1. Which Python structure is "unordered" and contains "unique" elements?

A. List

B. Tuple

C. Dictionary

D. Set

2. In "Hierarchical Clustering", the "Dendrogram" shows:

A. The density of points.

B. The sequence of cluster fusions and their dissimilarity.

C. The probability of class membership.

D. The gradient of the loss function.

3. "Oozie" is primarily used for:

A. Storing data.

B. Processing streams.

C. Scheduling workflows.

D. Querying data.

4. Which metric is used to measure the "impurity" in Decision Trees (though not explicitly named in text, concept is related to classification splits)?Question: "Accuracy" is defined as:

A. TP / Total

B. (TP + TN) / Total

C. TP / (TP + FP)

D. TP / (TP + FN)

5. "Cross-Entropy" quantifies:

A. Distance between points.

B. Difference between probability distributions.

C. Variance of a single distribution.

D. Correlation between variables.

6. "Bootstrapping" uses:

A. Sampling without replacement.

B. Sampling with replacement.

C. Stratified sampling.

D. Cluster sampling.

7. Which SQL clause "limits the range" of values (e.g. for dates)?

A. LIKE

B. IN

C. BETWEEN

D. DISTINCT

8. "Spark" uses which data structure for processing?

A. Tables

B. RDDs (Resilient Distributed Datasets)

C. JSON

D. Arrays

9. "Dropout" is a regularization technique that:

A. Adds a penalty to weights.

B. Removes random units during training.

C. Stops training early.

D. Normalizes the batch.

10. "HDFS" splits large files into:

A. Tables

B. Blocks

C. Rows

D. Objects

11. Which "Gradient Descent" type uses the "entire" training set for each step?

A. Stochastic

B. Mini-batch

C. Batch

D. Newton's

12. "Correlation" of -1 implies:

A. No relationship.

B. Perfect positive relationship.

C. Perfect negative relationship.

D. Nonlinear relationship.

13. "Mode" is the:

A. Average value.

B. Middle value.

C. Most frequent value.

D. Largest value.

14. "KNN" runtime complexity is:

A. O(n)

B. O(d)

C. O(nd)

D. O(1)

15. "Data Cleaning" issue "Apples to Oranges" refers to:

A. Missing values.

B. Data Compatibility.

C. Outliers.

D. Artifacts.

16. "Logistic Regression" output is transformed by:

A. ReLU

B. Sigmoid

C. Tanh

D. Step function

17. "F-Score" formula is:

A. 2 * (P*R)/(P+R)

B. (P+R)/2

C. P*R

D. P+R

18. "Type II Error" is:

A. False Positive

B. False Negative

C. True Positive

D. True Negative

19. "Power Law" distributions have:

A. Short tails.

B. Long tails.

C. No tails.

D. Bell shapes.

20. "MapReduce" Reduce step:

A. Splits data.

B. Combines results.

C. Filters data.

D. Sorts data.

21. Which of the following are "Python" structures?

(J) List

(K) Tuple

(L) Set

(M) Tree

Select the correct option:

A. J, K and L

B. All of them

C. J and K only

D. L and M only

a

22. Which of the following are "Hadoop" components?

(J) HDFS

(K) MapReduce

(L) YARN

(M) Pandas

Select the correct option:

A. J and K only

B. J, K and L

C. All of them

D. L and M only

b

23. Which of the following are "Metric" distances?

(J) Manhattan

(K) Euclidean

(L) Accuracy

(M) Precision

Select the correct option:

A. J and K only

B. L and M only

C. All of them

D. J, K and L

a

24. Which of the following are "Supervised" tasks?

(J) Classification

(K) Regression

(L) Clustering

(M) Dim Reduction

Select the correct option:

A. J and K only

B. L and M only

C. All of them

D. J and L only

a

25. Which of the following are "SQL" commands?

(J) SELECT

(K) INSERT

(L) UPDATE

(M) STOP

Select the correct option:

A. J, K and L

B. All of them

C. J and K only

D. L and M only

26. Match the Probability Term (Column-I) with Formula (Column-II):

Column-I	Column-II
(a) Independent	(I) P(A)P(B)
(b) Conditional	(II) P(A,B)/P(B)

(c) Bayes	(III) P(A|B)
(d) Joint	(IV) P(A,B)

Options:

A. a-I, b-II, c-III, d-IV

B. a-II, b-I, c-IV, d-III

C. a-III, b-IV, c-I, d-II

D. a-IV, b-III, c-II, d-I

27. Match the Data Type (Column-I) with Example (Column-II):

Column-I	**Column-II**
(a) Structured	(I) SQL Tables
(b) Unstructured	(II) Audio/Images
(c) Quantitative	(III) Height/Weight
(d) Categorical	(IV) Race/Color

Options:

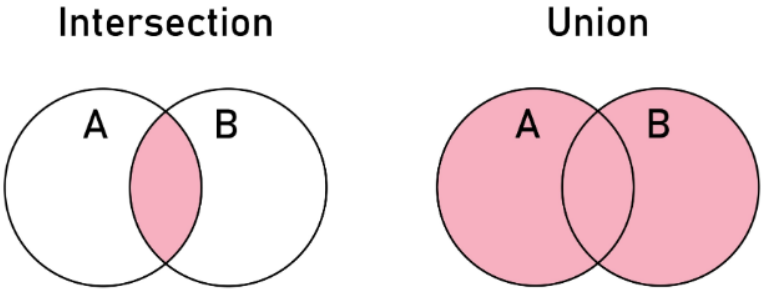
A. a-I, b-II, c-III, d-IV

B. a-II, b-III, c-IV, d-I

C. a-III, b-IV, c-I, d-II

D. a-IV, b-I, c-II, d-III

28. Match the SQL Join (Column-I) with Diagram Label (Column-II):



Column-I	Column-II
(a) Inner	(I) Intersection
(b) Left	(II) Left Circle + Intersection
(c) Right	(III) Right Circle + Intersection
(d) Full	(IV) Union of A and B

Options:

A. a-I, b-II, c-III, d-IV

B. a-II, b-I, c-IV, d-III

C. a-III, b-IV, c-I, d-II

D. a-IV, b-III, c-II, d-I

29. Match the Deep Learning (Column-I) with Concept (Column-II):

Column-I	Column-II
(a) CNN	(I) Images
(b) RNN	(II) Sequences

(c) GAN	(III) Fake/Real
(d) Dense	(IV) Connected

Options:

A. a-I, b-II, c-III, d-IV

B. a-II, b-III, c-IV, d-I

C. a-III, b-IV, c-I, d-II

D. a-IV, b-I, c-II, d-III

30. Assertion (A): "Precision" is useful when false positives are costly.

Reason (R): It measures the ratio of true positives to all predicted positives.

A. Both A and R are true, and R is the correct explanation of A

B. Both A and R are true, but R is NOT the correct explanation of A

C. A is true, but R is false

D. A is false, but R is true

31. Assertion (A): "HDFS" stores metadata on DataNodes.

Reason (R): "NameNode" manages the directory structure. (A is False).

A. Both A and R are true, and R is the correct explanation of A

B. Both A and R are true, but R is NOT the correct explanation of A

C. A is true, but R is false

D. A is false, but R is true

32. Assertion (A): "Linear Regression" uses Gradient Descent.

Reason (R): It minimizes the error function.

A. Both A and R are true, and R is the correct explanation of A

B. Both A and R are true, but R is NOT the correct explanation of A

C. A is true, but R is false

D. A is false, but R is true

33. Assertion (A): "Correlation" implies "Causation".

Reason (R): High correlation means one variable predicts the other. (A is False).

A. Both A and R are true, and R is the correct explanation of A

B. Both A and R are true, but R is NOT the correct explanation of A

C. A is true, but R is false

D. A is false, but R is true

34. Arrange "K-Means":

(J) Initialize.

(K) Assign.

(L) Update.

Options:

A. J - K - L

B. K - L - J

C. L - J - K

D. J - L - K

35. Arrange "MapReduce":

(J) Split.

(K) Map.

(L) Reduce.

Options:

A. J - K - L

B. K - L - J

C. L - K - J

D. J - L - K

36. Arrange "Data":

(J) Raw.

(K) Cleaned.

(L) Features.

Options:

A. J - K - L

B. K - L - J

C. L - J - K

D. J - L - K

37. Arrange "Modeling":

(J) Train.

(K) Test.

(L) Deploy.

Options:

A. J - K - L

B. K - L - J

C. L - J - K

D. J - L - K

38. Arrange "SQL":

(J) SELECT

(K) FROM

(L) LIMIT

Options:

A. K - J - L

B. J - K - L

C. L - K - J

D. K - L - J

39. Arrange "Gradient Descent":

(J) Calc Grad.

(K) Update.

(L) Repeat.

Options:

A. J - K - L

B. K - L - J

C. L - J - K

D. J - L - K

40. Arrange "Tree Build":

(J) Root.

(K) Branch.

(L) Leaf.

Options:

A. J - K - L

B. K - L - J

C. L - J - K

D. J - L - K

Answer Key

1.D 2.B 3.C 4.B 5.B 6.B 7.C 8.B 9.B 10.B 11.C 12.C 13.C 14.C 15.B 16.B 17.A 18.B 19.B 20.B 21.A 22.B 23.A 24.A 25.A 26.A 27.A 28.A 29.A 30.A 31.D 32.A 33.D 34.A 35.A 36.A 37.A 38.A 39.A 40.A