1. (True or false) Multimedia processing falls into the area of big data.

F/T

2. (True or false) Data mining only involves finding patterns in data that are already known and easily interpretable by humans.

F

3. (True or false) The sample space of an experiment represents all possible outcomes that can occur.

T

4. (True or false) K-means clustering is a supervised machine learning algorithm that assigns each observation to the cluster with the nearest mean.

F

5. (True or false) The parameter size will remain unchanged given a larger N in N-gram models because of the Markov assumption.F

6. (True or false) If a random variable X follows standard normal, its probability density function (PDF) is $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$. So we can infer the probability of observing $X = 0$ is $\frac{1}{\sqrt{2\pi}}$.

F

7. (True or false) The central limit theorem applies to any samples identically distributed, independent, and large enough, regardless of their population distribution.

F/T

8. (True or false) Sigmoid functions are commonly used for binary classification.

T

9. (True or false) In logistic regression, the parameters of the models can be seen as the weights over features.

T

10. (True or false) R allows the user to give an object a name that already exists, and R will not warn you when you use an existing name.

T

11. (Multi-choice) Which of the following is true about a random sample from a population?
    A) The sample consists of dependent random variables.
    B) The sample consists of identically distributed random variables.
    C) The outcomes of the experiment are fixed values.
    D) A statistic is a fixed value, not a random variable.

12. (Multi-choice) Suppose a bag contains 5 red balls and 7 blue balls. You randomly choose a ball from the bag, and without replacing it, you choose a second ball. What is the probability that the second ball is red, given that the first ball was blue?

    A) 5/12
    B) 1/3
    C) 5/11
    D) 2/3

13. (Multi-choice) Which of the following is a characteristic of data analytics?

    A. It relies solely on computer science to extract insights from data.
    B. It is not concerned with the amount of data used in the analysis.
    C. It involves the discovery of knowledge and information from data.
    D. It is an isolated field with no connection to other disciplines.

14. (Multi-choice) Which of the following BEST describes probabilistic language models?

    A) They are used to generate human-like responses in chatbots.
    B) They are based on a statistical analysis of large amounts of text data.
    C) They can only be trained on a small amount of data.

15. (Multi-choice) Suppose that a company produces two types of products, A and B. The probability of producing a defective product for type A is 0.1, and for type B is 0.15. The proportion of type A products produced is 0.6, and the proportion of type B products produced is 0.4. Given that a randomly selected product is defective, what is the conditional probability that the product is of type A?

    A) 0.32
    B) 0.40
    C) 0.50
    D) 0.60

16. (Multi-choice) Which of the following is the correct definition of the derivative of a function?
    A) The slope of the tangent line to the function at a specific point
    B) The area under the curve of the function between two points
    C) The average rate of change of the function over a specific interval
    D) The maximum value of the function over a specific interval

17. (Multi-choice) What is the value of the integral $\int (x^{\#} + 2x - 3)\, dx$ from $x = -1$ to $x = 2$?

    A) -3
    B) 0
    C) 3
    D) 6

18. (Multi-choice) Which symbol is used in R to represent missing values, and which symbol is used to represent impossible values?

    A) NaN represents missing values, and NA represents impossible values.
    B) NA represents missing values, and NaN represents impossible values.
    C) NA represents both missing and impossible values.
    D) NaN represents both missing and impossible values.

19. (Multi-choice) Which of the following statements regarding decision and loss functions in machine learning training is true?
    A) Decision functions make predictions, while loss functions measure the error between the predicted output and the true output.
    B) Decision functions measure the error between the predicted and true output, while loss functions make predictions.

C) Decision and loss functions are the same and are used interchangeably in machine learning training.
D) Decision and loss functions are unimportant in machine learning training.

20. (Multi-choice) Which of the following properties of vectors is NOT true?

A) Vector addition is commutative: $u + v = v + u$, where u and v are both vectors.
B) Vector multiplication by a scalar is distributive: $a(u + v) = au + av$, where u and v are both vectors and a is a scalar.
C) Vector multiplication by a scalar is associative: $a(bu) = (ab)u$, where u is a vector and a and b are both scalars.
D) The dot product of two vectors is always in the range of -1 and 1: $u'v \in [-1,1]$, where u and v are both vectors.

21. (Multi-answer) A researcher wants to test if the mean height of a population is 170 cm (null hypothesis H_0). The standard deviation of the population height is known to be 10 cm. He takes a random sample of 100 people and measures their heights. He finds that the sample mean is 172 cm. The standard normal distribution table is shown in the

following, where $\Phi(z)$ is the cumulative distribution function of the standard normal.

z	$\Phi(z)$	z	$\Phi(z)$
0.0	.5000	-1.2	.1151
-0.1	.4602	-1.4	.0808
-0.2	.4207	-1.6	.0548
-0.3	.3821	-1.8	.0359
-0.4	.3446	**-2.0**	**.0228**
-0.5	.3085	-2.2	.0139
-0.6	.2743	-2.4	.0082
-0.7	.2420	-2.6	.0047
-0.8	.2119	-2.8	.0026
-0.9	.1841	-3.0	.0013
-1.0	.1587	-3.2	.0007

Which of the following statements are correct? Select all that apply.

A) The researcher should reject $H_$ at the level of significance 0.01.
B) The researcher should reject $H_$ at the level of significance 0.03.
C) The researcher should reject $H_$ at the level of significance 0.06.
D) The researcher should reject $H_$ at the level of significance 0.09.

22. (Multi-answer) Which of the following statements are true regarding the properties of expected values and variance of discrete random variables? Select all that apply.

    A) The expected value of a constant is equal to the constant itself.
    B) The expected value of a sum of random variables is equal to the sum of their expected values.
    C) The variance of a constant is equal to zero.
    D) The variance of a sum of random variables is equal to the sum of their variances.

23. (Multi-answer) Which of the following statements about gradients are true? Select all that apply.

    A) Gradients are a vector quantity.
    B) Gradients 0 indicate that the corresponding point is a global optimal solution.
    C) Gradients can be used to optimize loss functions in machine learning.
    D) Gradients are closely related to derivatives.

24. (Multi-answer) Which of the following statements are true regarding vectors? Select all that apply.

    A) Vectors have length but no direction.
    B) Vectors can be added and subtracted using the head-to-tail method.

C) Vectors can be seen as the columns or rows of a matrix.
D) The dot product of two vectors of the same dimension always yields a scalar.

25. (Multi-answer) Which of the following statements about matrix multiplication are true? Select all that apply.

A) Matrix multiplication is not commutative in general, but it becomes commutative when one of the factors is an identity matrix, such that AB = BA.
B) The product of two matrices with dimensions m x n and p x q is a matrix with dimensions (m+n) x (p+q).
C) The product of two matrices is only defined if the number of columns in the first matrix is equal to the number of rows in the second matrix.

26. (Multi-answer) Which of the following are true regarding the data analysis process? Select all that apply.

A) The goal of data analysis is to discover useful information, inform conclusions, and support decision-making.
B) Data analysis involves only inspecting and cleaning data.
C) The modelling stage of data analysis is not important in discovering useful information.
D) The data analysis process involves transforming data to make it more useful.

27. (Multi-answer) Which of the following statements are true about cosine similarity of two data samples in vector representation? Select all that apply.

A) Cosine similarity is a measure of the angle between two vectors.
B) Cosine similarity is always between 0 and 1.
C) Cosine similarity is highly sensitive to norm of the two vectors.
D) Cosine similarity can reflect the data similarity.

28. (Multi-answer) Which of the following statements are true about using R? Select all that apply.

    A) You can only enter commands one at a time at the command prompt (>)
    B) You can run a set of commands from a source file
    C) R only supports numerical data types such as vectors
    D) R can support a wide variety of data types such as matrices, dataframes, and lists.

29. (Multi-answer) Which of the following statements are true about decision function, loss function, machine learning goal, and gradient descent? Select all that apply.
    A) The decision function can be used to map data samples to the classification labels.
    B) The loss function measures the accuracy of the model on the test data.
    C) The machine learning goal is to minimize the loss function on the training data.
    D) Gradient descent is an optimization algorithm used to find the optimal parameters of the model.

30. (Multi-answer) Which of the following statements are true regarding Naive Bayes classifier? Select all that apply.

A) Naive Bayes assumes that the features are conditionally independent given the class.
B) Naive Bayes can be considered as a linear classifier.
C) Naive Bayes is commonly used for unsupervised learning tasks.
D) Naive Bayes can possibly handle missing features (those present in the test set whereas absent in the training set).