SUBJECT::- DATA SCIENCE   TYBSc CS
This is a Sample Question Bank.

R.No	Questions	A	B	C	D	Correct Answer
1	When the sampling is driven by the subjective judgment of the expert, it's called _____.	quota sampling	purposive sampling	convenience sampling	random sampling	purposive sampling
2	The most likely or frequently occurring value is called as_____.	Mean	Median	Mode	Skewness	Mode
3	Data published in books , magzines, newspapers and journals is called as _____.	Primary data	Secondary data	Unknown data	Nominal data	Secondary data
4	_____ is a Hadoop environment non-relational database system use for distributed computing.	HBase	MapReduce	CouchDB	FlockDB	HBase
5	_____is the on-going management of data through its lifecycle.	Data Curation	Data Transformation	Data Visualization	Data wrangling	Data Curation
6	In data curation, _____is the action that evaluates and selects data for long-term preservation.	Ingest	Transform	Appraise	Curate	Appraise
7	_____ type of decision tree when the predicted outcome is the class to which the data belongs.	Classification Tree	Random Forest	Regression Tree	SVM	Classification Tree
8	The distance between nearest data points of each class, and the hyperplane is called_____.	Plane	Margin	Graph	Chart	Margin
9	Overfitting models have _____ bias and _____ variance.	low, high	high,low	low,low	high,high	low, high
10	L1 Regularization is also called _____.	Ridge regression	Lasso regression	Logistic regression	Classification	Lasso regression
11	The second central moment is known as _____.	variance	mean	median	Skewness	variance

#	Question					
12	_____ is a measure of peakedness and tailedness of the probability distribution of a random variable.	skewness	covariance	variance	kurtosis	kurtosis
13	_____ package that reshapes the layout of tables.	tidyr	shape	dplyr	reshape	tidyr
14	The_____ is defined as the middle number between the smallest number and the median of the data set.	First quartile	Second quartile	Third quartile	q-quartile	First quartile
15	_____ is a Occasional action in Data Curation.	Migrate	Curate	Transform	Ingest	Migrate
16	_____ is the process of converting data from one format to another.	Data Curation	Data Transformation	Data Visualization	Data wrangling	Data Transformation
17	XQuery is a functional query language used to retrieve information stored in _____format.	HTML	XML	UML	XSL	XML
18	XPath specification has _____ types of nodes	Four	Five	Six	Seven	Seven
19	_____ is Unsupervised Learning Technique.	Clustering	SVM	Linear Regression	Logistic Regression	Clustering
20	_____ is a collection of observations obtained through repeated measurements over time.	Sample	Time series data	Probability data	Random data	Time series data
21	When the model tries to learn or fit every noise data point in the data set, it results in _____.	Overfitting	Underfitting	Smoothing	Aggregation	Overfitting
22	Qualitative data is also called as _____data.	Continuous	Numerical	Categorical	Discrete	Categorical
23	_____ measures asymmetry about the mean of the probability distribution of a random variable.	skewness	covariance	variance	kurtosis	skewness
24	_____ measures the degree of the relationship between two random va	Standard Deviation	Skewness	Correlation	kurtosis	Correlation
25	_____ is column-oriented database system.	HBase	MapReduce	CouchDB	FlockDB	HBase

munotes.in

#	Question	A	B	C	D	Answer
26	A Collection in MongoDB is composed of _____.	Tables	Documents	Rows	Columns	Documents
27	Amazon _____ is a web service interface that provides resizable compute capacity in the AWS cloud.	EC1	EC2	EC3	EC4	EC2
28	A supervised machine learning algorithm should aim to have _____ bias and _____ variance.	low, high	high,low	low,low	high,high	low,low
29	A good machine learning model is the one that has _____AIC among all the models.	Constant	Minimum	Average	Maximum	Minimum
30	L2 Regularization is also called _____.	Ridge regression	Lasso regression	Logistic regression	Classification	Ridge regression