

Scheme	R2012
Semester	VIII
Course code	CPE8031
Course Name	Machine Learning

Question No.	Question	a	b	c	d
1	Machine Learning comes under which of the following domain?	Artificial Intelligence	Network Security	Engineering sciences	System programming
2	Choose the correct tree based learner.	Rule based	Hidden markov model	Bayesian classifier	CART
3	Support Vector Machine(SVM) can be used for both classification or regression challenges. Which kind of learning technique SVM uses?	supervised	unsupervised	reinforced	clustered
4	Support Vector Machine(SVM) performs well in _____ dimension spaces.	high	low	wide	single
5	What is the approach of basic algorithm for decision tree induction?	Greedy	Top Down	Procedural	Step by Step
6	Principal component analysis(PCA) is used for _____.	Dimensionality Enhancement	LU Decomposition	QR Decomposition	Dimensionality Reduction
7	Various _____ methods and techniques are used for calculation of the outliers.	distance calculation	prediction	optimization	integration
8	What is true about the discount factor in reinforcement learning?	discount factor should be greater than 1	discount factor should always be negative	discount factor should be in range of 0 and 1	discount factor can be any real number
9	Choose correct applications of reinforcement learning?	Aircraft Control	Sentimental analysis	House price prediction	Spam Email Filtering
10	Which of the following problems can be solved by supervised learning too? Assume appropriate dataset is available.	From a large collection of spam emails, discover if there are sub types of spam emails.	Given data on how 1000 medical patients respond to an experimental medicine, discover whether there are different categories of patients in terms of how they respond to, and if so what are these categories	Given a large dataset of medical records from patients suffering from heart disease, try to learn whether there might be different groups of patients for which customised treatment is required	Given genetic (DNA) data from a person, predict the odds of the person developing diabetes over the next 10 years
11	Which of the following is a clustering algorithm in machine learning?	Expectation Maximization	CART	Gaussian Naïve Bayes	Apriori
12	Consider a point that is correctly classified and distant from the decision boundary. Which of the following methods will be unaffected by this point?	Nearest neighbor	Support Vector Machine	Logistic regression	Linear regression
13	K-fold cross-validation is _____.	linear in K	quadratic in K	cubic in K	exponential in K
14	What are support vectors?	These are the datapoints which help the SVM to generate optimal hyperplane.	It is an intermediate vector generated during calculation of optimal hyperplane	In SVM all the data points are called support vectors.	This are predefined vectors used in calculating hyperplane
15	Why SVM's are more accurate than logistic regression?	SVM gives more weightage to wrongly classified data points.	SVM gives more weightages to data points which are correctly classified.	SVM uses all the data points assuming a probabilistic model.	SVM uses concept of large margin separator and for non linearity it uses kernel functions
16	While comparing reinforcement learning and supervised learning, which of the following statement is true?	Both in reinforcement and supervised learning decisions are taken sequentially	Supervised learning is best suited where human interaction is prevalent whereas reinforcement learning is best suited for software systems.	Reinforcement learning works by interacting with environment whereas supervised learning works on sample data	Both in reinforcement and supervised learning decisions taken at one time step is independent with respect to previous timestep.
17	You ran gradient descent for 20 iterations with learning rate=0.2 and compute cost for each iteration. You observe that cost decreases after each iteration. Based on this which conclusion is more suitable.	Try smaller values of learning rate like 0.01	0.2 is effective choice of learning rate.	Try larger values of learning rate like 1.	Try any negative value for learning rate
18	You are training an RBF SVM with the following parameters: C (slack penalty) and $\gamma = 1/2\sigma^2$ (where σ^2 is the variance of the RBF kernel). How should you tweak the parameters to reduce overfitting?	Increase C and/or reduce γ	Reduce C and/or increase γ	Reduce C and/or reduce γ	Reduce C only (γ has no predictable effect on overfitting)

19	For a trained logistic classifier given a sample x , it gives prediction as 0.8. This means that ____.	$P(Y=0 x)=0.8$	$P(Y=1 x)=0.8$	$P(Y=0 x)=0.2$	$P(Y=1 x)=0.2$
20	What are two steps of pruning in decision tree ?	Pessimistic pruning and Optimistic pruning	Postpruning and Prepruning	Cost complexity pruning and time complexity pruning	High pruning and low pruning
21	Which algorithm is used for performing probabilistic reasoning on temporal data?	Hill-climbing search	Hidden markov model	Naïve Method	Support Vector Machine
22	In Logistic regression predicted value of Y lies within ____ range.	0 to 1	0 to $-\infty$	$-\infty$ to $+\infty$	-1 to 1
23	In principal component analysis ,if eigenvalues are equal.What does it mean?	PCA will perform outstandingly	PCA will perform badly	No effect	Model will be unstable
24	_____ phenomenon refers that a model is neither trained on training data nor generalized properly on new data.	good fitting	overfitting	moderate fitting	underfitting
25	Choose the reason for pruning a Decision Tree?	To save computing time during testing	To avoid overfitting the training set	To save space for storing the Decision Tree	To make the training set error smaller