

Name:

IDNO:

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
 Department of Computer Science and Information Systems
I SEMESTER 2017-2018
BITS F464 – Machine Learning

08th December, 2017

Comprehensive Examination

Weightage: 40%

- PART A (Closed Book) - Weightage 25%
- 20 Multiple Choice Questions carry 0.5 mark each. (10 marks).
- Mark the most appropriate choice in the grid below in CAPITAL LETTERS
- Maximum Time allowed for Part A: 90 Minutes

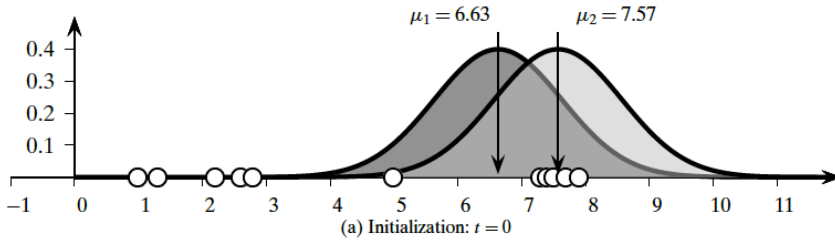
Q.→	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
Ans. →																				

SET A

1. Uncertainty sampling, which of the following takes into account the entire class distribution?
 - (a) Least confident
 - (b) Margin sampling
 - (c) Entropy
 - (d) None of the above consider the entire class distribution
2. A greedy active learning scenario:
 - (a) Query-member synthesis
 - (b) Pool-based sampling
 - (c) Stream-based sampling
 - (d) None of the above is recommended
3. Query by committee (QBC):
 - (a) Maximizes the version space so that we have large number of data points to select for labeling
 - (b) Minimizes the version space so that we have a small number of data points to select for labeling
 - (c) Does not alter the version space
 - (d) Defines version space
4. Version space is a region of:
 - (a) Certainty
 - (b) Uncertainty
 - (c) Unlabeled data points
 - (d) Labelled data points
5. Recommended active learning scenario when memory and/or processing power is limited and you are dealing with a very large corpus:
 - (a) Query-member synthesis
 - (b) Pool-based sampling
 - (c) Stream-based sampling
 - (d) None of the above is recommended
6. Query member synthesis is most suited for:
 - (a) Speech recognition
 - (b) NLP
 - (c) Handwriting recognition
 - (d) Not suited for any of the above applications
7. In reinforcement learning, an element of exploration can be introduced by:
 - (a) Greedy and e-greedy
 - (b) Greedy and softmax
 - (c) E-greedy and softmax
 - (d) Greedy, e-greedy, and softmax
8. Main disadvantage of Q-learning algorithm:
 - (a) Complex and difficult to implement
 - (b) Does not work for continual tasks
 - (c) Memory intensive
 - (d) Focusses on maximizing intermediate reward
9. In EM algorithm:
 - (a) Every data point is assigned to exactly only cluster
 - (b) Every data point is probabilistically assigned to multiple clusters with equal probability
 - (c) Works only for mixture of Gaussians
 - (d) Every data point is probabilistically assigned to multiple clusters

10. The concept of soft-margin SVM makes sense for:
 - (a) Both Linear and Non-linear SVM
 - (b) Neither Linear nor Non-linear SVM
 - (c) Only to Non-linear SVM
 - (d) Linear SVM but not to Non-linear SVM
11. In k-fold cross validation:
 - (a) It depends on the value of k, whether all data points will participate in both training and testing
 - (b) Irrespective of the value of k, all data points participate in both training and testing
 - (c) Only one, randomly chosen, partition participates in testing
 - (d) All partitions participate only in training
12. Increasing order of complexity:
 - (a) Perceptron, McCullach and Pitts model, ANN
 - (b) ANN, McCullach and Pitts model, Perceptron,
 - (c) McCullach and Pitts model, Perceptron, ANN
 - (d) All models can be made as complex as we want
13. Kernel functions work in:
 - (a) Both input space and feature space
 - (b) Only in feature space
 - (c) Only in input space
 - (d) Neither in input space nor in feature space
14. Correct statement about kernel functions used in SVM:
 - (a) They uniquely determine the feature space
 - (b) There are finite number of kernel functions
 - (c) Kernels must satisfy Mercer's theorem
 - (d) Kernels must satisfy KKT conditions
15. PCA & FLD:
 - (a) Both are unsupervised learning algorithms
 - (b) Both are supervised learning algorithms
 - (c) PCA is supervised whereas, FLD is unsupervised
 - (d) FLD is supervised whereas, PCA is unsupervised
16. Most appropriate statement about boosting:
 - (a) It focusses on difficult to classify examples
 - (b) It gives more weightage to good classifiers
 - (c) It uses majority voting
 - (d) It focusses on difficult to classify examples and gives more weightage to good classifiers
17. Correct statement about decision tree classifiers:
 - (a) Incapable of finding linear decision boundaries
 - (b) Capable of finding any kind of non-linear boundary
 - (c) Susceptible to overfitting, but not underfitting
 - (d) Susceptible to both overfitting and underfitting
18. Problems associated with CoD:
 - (a) Exponentially increasing time complexity of Machine Learning algorithms
 - (b) Concentration effect of L_p -norm
 - (c) Data sparsity
 - (d) All the above
19. The perceptron classifier for binary classification problems:
 - (a) Will always converge
 - (b) Will always have some training error
 - (c) Convergence not guaranteed for non-linearly separable data
 - (d) Convergence depends on the order in which data points are processed
20. Criticism of Bayesian approach to probability in context of Machine Learning:
 - (a) Subjective priors
 - (b) Scalability issues
 - (c) Non-informative priors
 - (d) (a) and (b)

1. The figure below is the initialization of the EM algorithm for a set of 1D points. Pictorially illustrate the converged solution and write the approximate mean and variance of the final distribution. Initial variances for both distributions is 1 and initial means are marked in the figure.



$\mu_1 =$
 $\sigma_1^2 =$

$\mu_2 =$
 $\sigma_2^2 =$

2. Mention two conditions (on $J(\mathbf{w})$) for which FLD will fail.

3. As a caring teacher, I want to know whether or not you are actually working towards the exam. I know that there are four things that students do in the evenings (go to the pub, watch, party, study) and I want to work out whether you are studying. However, I can't ask you this directly, because you would probably lie to me. So, all I can do is to try to make observations about your behavior and appearance. I can probably work out if you look tired, hungover, scared, or fine. I want to use these observations to try to work out what you did last night. The problem is that, I don't even know why you look the way you do, but I can guess by assigning probabilities to these things.

(a) How would you model the problem?

(b) Give the structure of probability table(s) that you would require (No need to give probability values)

4. A dataset for a binary classification problem has 3 attributes and each attribute can take 5 distinct values. How many probability calculations in Naïve Bayes Classifier?

How many probability calculations when conditional independence is not assumed?

5. The class posterior probabilities for 3 classes are given as:

Data Point	C1	C2	C3
x	0.7	0.2	0.1
y	0.6	0.35	0.05
z	0.5	0.4	0.1

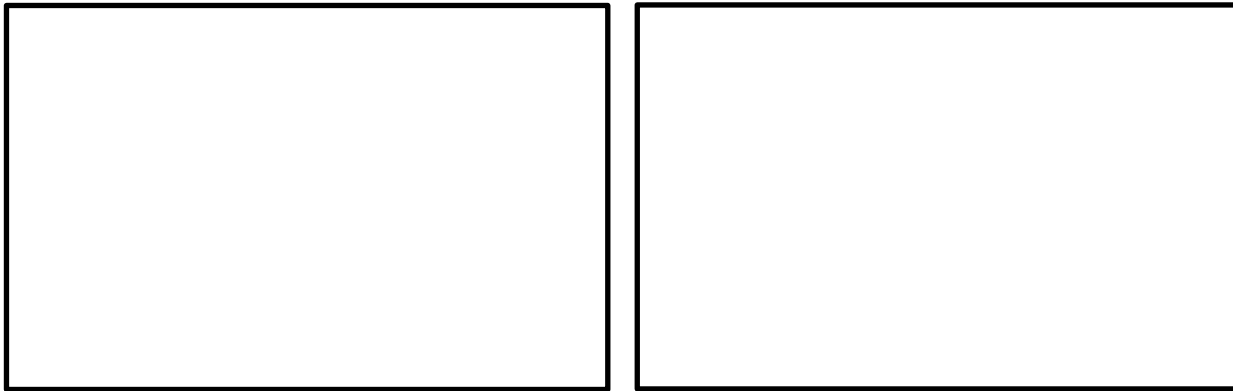
Data point picked up for labeling by Least Confident is _____, by Margin Sampling is _____, and by Entropy is _____.

6. A hard-margin linear SVM is used to classify 1.3 million points belonging to 2 classes. An active learning approach is used to select points for labeling. Under what circumstances, the active learning solution would match the passive learning solution.

7. If there are 4 options to choose from in Reinforcement learning, write suitable probabilities for each option, given that the best option has a probability of 0.8:

Greedy		ϵ -Greedy		Softmax	
0.8		0.8		0.8	

8. Partition a 2-dimensional space using 4-means and Kernel 4-means clustering algorithms.



9. For the three scatter plots given below, determine whether FLD and PCA would work (Yes) or not (No):

	Data	PCA	FLD
Class 1			
Class 2			