BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
Department of Computer Science and Information Systems
I SEMESTER 2023-2024
BITS F464 – Machine Learning
18th December, 2023   Comprehensive Exam – PART A (CLOSED BOOK)   Weightage: 20% (MM:40)
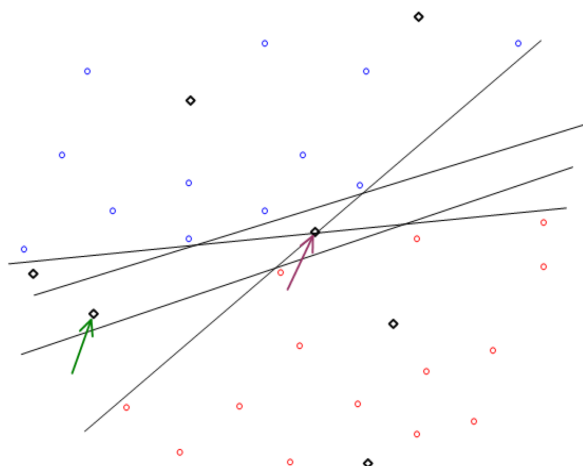Max. Time: 100 Minutes

1. Provide an upper bound to the generalization error of a classifier in terms of its training error, complexity, and size of training data. Is the bound deterministic? Why or why not? How can we quantify the complexity of a model? Use the upper bound to justify maximization of margin in SVM.

2. Pictorially illustrate the problem faced by 1 vs. 1 and 1 vs. rest approaches to solve multiclass classification problems using binary classifiers. Propose a solution and pictorially illustrate how it solves the problem.

3. Rank greedy, ε-greedy, and softmax action selection strategies in increasing order of (a) exploitation and (b) exploration. Justify your answer. No marks without proper justification.

4. Suppose you want to cluster N points into k clusters using the EM clustering algorithm. Give a template of the output which gives clustering information for each of the N points, $\{x_1, x_2, \ldots x_N\}$ in such a way that the same template can also be used for the output of K-means clustering algorithm. Use the common template to provide clustering information for x_1 and x_2 separately for EM and K-means clustering.

5. In Query-By-Committee (QBC), there are 9 committee members, C1, C2, …, & C9. There are 3 classes: A, B, & C. Which among the following points will be selected by QBC using vote entropy? X1 (ABCBBCBBB); X2(BCCBABBAC); X3(CABABABCA), and X4(AAAAAAAAA). The sequence in bracket represents the class label assigned by the committee members C1, C2, …, & C9, respectively. Vote entropy is given by:

$$x^*_{VE} = \operatorname*{argmax}_x - \sum_i \frac{V(y_i)}{C} \log \frac{V(y_i)}{C},$$

Show all the calculations.

6. What are the main aspects of Active Learning which differentiates it from Passive Learning? Compare active learning with boosting.

7. Consider a set of 7 interval classifiers. Points inside the interval are positive and points outside are negative. Pictorially illustrate situations in which the version space is (a) empty and (b) non-empty

8. Incorporation of priors is one of the key concepts in the Bayesian approach to decision making. Link the Bayesian Inference Engine (BIE) to Naïve Bayes' Classifier (NBC) by using a block diagram which connects concepts like priors, likelihoods, posteriors, and additional information using a classification problem having 4 classes.

9. Give the simplest neural network which can solve the 2D XOR problem. How can we introduce non-linearity in a neural network?

10. Using Query-By-Committee query strategy framework, which of the two points (pointed to by arrows) you would pick for labeling? (a) Mark the version space (b) mark the points which are informative and will be picked up for labeling by an active learner (c) mark the modified version spaces for the points pointed to by the left and right arrows when they are picked for labeling & (d) mark the point which will be picked up for labeling if we have budget to label only 1 point.

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
Department of Computer Science and Information Systems
I SEMESTER 2023-2024
BITS F464 – Machine Learning
18th December 2023   Comprehensive Exam – PART B (OPEN BOOK*)   Weightage: 15% (MM 30)

OPEN BOOK – Text & reference books mentioned in the handout. Handwritten/photocopies of class notes. Lecture slides. Calculators are allowed!

1. How the value of the hyperparameter is set in soft-margin SVM? Discuss the behavior of SVM at extreme values (zero to infinity) of the hyperparameter. Which type of SVM you will use for the following binary classification problem? Justify your answer.
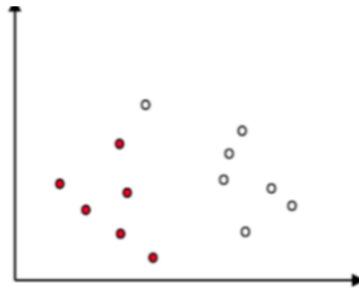
[6]

2. Suppose there is a business with 1024 offices. Each office has 100 GB of data about its customers. Currently you are using a cloud service provider to host all your data and you have a use case which requires you to classify data into 10 classes. The classification model has 4096 parameters, each 8-bytes long. You decide to switch to federated learning (FL) to address privacy concerns. Calculate the yearly savings in communication costs if your FL model required 10 rounds and you need to classify your data once every month. Assume that all offices participate in all rounds (this is called cross-silo FL). Every month you send 1 GB of data from each office to the FL server (hosted at the same cloud service provider) for building the skeleton model.

[8]

3. The Election Commission of India (ECI) has started planning for the 2024 Lok Sabha elections. To ensure free and fair elections, ECI conducts elections in multiple phases. As a ML expert, how would you help ECI in deciding on the number of phases and which constituencies go to poll in which phase. Use the techniques learned in the course to suggest a solution. Also identify the kind of data you would use. How would you help ECI in terms of visualization.

[6]

4. Mahalanobis Distance between two points in n-dimensional space is defined as:

$$mahalanobis(p,q) = (p-q)\Sigma^{-1}(p-q)^{T}$$

where Σ is the Covariance matrix of the input data X.

$$\Sigma_{j,k} = \frac{1}{n-1}\sum_{i=1}^{n}(X_{ij} - \overline{X}_{j})(X_{ik} - \overline{X}_{k})$$

What is the connection between Mahalanobis and Euclidean distance?

[4]

5. You are given four points, (0,6), (2,3), (3,2), & (6,0). First and fourth points belong to one class whereas, the middle two belong to another class. The non-linear function y=1/x separates the two classes. Get an idea from this function to find a transformation $f:(x,y) \rightarrow (x,y,z)$ to map the points to a 3-dimensional space so that the points become linearly separable. Also, find the equation of a plane which separates the two classes. For the chosen mapping f, can you find a kernel?

[6]