in **two** pages.

Question (1) (25 points)

- A. (6 points)** State the principle of Occam Razor. Why do you think it is important in the classification problem.
- B. (9 points)** What are the practical concerns that may arise when building a decision tree? Explain how you may deal with them.
- C. (10 points)** Consider the problem of deciding whether or not to go on a trip, based on various attributes of the day. Here is a set of examples, classified according to whether it was or was not a good idea to go on the trip. And next to it one possible decision tree for this problem

Example	Rainy	Sunny	Warm	Summer	Friday	Trip
Day1	T	F	F	F	F	No
Day2	F	T	F	F	T	Yes
Day3	F	T	T	T	T	No
Day4	F	T	T	F	T	No
Day5	T	F	F	F	T	No
Day6	F	T	F	F	T	No



- For this tree, which day(s) are false positives? (i.e. the decision “No” in the tree matches the decision No in the table)
- For this tree, which day(s) are false negatives? (i.e. the decision “No” in the tree does not match the decision No in the table)
- Suppose you could turn one leaf into a new attribute test; which would it be, the leaf marked (1) or leaf marked (2) or leaf marked (3)?
- Which new attribute test (Warm or Summer or Friday) provides the highest information gain?
- (True / False) There is a consistent decision tree for this training set. Explain why

Question (2) (25 points)

- A. (15 Points)** Consider the information system $S = \{U, A, V, f\}$:
- Define the lower and upper approximation of a set $X \subseteq U$
 - What is the approximation accuracy.
 - How can you determine the quality of an approximation of X
 - Show that: $\overline{apr}_p(X \cup Y) = \overline{apr}_p(X) \cup \overline{apr}_p(Y)$

- v. Show that: $\overline{apr_p}(X \cap Y) \subseteq \overline{apr_p}(X) \cap \overline{apr_p}(Y)$
vi. Show that: $\underline{apr_p}(X \cup Y) \supseteq \underline{apr_p}(X) \cup \underline{apr_p}(Y)$

B. (10 Points)

For the table shown :

- Extract the information system S
- Why is the attribute set in this table considered dependent.
- Find the information system S' after removing redundancy.

U	a ₁	a ₂	a ₃	a ₄
x ₁	1	1	1	1
x ₂	1	0	1	2
x ₃	0	0	1	0
x ₄	0	0	1	0
x ₅	1	0	0	2

- Remove attribute a₃ from S', then find the indiscernible elements.

Question (3) (25 points)

- A. (5 points)** How do you calculate the discernable matrix, why do you need to calculate only the lower triangle matrix.

B. (20 points)

For the information system, S, shown by the table:

- Calculate the equivalence classes of the system
- For $Y_1 = \{x_1, x_2, x_3, x_7, x_{10}\}$
 $Y_2 = \{x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$,
 $Y_3 = \{x_1, x_2, x_3, x_7, x_8\}$, $Y_4 = \{x_1, x_3, x_9\}$,
Show that these sets can be discernible in S
- Give an example of an indiscernible set in S.
- Find the equivalence classes in the system
- Construct the discernable matrix.
- Find the information system generated by the reduct {a₁, a₃}

U	a ₁	a ₂	a ₃
x ₁	0	1	2
x ₂	1	0	0
x ₃	1	0	0
x ₄	0	0	1
x ₅	0	0	1
x ₆	3	1	0
x ₇	1	0	0
x ₈	3	1	0
x ₉	3	1	3
x ₁₀	0	1	2

- Use the table to classify the following objects,
 $Y_1 = \{x_2, x_3, x_7\}$, $Y_2 = \{x_1, x_6, x_8\}$, $Y_3 = \{x_1, x_2, x_4, x_5, x_{10}\}$
Then, find the quality of your classification

Question (4) (25 points)

- A. (15 points)** How may the Genetic Algorithm be used in attribute reduction? Explain clearly the role of the crossover and mutation operators.

- B. (10 points)** Differentiate between the definition of Rough Fuzzy Set and Fuzzy Rough Set

Best Wishes
Areeg Abdalla