

Shrimad Rajchandra Institute of Management and Computer Application
Uka Tarsadia University
MCA Semester3
AI8001 – Machine Learning
Question List

Unit-1

=====

1.	What is machine learning? Draw the diagram and discuss relationship among artificial intelligence, machine learning, and deep learning.																								
2.	List any four applications of machine learning and discuss how machine learning is useful in any two of the listed applications.																								
3.	With appropriate example, explain supervised machine learning.																								
4.	What is unsupervised learning? Explain with suitable example.																								
5.	Compare supervised versus unsupervised machine learning. Discuss Similarities and differences.																								
6.	What are the objectives of training and testing dataset? Explain cross validation with appropriate diagram.																								
7.	Explain accuracy and recall as performance measures used to evaluate machine learning algorithms.																								
8.	What is data preprocessing in machine learning? Use any two suitable examples and explain encoding with respect to data preprocessing.																								
9.	What are inconsistent values and duplicate values with respect to quality of data? Write two examples of each. Also discuss solution to handle inconsistent, duplicate values to improve data.																								
10.	Define data preprocessing. Explain missing values and out-of-range values with respect to quality of data.																								
11.	What are categorical and numerical data? Give suitable example and explain nominal, ordinal, interval, and ratio.																								
12.	<p>What is categorical data? Which are the categorical data in following example? Explain, how encoding is done using LabelEncoder class in this example.</p> <table><tr><th>City</th><th>Age</th><th>Salary</th><th>Purchased</th></tr><tr><td>Bardoli</td><td>45</td><td>151000</td><td>Yes</td></tr><tr><td>Surat</td><td>51</td><td>164000</td><td>Yes</td></tr><tr><td>Valsad</td><td>42</td><td>134000</td><td>No</td></tr><tr><td>Surat</td><td>24</td><td>89000</td><td>Yes</td></tr><tr><td>Valsad</td><td>36</td><td></td><td>No</td></tr></table>	City	Age	Salary	Purchased	Bardoli	45	151000	Yes	Surat	51	164000	Yes	Valsad	42	134000	No	Surat	24	89000	Yes	Valsad	36		No
City	Age	Salary	Purchased																						
Bardoli	45	151000	Yes																						
Surat	51	164000	Yes																						
Valsad	42	134000	No																						
Surat	24	89000	Yes																						
Valsad	36		No																						
13.	Compare LabelEncoder with OneHotEncoder for encoding categorical data. Discuss one advantage and one limitation of OneHotEncoder.																								
14.	Explain feature scaling with respect to data preprocessing.																								
15.	Write and explain python code snippet to create and print a dataframe for Employee_Salary dataset.																								
16.																									
17.																									

2025 Question List of AI8001 - Machine Learning

18.	
19.	
20.	

Unit-2

=====

1.	Draw the diagram and explain simple linear regression.
2.	What is regression? Explain best fitting line in the context of simple linear regression.
3.	Explain multiple linear regression with appropriate example.
4.	Draw flow chart and explain backward elimination method for multiple linear regression.
5.	What is multiple linearity and dummy variable trap? Explain backward elimination methods with respect to multiple linear regression.
6.	In $Y = b_0 + b_1x$, what is the importance of b_0 and b_1 ? What is best fitting line? Discuss using graph.
7.	List out steps for bidirectional elimination method for building a multiple linear regression model. If we will consider all possible multiple linear regression models, how many models are possible for dataset having 10 columns?
8.	Explain backward elimination and forward selection method for building a multiple linear regression model.
9.	Below are the 10 records for column state. In multiple linear regression, how many dummy variables one need to add in this example? Explain. (State: Gujarat, Maharashtra, Bihar, Goa, Gujarat, Goa, Maharashtra, Gujarat, Goa, Bihar)
10.	If p value of X variable is 0.06 and your significance level is 5%, will you keep X variable in model? Why?
11.	How to select only statistically significant independent variables in multiple linear regression? Explain step-by-step.
12.	Why one need to remove some of the independent variables while creating model? Write down five methods for building a model in multiple linear regression and explain any one in detail.
13.	How can we add all dummy variables in multiple linear regression equations? Write usage of it.
14.	Write the situation where polynomial regression is useful. Explain polynomial regression with appropriate example.
15.	What is linear regression? When one need to use polynomial regression? Why it is called linear?
16.	With suitable example and python code snippet, explain decision tree algorithm.
17.	In which situations, decision tree regression model is best suited and why? Explain any three advantages of decision tree algorithm.
18.	Random forest uses boosting or bagging technique? Explain bagging with respect to ensemble learning.
19.	Draw the diagram and compare bagging and boosting techniques.
20.	Explain any five advantages of random forest algorithm.
21.	Draw flow chart and explain random forest regression algorithm.
22.	Draw the diagram and explain the creation of decision tree regression model.

23.	What is ensemble learning? Explain any four limitations of random forest algorithm.
24.	
25.	
26.	
27.	
28.	

Unit-3

=====

1.	What is classification? Is it supervised or unsupervised machine learning? Explain binary and multi-class classification with suitable examples.
2.	Can we use binary classification algorithms for multi-class classification? Discuss.
3.	Draw the diagram and explain One-vs-Rest and One-vs-One with respect to multi-class classification.
4.	Explain multi-label classification and imbalanced classification with suitable examples.
5.	Write and explain python code snippet for logistic regression.
6.	Draw the diagrams and compare linear regression and logistic regression.
7.	Write and explain two assumptions used in logistic regression algorithm. Also explain any one parameter of LogisticRegression class (sklearn.linear_model).
8.	K-Nearest Neighbors is a supervised or unsupervised algorithm? Why it is called non-parametric and lazy learner algorithm? Discuss.
9.	Using sample data, explain the impact of very low and very high value for parameter K in K-Nearest Neighbors algorithm. How to select value of parameter K? Discuss.
10.	Develop a classification template using python code for K-NN algorithm.
11.	What is Euclidian distance? Write down and discuss steps of KNN algorithm.
12.	What is kernel function in Support Vector Machine? List any three kernel functions and write mathematical formula for each.
13.	Support Vector Machine is supervised or unsupervised machine learning algorithm? Draw the diagram and explain hyperplane, margin, and support vectors.
14.	Write python code snippet and explain Support Vector Machine as a classifier.
15.	Explain Support vector machine algorithm with a proper graphical representation and mathematical notations.
16.	What are support vectors? How SVM is different from other classifiers? Discuss with suitable example.
17.	How support vector machine works? Explain.
18.	Write and explain python code snippet for Naive Bayes classifier.
19.	When to use naive Bayes classifier? Discuss with appropriate examples.
20.	What is the role of conditional probability in Naive Bayes? Discuss using proper example.
21.	Assume that you have the following data set. Calculate and find the individual conditional probabilities for each feature and calculate the final probability of whether a match shall be scheduled to play or not using Naïve Bayes.

	Outlook	Temperature	
	Sunny	Hot	
	Rainy	Mild	
	Overcast	Cold	
	Rainy	Cold	
	Sunny	Hot	
	Overcast	Mild	
	Rainy	Mild	
22.			
23.			
24.			
25.			
26.			
27.			
28.			

Unit-4

=====

1.	What is unsupervised machine learning? What are clustering and association? List two algorithms for each.
2.	Explain working of clustering and association as unsupervised machine learning.
3.	List four clustering categories and explain any one.
4.	Explain density based and hierarchical based clustering.
5.	Compare partitioning based and grid-based clustering. Explain both methods.
6.	Draw the flow chart and explain K-Means clustering.
7.	What is clustering? Discuss different convergence criteria with respect to K-Means clustering.
8.	Draw diagrams and explain how K-Means clustering algorithm is sensitive to outliers and initial seeds.
9.	Discuss any two methods to choose value of K in K-Means clustering algorithm.
10.	Write python code snippet and explain elbow method in K-Means clustering algorithm. Draw suitable diagram.
11.	Define clustering. Explain any four limitations of K-Means clustering algorithm.
12.	What are agglomerative and divisive hierarchical clustering? Write steps for divisive hierarchical clustering.
13.	Draw the flow chart and step-by-step explain working of agglomerative hierarchical clustering.
14.	Write python code and explain agglomerative hierarchical clustering.
15.	What is dendrogram? Explain how to select number of clusters in hierarchical clustering using dendrogram.
16.	In hierarchical clustering, how to identify closest clusters? Draw the diagrams and explain any two linkage methods.
17.	With diagram explain, minimum linkage clustering and centroid linkage clustering.
18.	Explain maximum linkage clustering and average linkage clustering with suitable diagram.
19.	Explain support and confidence metrics with respect to measure association.
20.	Step-by-step explain Apriori algorithm.
21.	Write basic assumptions of Apriori algorithm. Explain any four applications of Apriori algorithm.
22.	
23.	
24.	
25.	
26.	
27.	
28.	

29.	
30.	
31.	
32.	

Unit-5

=====

1.	What is reinforcement learning? Explain agent, environment, action, state, reward with suitable diagram.
2.	Draw the diagram and explain reinforcement learning.
3.	Define exploration and exploitation. Which one is better compare to other with respect to reinforcement learning? Discuss.
4.	With example, explain exploration versus exploitation with respect to reinforcement learning.
5.	Explain multi-armed bandit problem with suitable example.
6.	Your marketing department has received five different advertisements for branding of the new product. Discuss impact of “no exploration”, “exploration at random”, and “exploration smartly with preference to uncertainty”.
7.	Explain working of upper confidence bound algorithm with appropriate diagram.
8.	Write and explain python code snippet to implement upper confidence bound.
9.	Write mathematical formula of upper confidence bound algorithm and explain each term of it.
10.	What is the use of upper confidence bound algorithm? Explain following formula: $A_t = \underset{a}{\operatorname{argmax}} (Q_t(a) + \sqrt{\frac{2 \ln t}{N_a(t)}})$
11.	How to solve multi-armed bandit problem with Thompson sampling? Discuss.
12.	Explain Thompson sampling algorithm with appropriate python code snippet.
13.	What is a Artificial Neural Network (ANN)? Which types of problems are suitable to solve using ANN? Discuss.
14.	Explain Hidden Layer with suitable example.
15.	With a suitable example explain back propagation in Artificial Neural Network.
16.	Explain Artificial Neural Network with the help of appropriate diagram.
17.	Which activation functions are commonly applied in hidden layer and output layer? Discuss them in detail.
18.	What is back propagation? What are the advantages of it? List out steps for training ANN.
19.	What is activation function? List out any four activation. If dependent variables having value 0 or 1, which activation function is more suitable? Why?
20.	What is neuron? Draw the structure of single neuron. Discuss in detail.
21.	What is pooling and why one need it? Discuss pooling with the help of example feature map.
22.	Draw the basic architecture of CNN and explain it in detail.
23.	List different types of pooling methods. What is the importance of

	pooling layer? How it works?
24.	

Unit-6

=====

1.	Define dimensionality reduction. Draw the diagram and discuss curse of dimensionality. Is dimensionality reduction good or bad? Discuss.
2.	What is dimensionality reduction? Explain any four advantages of dimensionality reduction.
3.	What is dimensionality reduction? Why it is needed? List out any three dimensionality reduction techniques and explain briefly.
4.	Discuss dimensionality reduction in detail with suitable example.
5.	Draw the diagram and explain feature selection and feature extraction. Write two advantages of each.
6.	Explain dimensionality, orthogonal, eigenvectors, covariance matrix, and correlation with respect to Principal Component Analysis (PCA).
7.	Write and explain python code for implementing PCA.
8.	Principal Component Analysis (PCA) is supervised or unsupervised technique? What are the advantages of PCA? Discuss.
9.	Write two LDA assumptions about data and explain working of Linear Discriminant Analysis.
10.	In which situation Linear Discriminant Analysis is best suited? Discuss with suitable example.
11.	Write python script for implementing Linear Discriminant Analysis and discuss it.
12.	Linear Discriminant Analysis (LDA) is supervised or unsupervised technique? Explain two advantages and two limitations of Linear Discriminant Analysis.
13.	Draw the diagrams and explain working of PCA and LDA. Discuss when one algorithm is more suitable than other.
14.	When to use Kernel PCA? Explain working of it.
15.	List any four kernel functions used in Kernel PCA and explain any two kernel functions.
16.	Write and explain python code for implementing Kernel PCA.
17.	Kernel PCA is supervised or unsupervised technique? Explain two advantages and two limitations of Kernel PCA.
18.	Discuss advantage and limitations of LDA compare to kernel PCA.
19.	
20.	