

## Sample Questions Semester V Machine Learning

Sr. No.	Question	Module No.	Option a	Option b	Option c	Option d
1	. Regarding bias and variance, which of the following statements are true? (Here 'high' and 'low' are relative to the ideal model.)	1	Models which overfit have a high bias.	Models which overfit have a low bias.	Models which underfit have a high variance.	Models which overfit have a low variance.
2	What is Machine learning?	1	The autonomous acquisition of knowledge through the use of computer programs	The autonomous acquisition of knowledge through the use of manual programs	The selective acquisition of knowledge through the use of computer programs	The selective acquisition of knowledge through the use of manual programs
3	ML is a field of AI consisting of learning algorithms that?	1	Improve their performance	At executing some task	Over time with experience	Improve their performance (P), At executing some task (T), Over time with experience (E).
4	Teaching a machine on the basis of a policy is called as	1	unsupervised learning	supervised learning	reinforcement learning	dimension reduction
5	The set of all legal hypothesis is called as	1	set of hypothesis	hypothesis space	superset of hypothesis	hypothesis
6	Below are the 8 actual values of target variable in the train file. [0,0,0,1,1,1,1,1] What is the entropy of the target variable?	2	$-(5/8 \log(5/8) + 3/8 \log(3/8))$	$5/8 \log(5/8) + 3/8 \log(3/8)$	$3/8 \log(5/8) + 5/8 \log(3/8)$	$5/8 \log(3/8) - 3/8 \log(5/8)$

7	<p>Imagine, you are solving a classification problems with highly imbalanced class. The majority class is observed 99% of times in the training data.</p> <p>Your model has 99% accuracy after taking the predictions on test data. Which of the following is true in such a case?</p> <p>1.Accuracy metric is not a good idea for imbalanced class problems.  2.Accuracy metric is a good idea for imbalanced class problems.  3.Precision and recall metrics are good for imbalanced class problems.  4.Precision and recall metrics aren't good for imbalanced class problems.</p>	2	1 and 3	1 and 4	2 and 3	2 and 4
8	Which of the following is NOT supervised learning?	2	PCA	Decision Tree	Linear Regression	Naive Bayesian
9	Which of the following statements about Naive Bayes is incorrect?	2	Attributes are statistically dependent of one another given the class value.	Attributes are statistically independent of one another given the class value.	Attributes are equally important.	Attributes can be nominal or numeric
10	Predicting the amount of rainfall in a region based on various features is a _____ problem.	2	Supervised learning	Unsupervised learning	Clustering	PCA
11	Which of the following statement is true about outliers in Linear regression?	3	Linear regression is not sensitive to outliers	Linear regression is sensitive to outliers	linear regression eliminates outliers	Linear regression accepts outliers
12	A multiple regression model has the form: $y = 2 + 3x_1 + 4x_2$ . As $x_1$ increases by 1 unit (holding $x_2$ constant), $y$ will	3	increase by 3 units	decrease by 3 units	increase by 4 units	decrease by 4 units

13	A measure of goodness of fit for the estimated regression equation is the	3	A measure of goodness of fit for the estimated regression equation is the	mean square due to error	mean square due to regression	root mean squared error
14	Another name for an output attribute.	3	predictive variable	independent variable	estimated variable	dependent variable
15	Simple regression assumes a _____ relationship between the input attribute and output attribute.	3	quadratic	linear	inverse	reciprocal
16	Logistic regression is a _____ regression technique that is used to model data having a _____ outcome.	3	linear, numeric	linear, binary	nonlinear, numeric	nonlinear, binary
17	Which of the following is/are true regarding an SVM?	4	For two dimensional data points, the separating hyperplane learnt by a linear SVM will be a straight line.	In theory, a Gaussian kernel SVM cannot model any complex separating hyperplane.	For every kernel function used in a SVM, one can obtain an equivalent closed form basis expansion.	Overfitting in an SVM is not a function of number of support vectors.
18	Which of the following can only be used when training data are linearly separable?	4	Linear hard-margin SVM.	Linear Logistic Regression.	Linear Soft margin SVM.	The centroid method.
19	A line that linearly separates and classifies a set of data is called _____.	4	support vector	best fit line	least fit line	margin
20	“Support Vector Machine” is _____ algorithm	4	unsupervised machine learning	supervised machine learning	recommender system	prediction
21		4				
22	In adaboost the Weights of the correctly classified data are _____	5	decreased	increased	nochange	sometimes increased, sometimes decreased
23		5				

	In ensemble learning, you aggregate the predictions for weak learners, so that an ensemble of these models will give a better prediction than prediction of individual models.					
24	Which of the following statements is / are true for weak learners used in ensemble model?  1.They don't usually overfit. 2.They have high bias, so they cannot solve complex learning problems 3. They usually overfit.	5	1 and 2	1 and 3	2 and 3	Only 1
25	Bootstrapping allows us to	5	choose the same training instance several times.	choose the same test set instance several times.	build models with alternative subsets of the training data several times.	test a model with alternative subsets of the test data several times.
26	Which of the following can be true for selecting base learners for an ensemble? 1. Different learners can come from same algorithm with different hyper parameters 2. Different learners can come from different algorithms 3. Different learners can come from different training spaces	5	1	2	1 and 3	1,2 and 3
27		6				
28	In which of the following cases will K-means clustering fail to give good results? 1) Data points with outliers 2) Data points with different densities 3) Data points with nonconvex shapes	6	1 and 2	2 and 3	1, 2, and 3	1 and 3

29	Suppose we would like to perform clustering on spatial data such as the geometrical locations of houses. We wish to produce clusters of many different sizes and shapes. Which of the following methods is the most appropriate?	6	Decision Trees	Density-based clustering	Model-based clustering	K-means clustering
30	This clustering algorithm initially assumes that each data instance represents a single cluster.		agglomerative clustering	conceptual clustering	K-Means clustering	expectation maximization
31	This unsupervised clustering algorithm terminates when mean values computed for the current iteration of the algorithm are identical to the computed mean values for the previous iteration.	6	agglomerative clustering	conceptual clustering	K-Means clustering	expectation maximization
32	What is PCA components	7	Set of all eigen vectors for the projection space	Matrix of principal components	Result of the multiplication matrix	result of inverse of the matrix
33	When performing regression or classification, which of the following is the correct way to preprocess the data?	7	PCA → normalize PCA output	Normalize the data → PCA → training	Normalize the data → PCA → normalize PCA output	PCA--> normalize the data--> normalize the output
34	Anomaly detection is also referred to as	7	supervised learning	outlier detection	unsupervised learning	dimension reduction
35	A recommender system, seeks to predict _____ a user would give to an item.	7	class	preference	numerical value	weights
36	Dimensionality reduction, transforms a ____ into a _____ retaining the meaningful properties of the original data	7	high-dimensional space, low-dimensional space	high-dimensional space, high-dimensional space	low-dimensional space, high-dimensional space	low-dimensional space, low-dimensional space