IN-CLASS EXERCISE (I5)

Student ID:..... Duration: 15 mins Date: 14/08/2023

Student name: ..... Score: ...../ 3

Q1 (2pts) A nutritional food company is developing a new product in the ABC country. Thus, they set out to study how age and lifestyle affect an individual's body shape. The following is the training data set.

#	Age<30	Eat Pizza	Exercise	Result
1	Yes	Yes	Yes	Fit
2	Yes	Yes	No	Fit
3	Yes	No	Yes	Fit
4	Yes	No	Yes	Fit
5	No	Yes	Yes	Unfit
6	No	Yes	No	Unfit
7	No	No	Yes	Fit
8	No	No	No	Fit

a) (1.5pts) Use **ID3 algorithm** to build a decision tree from the given data. Fill in the following tables with your numerical results, each of which is rounded to three decimal places. Attributes having the same best metric values are chosen in alphabetical order.

Evaluate all attributes and circle the attribute that is selected for the root node of the ID3 decision tree.

	Whole dataset	Age<30		Eat Pizza		Exercise	
		No	Yes	No	Yes	No	Yes
Entropy							
Average Entropy							
Information Gain							

For the branch that still contains a mixture of positive and negative examples, evaluate the remaining attributes and circle the attribute that is selected for the next node (you may leave irrelevant cells blank).

	Whole subset	Age<30		Eat Pizza		Exercise	
		No	Yes	No	Yes	No	Yes
Entropy							
Average Entropy							
Information Gain							

b) (0.5pt) Draw the complete decision tree.

Q2 (1pt) Identify which type of learning should be used to solve the following task. Explain your choice.
In the Pentagon (US) meeting room, there is a cabinet with ten drawers, each containing confidential documents for a specific department. Documents for different departments have specific patterns and colors. The assistant knows this rule and thus he is in charge of delivering incomings documents to the drawers.

.....

.....

.....

.....

IN-CLASS EXERCISE (I5)

Student ID:..... Duration: 15 mins Date: 14/08/2023

Student name: ..... Score: ...../ 3

Q1 (2pts) We have some data about when people go hiking. The data takes into effect, whether the hike is on a weekend or not, if the weather is rainy or sunny, and if the person has company during the hike.

#	Weekend?	Company?	Weather	Go Hiking?
1	Y	N	R	N
2	Y	Y	R	N
3	Y	Y	S	Y
4	Y	N	S	Y
5	Y	Y	R	N
6	Y	Y	S	Y
7	N	Y	S	N
8	N	Y	R	N
9	N	N	S	N

a) (1.5pts) Use **ID3 algorithm** to build a decision tree from the given data. Fill in the following tables with your numerical results, each of which is rounded to three decimal places. Attributes having the same best metric values are chosen in alphabetical order.

Evaluate all attributes and circle the attribute that is selected for the root node of the ID3 decision tree.

	Whole dataset	Company?		Weather		Weekend?	
		N	Y	R	S	N	Y
Entropy							
Average Entropy							
Information Gain							

For the branch that still contains a mixture of positive and negative examples, evaluate the remaining attributes and circle the attribute that is selected for the next node (you may leave irrelevant cells blank).

	Whole subset	Company?		Weather		Weekend?	
		N	Y	R	S	N	Y
Entropy							
Average Entropy							
Information Gain							

b) (0.5pt) Draw the complete decision tree.

Q2 (1pt) Identify which type of learning should be used to solve the following task. Explain your choice.
A teaching assistant is grading his students' coding assignments. He recognizes that several works have similar writing styles and bugs committed. He does not know precisely how many such groups. However, every time he finds more than two similar assignments, he marks them as a new group.

.....

.....

.....

.....

IN-CLASS EXERCISE (I5)

Student ID:..... Duration: 15 mins Date: 14/08/2023

Student name: ..... Score: ...../ 3

Q1 (2pts) A student is considering whether to go to the party or not. He is thinking about whether he may have a terrible hangover the next morning, whether the party is held at weekend, and how difficult the incoming exam is.

#	Hangover	Exam	Weekend	Party
1	No	Easy	No	Yes
2	No	Hard	No	No
3	No	No	No	Yes
4	No	No	Yes	Yes
5	Yes	Easy	No	No
6	Yes	Hard	No	No
7	Yes	No	No	No
8	Yes	No	Yes	No

a) (1.5pts) Use **ID3 algorithm** to build a decision tree from the given data. Fill in the following tables with your numerical results, each of which is rounded to three decimal places. Attributes having the same best metric values are chosen in alphabetical order.

Evaluate all attributes and circle the attribute that is selected for the root node of the ID3 decision tree.

	Whole dataset	Hangover		Weekend		Exam		
		No	Yes	No	Yes	Easy	Hard	No
Entropy								
Average Entropy								

For the branch that still contains a mixture of positive and negative examples, evaluate the remaining attributes and circle the attribute that is selected for the next node (you may leave irrelevant cells blank).

	Whole subset	Hangover		Weekend		Exam		
		No	Yes	No	Yes	Easy	Hard	No
Entropy								
Average Entropy								

b) (0.5pt) Draw the complete decision tree.

Q2 (1pt) Identify which type of learning should be used to solve the following task. Explain your choice.
You are training your dog to get the stick. Each time the dog returns a stick successfully, you offer it favorite treats. Eventually, the dog understands that whenever the master throws a stick, it should get it as early as possible to gain a reward (a bone) from a master in a lesser time.

.....

.....

.....

.....

IN-CLASS EXERCISE (I5)

Student ID:..... Duration: 15 mins Date: 14/08/2023

Student name: ..... Score: ...../ 3

Q1 (2pts) Let's say after your encounter with several people, you don't want vampires to be your friend in future. So you made a list of several people you met, their characteristics and if they turned out to be a vampire or not.

a) (1.5pts) Use **ID3 algorithm** to build a decision tree from the given data.

#	Casts shadows	Eat garlic	Skin complexion	Vampire
1	Dont know	Yes	Pale	No
2	Yes	Yes	Ruddy	No
3	Dont know	No	Ruddy	Yes
4	No	No	Average	Yes
5	Dont know	No	Average	Yes
6	Yes	No	Pale	No
7	Yes	No	Average	No
8	Dont know	Yes	Ruddy	No

Fill in the following tables with your numerical results, each of which is rounded to three decimal places. Attributes having the same best metric values are chosen in alphabetical order.

Evaluate all attributes and circle the attribute that is selected for the root node of the ID3 decision tree.

	Whole dataset	Casts shadow			Eat garlic		Skin complexion		
		Dont know	Yes	No	No	Yes	Average	Pale	Ruddy
Entropy									
Average Entropy									
Information Gain									

For the branch that still contains a mixture of positive and negative examples, evaluate the remaining attributes and circle the attribute that is selected for the next node (you may leave irrelevant cells blank).

	Whole subset	Casts shadow			Eat garlic		Skin complexion		
		Dont know	Yes	No	No	Yes	Average	Pale	Ruddy
Entropy									
Average Entropy									
Information Gain									

b) (0.5pt) Draw the complete decision tree.

Q2 (1pt) Identify which type of learning should be used to solve the following task. Explain your choice.
A robot enters a maze and finds a way to exit. He has not known what is at a specific square of the maze until he reaches it. He needs to explore different paths in the maze and sometimes he gets stuck with a dead end. However, he soon remembers the failures and tries not to repeat them.

.....

.....

.....

.....

SOLUTION

IN-CLASS EXERCISE (I5)

Student ID:..... Duration: 15 mins Date: 14/08/2023

Student name: ..... Score: ...../ 3

Q1 (2pts) A nutritional food company is developing a new product in the ABC country. Thus, they set out to study how age and lifestyle affect an individual's body shape. The following is the training data set.

#	Age<30	Eat Pizza	Exercise	Result
1	Yes	Yes	Yes	Fit
2	Yes	Yes	No	Fit
3	Yes	No	Yes	Fit
4	Yes	No	Yes	Fit
5	No	Yes	Yes	Unfit
6	No	Yes	No	Unfit
7	No	No	Yes	Fit
8	No	No	No	Fit

a) (1.5pts) Use **ID3 algorithm** to build a decision tree from the given data. Fill in the following tables with your numerical results, each of which is rounded to three decimal places. Attributes having the same best metric values are chosen in alphabetical order.

Evaluate all attributes and circle the attribute that is selected for the root node of the ID3 decision tree.

	Whole dataset	Age<30		Eat Pizza		Exercise	
		No	Yes	No	Yes	No	Yes
Entropy	0.811	1	0	0	1	0.918	0.722
Average Entropy		0.5		0.5		0.796	
Information Gain		0.311		0.311		0.015	

For the branch that still contains a mixture of positive and negative examples, evaluate the remaining attributes and circle the attribute that is selected for the next node (you may leave irrelevant cells blank).

	Whole subset	Age<30		Eat Pizza		Exercise	
		No	Yes	No	Yes	No	Yes
Entropy	1			0	0	1	1
Average Entropy				0		1	
Information Gain				1		0	

b) (0.5pt) Draw the complete decision tree.

Age<30 = No
|------- Eat Pizza = No: Fit
|------- Eat Pizza = Yes: Unfit
Age < 30 = Yes: Fit

Q2 (1pt) Identify which type of learning should be used to solve the following task. Explain your choice.
In the Pentagon (US) meeting room, there is a cabinet with ten drawers, each containing confidential documents for a specific department. Documents for different departments have specific patterns and colors. The assistant knows this rule and thus he is in charge of delivering incomings documents to the drawers.

Supervised learning. The number of classes is defined, which is ten drawers, each of which corresponds to a department. The attributes characterizing each department is the pattern and color of envelopes. Each envelope can only go to one of the designated drawers following its pattern and color.

IN-CLASS EXERCISE (I5)

Student ID:..... Duration: 15 mins Date: 14/08/2023

Student name: ..... Score: ...../ 3

Q1 (2pts) We have some data about when people go hiking. The data takes into effect, whether the hike is on a weekend or not, if the weather is rainy or sunny, and if the person has company during the hike.

#	Weekend?	Company?	Weather	Go Hiking?
1	Y	N	R	N
2	Y	Y	R	N
3	Y	Y	S	Y
4	Y	N	S	Y
5	Y	Y	R	N
6	Y	Y	S	Y
7	N	Y	S	N
8	N	Y	R	N
9	N	N	S	N

a) (1.5pts) Use **ID3 algorithm** to build a decision tree from the given data. Fill in the following tables with your numerical results, each of which is rounded to three decimal places. Attributes having the same best metric values are chosen in alphabetical order.

Evaluate all attributes and circle the attribute that is selected for the root node of the ID3 decision tree.

	Whole dataset	Company?		Weather		Weekend?	
		N	Y	R	S	N	Y
Entropy	0.918	0.918	0.918	0	0.971	0	1
Average Entropy		0.918		0.539		0.667	
Information Gain		0		0.379		0.252	

For the branch that still contains a mixture of positive and negative examples, evaluate the remaining attributes and circle the attribute that is selected for the next node (you may leave irrelevant cells blank).

	Whole subset	Company?		Weather		Weekend?	
		N	Y	R	S	N	Y
Entropy	0.971	1	0.918			0	0
Average Entropy		0.951				0	
Information Gain		0.2				0.971	

b) (0.5pt) Draw the complete decision tree.

Weather = S
|------- Weekend? = No: No
|------- Weekend? = Yes: Yes
Weather = R: No

Q2 (1pt) Identify which type of learning should be used to solve the following task. Explain your choice.
A teaching assistant is grading his students' coding assignments. He recognizes that several works have similar writing styles and bugs committed. He does not know precisely how many such groups. However, every time he finds more than two similar assignments, he marks them as a new group.

Unsupervised learning. The number of classes is not defined in advance. A different combination of writing style and bug introduce a new group and more combinations can be found during the grading.

IN-CLASS EXERCISE (I5)

Student ID:..... Duration: 15 mins Date: 14/08/2023

Student name: ..... Score: ...../ 3

Q1 (2pts) A student is considering whether to go to the party or not. He is thinking about whether he may have a terrible hangover the next morning, whether the party is held at weekend, and how difficult the incoming exam is.

#	Hangover	Exam	Weekend	Party
1	No	Easy	No	Yes
2	No	Hard	No	No
3	No	No	No	Yes
4	No	No	Yes	Yes
5	Yes	Easy	No	No
6	Yes	Hard	No	No
7	Yes	No	No	No
8	Yes	No	Yes	No

a) (1.5pts) Use **ID3 algorithm** to build a decision tree from the given data. Fill in the following tables with your numerical results, each of which is rounded to three decimal places. Attributes having the same best metric values are chosen in alphabetical order.

Evaluate all attributes and circle the attribute that is selected for the root node of the ID3 decision tree.

	Whole dataset	Hangover		Weekend		Exam		
		No	Yes	No	Yes	Easy	Hard	No
Entropy	0.954	0.811	0	0.918	1	1	0	1
Average Entropy		0.406		0.939		0.75		
		0.548		0.015		0.204		

For the branch that still contains a mixture of positive and negative examples, evaluate the remaining attributes and circle the attribute that is selected for the next node (you may leave irrelevant cells blank).

	Whole subset	Hangover		Weekend		Exam		
		No	Yes	No	Yes	Easy	Hard	No
Entropy	0.811			0.918	0	0	0	0
Average Entropy				0.689		0		
				0.123		0.811		

b) (0.5pt) Draw the complete decision tree.

Hangover = No
|------- Exam = Easy: Yes
|------- Exam = Hard: No
|------- Exam = No: No
Hangover = Yes: No

Q2 (1pt) Identify which type of learning should be used to solve the following task. Explain your choice.
You are training your dog to get the stick. Each time the dog returns a stick successfully, you offer it favorite treats. Eventually, the dog understands that whenever the master throws a stick, it should get it as early as possible to gain a reward (a bone) from a master in a lesser time.

Reinforcement learning. It is a trial-and-error process in which the agent only receives signal indicating success (found stick -> receive treats) or failure (not found stick -> no treats) and itself determines how to update the policy (find sticks in lesser time to get treats sooner)

IN-CLASS EXERCISE (I5)

Student ID:..... Duration: 15 mins Date: 14/08/2023

Student name: ..... Score: ...../ 3

Q1 (2pts) Let's say after your encounter with several people, you don't want vampires to be your friend in future. So you made a list of several people you met, their characteristics and if they turned out to be a vampire or not.

a) (1.5pts) Use **ID3 algorithm** to build a decision tree from the given data.

#	Casts shadows	Eat garlic	Skin complexion	Vampire
1	Dont know	Yes	Pale	No
2	Yes	Yes	Ruddy	No
3	Dont know	No	Ruddy	Yes
4	No	No	Average	Yes
5	Dont know	No	Average	Yes
6	Yes	No	Pale	No
7	Yes	No	Average	No
8	Dont know	Yes	Ruddy	No

Fill in the following tables with your numerical results, each of which is rounded to three decimal places. Attributes having the same best metric values are chosen in alphabetical order.

Evaluate all attributes and circle the attribute that is selected for the root node of the ID3 decision tree.

	Whole dataset	Casts shadow			Eat garlic		Skin complexion		
		Dont know	Yes	No	No	Yes	Average	Pale	Ruddy
Entropy	0.954	1	0	0	0.971	0	0.918	0	0.918
Average Entropy		0.5			0.607		0.689		
Information Gain		0.454			0.348		0.266		

For the branch that still contains a mixture of positive and negative examples, evaluate the remaining attributes and circle the attribute that is selected for the next node (you may leave irrelevant cells blank).

	Whole dataset	Casts shadow			Eat garlic		Skin complexion		
		Dont know	Yes	No	No	Yes	Average	Pale	Ruddy
Entropy	1				0	0	0	0	1
Average Entropy					0		0.375		
Information Gain					1		0.625		

b) (0.5pt) Draw the complete decision tree.

Casts shadows? = Dont know
|------- Eat garlic? = Yes: No
|------- Eat garlic? = No: Yes
Casts shadows? = Yes: No
Casts shadows? = No: Yes

Q2 (1pt) Identify which type of learning should be used to solve the following task. Explain your choice.
A robot enters a maze and finds a way to exit. He has not known what is at a specific square of the maze until he reaches it. He needs to explore different paths in the maze and sometimes he gets stuck with a dead end. However, he soon remembers the failures and tries not to repeat them.

Reinforcement learning. It is a trial-and-error process in which the agent must explore different paths and sometimes he gets stuck with a dead end. The more paths he has explored, the more knowledge he gets to update the policy (remember the failures and not repeat them)