ITE102 – Artificial Intelligence
Midterm Examination

Direction: Read the questions carefully. Choose the correct answer/s from the choices provided.

1. Which technology enables computers and machines to simulate human intelligence and problem-solving capabilities?

A. Artificial Intelligence
B. Blockchain
C. Quantum Computing
D. Virtual Reality

2. Which of the following describes AI systems designed for specific tasks and is most common today?

A. Weak AI
B. Strong AI

3. In 1950, Alan Turing proposed the Turing Test. What does it assess?

A. To see if a computer can win at chess
B. To test a computer's memory capacity
C. To check if a computer can be considered intelligent
D. To assess a computer's ability to perform calculations

4. Which of the following best defines "AI winter"?

A. Rapid AI advancements
B. Widespread AI adoption
C. Reduced AI funding and interest
D. Significant machine learning breakthroughs

5. Large Language Models (LLMs) are AI systems that can understand and generate text. Which of these can LLMs not do?

A. Create images
B. Summarize text
C. Translate languages
D. Write computer code

6. Which of the following describes a subfield of AI where computers learn from data to make decisions or predictions without explicit programming?

A. Robotics
B. Computer Vision
C. Machine Learning
D. Natural Language Processing

7. Which machine learning approaches involves training a model using an unlabeled dataset?

A. Supervised Learning
B. Unsupervised Learning

8. Which machine learning approaches would you use to compress an image by reducing its color range?

A. Supervised Learning
B. Unsupervised Learning

9. Which machine learning approaches would you use to identify a digit (0-9) from a handwritten image?

A. Supervised Learning
B. Unsupervised Learning

10. Which supervised learning algorithm can be used to predict house prices?

A. Linear Regression
B. Logistic Regression
C. K-means Clustering

11. Which supervised learning algorithm can be used to predict weight based on height?

A. Linear Regression
B. Logistic Regression
C. K-means Clustering

12. Which of the following is not considered unstructured data?

A. Images
B. Audio files
C. Spreadsheets
D. Text documents

13. What steps should you follow to build a machine learning model?

A. Collect Data → Clean Data → Split Data → Test Model → Train Model → Deploy Model
B. Collect Data → Clean Data → Split Data → Train Model → Test Model → Deploy Model
C. Collect Data → Split Data → Clean Data → Test Model → Train Model → Deploy Model
D. Collect Data → Split Data → Clean Data → Train Model → Test Model → Deploy Model

14. If you're building a model to predict Titanic passenger survival based on ticket class, age, and sex, what columns can you use as features? Check all that apply.

Ticket Class	Age	Sex	Survived?
3	17	Female	Yes
1	50	Male	Yes
3	34	Male	No
3	39	Female	No
3	27	Male	No

A. Ticket Class
B. Age
C. Sex
D. Survived

15. Consider the dataset below that you are using to build a model. How many duplicate rows need to be removed?

Ticket Class	Age	Sex	Survived?
3	30	Female	Yes
1	28	Male	Yes
1	28	Male	Yes
1	26	Male	Yes
3	4	Female	Yes
2	28	Male	No
2	28	Male	No
3	18	Female	No
3	41	Female	No
1	47	Male	No
3	9	Female	No
3	16	Female	Yes
3	19	Male	No
3	19	Male	No
1	37	Male	No

A. 1 B. 2 C. 3 D. 4 E. 5

16. There are three rows with missing values in the Age column. What value should we impute if we want to use the average age?

Ticket Class	Age	Sex	Survived?
2	39	Male	No
2	30	Female	Yes
3	19	Male	No
1	42	Female	Yes
3		Female	Yes
1	38	Male	No
1	63	Female	Yes
2	37	Male	No
2	19	Male	No
3		Female	Yes
2	18	Male	No
3		Male	No
3	11	Male	No
3	26	Female	Yes
3	44	Male	No

A. 19 B. 26 C. 32 D. 34

17. Machine learning algorithms require numerical input. How can you convert the Sex column to numeric values using one-hot encoding?

A. Update the Sex column: assign 1 for Male and 2 for Female.

Ticket Class	Age	Sex	Survived?
1	27	1	Yes
1	55	1	No
1	64	1	No
3	14	2	Yes
2	34	2	Yes
1	58	1	No
3	4	2	Yes
3	36	1	No
3	19	1	No
3	15	2	Yes
1	45	1	No
3	28	1	No
3	3	1	Yes
1	40	1	No
3	19	1	No

B. Add two columns: set 1 in Sex=Male if male, and 1 in Sex=Female if female.

Ticket Class	Age	Sex=Male	Sex=Female	Survived?
1	27	1	0	Yes
1	55	1	0	No
1	64	1	0	No
3	14	0	1	Yes
2	34	0	1	Yes
1	58	1	0	No
3	4	0	1	Yes
3	36	1	0	No
3	19	1	0	No
3	15	0	1	Yes
1	45	1	0	No
3	28	1	0	No
3	3	1	0	Yes
1	40	1	0	No
3	19	1	0	No

18. If your dataset has 1500 instances and you want to split it 80% for training and 20% for testing, how many instances go into each set?

A. Train set: 300 ; Test set: 1200
B. Train set: 450 ; Test set: 1050
C. Train set: 1050 ; Test set: 450
D. Train set: 1200 ; Test set: 300

19. Which supervised learning algorithm predicts a continuous value by finding the best-fit line or surface?

A. Linear Regression
B. Logistic Regression
C. K-means Clustering
D. Neural Networks

20. [2 pts] Suppose you are developing a model to predict points per game based on the number of training hours per week for players. Which of the following lines best fits the data when evaluating the Mean Squared Error (MSE) between the actual data points and the line?

You may use the table below as a guide for calculating the error of each line.

x (Training Hours Per Week)	y (Points Per Game)	$y_1 = x + 4$	$(\|y - y_1\|)^2$	$y_2 = x + 5$	$(\|y - y_2\|)^2$	$y_3 = x + 6$	$(\|y - y_3\|)^2$
10	15						
12	18						
15	20						
20	25						
18	22						
14	17						
22	28						
11	16						
16	19						
13	21						
		Sum of $(\|y - y_1\|)^2$		Sum of $(\|y - y_2\|)^2$		Sum of $(\|y - y_3\|)^2$	

A. $y = x + 4$
B. $y = x + 5$
C. $y = x + 6$

21. [2 pts] What is the computed Mean Squared Error for each of the following lines, in order:

1. y = x + 4
2. y = x + 5
3. y = x + 6

A. MSE of Line 1: 20 ; MSE of Line 2: 30 ; MSE of Line 3: 30
B. MSE of Line 1: 30 ; MSE of Line 2: 20 ; MSE of Line 3: 30
C. MSE of Line 1: 30 ; MSE of Line 2: 30 ; MSE of Line 3: 20

22. Which supervised learning algorithm predicts the probability of an observation belonging to a class for binary classification?

A. Linear Regression
B. Logistic Regression
C. K-means Clustering
D. Neural Networks

23. Suppose you are predicting whether an email is spam, and you set the classification threshold at 80%. If the predicted probability of an email being spam is 0.75, how would this email be classified?

A. Spam
B. Not Spam

24. When classifying multiple labels, the label with the highest probability is chosen. Given the prediction probabilities for the handwritten image below, which number (0-9) is predicted?

Label	Probability
0	0.00
1	0.00
2	0.05
3	0.00
4	0.00
5	0.00
6	0.00
7	0.00
8	0.91
9	0.04

A. 0
B. 1
C. 2
D. 3
E. 4
F. 5
G. 6
H. 7
I. 8
J. 9

25. Suppose you are predicting whether an image contains a cat. In your dataset, non-cats are dogs. Given the following results:

- 615 images were correctly identified as cats.
- 570 images were correctly identified as dogs.
- 195 images of dogs were incorrectly identified as cats.
- 120 images of cats were incorrectly identified as dogs.

What does the confusion matrix look like?

A.

		Predicted	
		Positive	Negative
Actual	Positive	570	120
	Negative	195	615

C.

		Predicted	
		Positive	Negative
Actual	Positive	615	120
	Negative	195	570

B.

		Predicted	
		Positive	Negative
Actual	Positive	570	195
	Negative	120	615

D.

		Predicted	
		Positive	Negative
Actual	Positive	615	195
	Negative	120	570

For the next four questions, evaluate the performance metrics using the confusion matrix provided below.

		Predicted	
		Positive	Negative
Actual	Positive	800	200
	Negative	100	900

26. What is the model's accuracy?

A. 0.80
B. 0.84
C. 0.85
D. 0.88

27. What is the model's precision?

A. 0.80
B. 0.84
C. 0.85
D. 0.88

28. What is the model's recall?

A. 0.80
B. 0.84
C. 0.85
D. 0.88

29. What is the model's F1-score?

A. 0.80
B. 0.84
C. 0.85
D. 0.88

30. Which unsupervised learning algorithm groups unlabeled data into clusters?

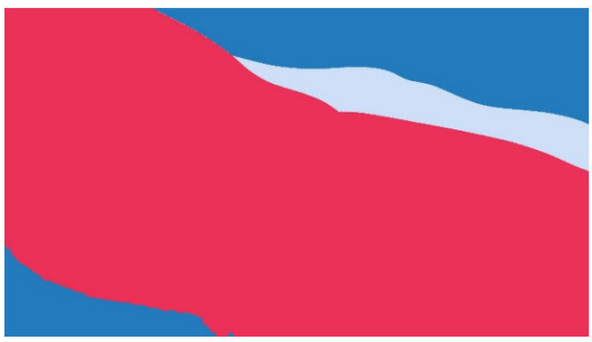
A. Linear Regression
B. Logistic Regression
C. K-means Clustering
D. Neural Networks

31. Suppose you are reducing the file size by decreasing the number of colors through clustering. Given the original photo, which of the following has been compressed to 4 clusters?

A.

C.

B.



D.

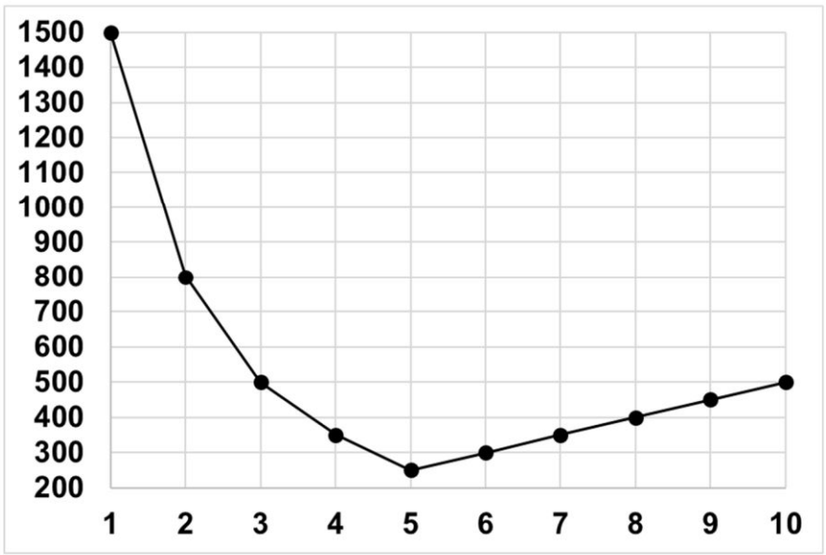
32. Given this group of data points, what is the centroid of the cluster?

Petal Width (cm)	Petal Length (cm)
1.4	3.5
1.5	3.7
1.3	3.6
1.7	4.0
1.5	3.9
1.4	3.8
1.6	4.1
1.8	4.2
1.5	3.7
1.6	4.0

A. Petal width: 1.53 ; Petal length: 3.85
B. Petal width: 1.80 ; Petal length: 4.20
C. Petal width: 3.85 ; Petal length: 1.53

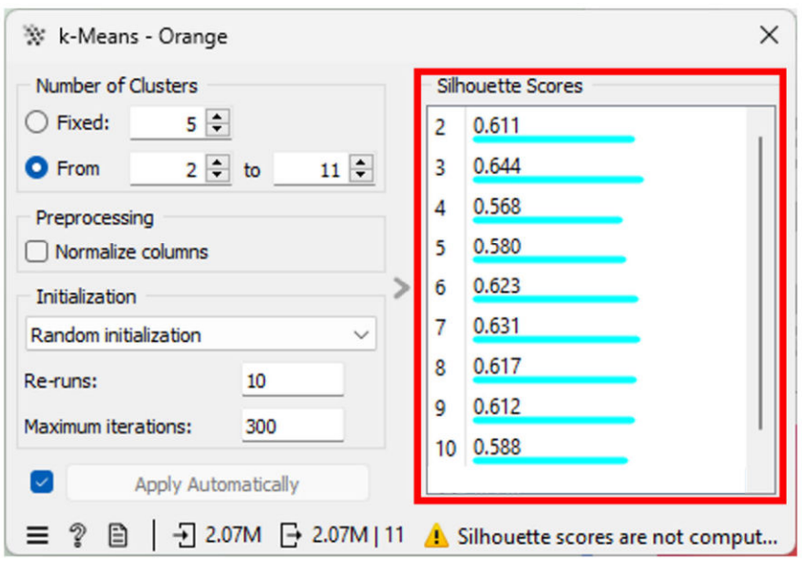
D. Petal width: 4.20 ; Petal length: 1.80

33. Suppose we want to determine the optimal number of clusters between 1 and 10 using Elbow Method. Which number of clusters should we choose?



A. 1
B. 2
C. 3
D. 4
E. 5
F. 6
G. 7
H. 8
I. 9
J. 10

34. Suppose we want to determine the optimal number of clusters between 2 and 10 using Silhouette Scores. Which number of clusters should we choose?



A. 2
B. 3
C. 4
D. 5
E. 6
F. 7
G. 8
H. 9
I. 10

35. What is the backbone of deep learning algorithms that enables the existence of ChatGPT and Stable Diffusion?

A. Linear Regression
B. Logistic Regression
C. K-means Clustering
D. Neural Networks

36. Which statement about neural networks is false?

A. Hidden layer nodes in a neural network must use non-linear activation functions.
B. A neural network has one input layer, one output layer, and can have multiple hidden layers.
C. Backpropagation is the process where errors are sent backward through the network to adjust weights.
D. A network where values flow forward and gradients flow backward is called a Recurrent Neural Network.

37. What are Generative Adversarial Networks not used for?

A. Data Augmentation
B. Image Synthesis
C. Object Detection
D. Style Transfer

38. What does the "T" in GPT stand for?

A. Technology
B. Text
C. Transformer
D. Troubleshoot