JK Lakshmipat University, Jaipur

Institute of Engineering and Technology

Machine Learning (CS1138)

Practice Questions III, BTech, IV Semester

Instructor: Dr. Arpan Gupta Date: 25 - 04 - 2025

Practice Questions. Some questions have already been discussed in class

1. Consider a neural network with input $x \in \mathbb{R}^3$, hidden layers sizes of 4 and 3 respectively. The output is a single unit for binary classification task. Write down the chain rule expression as product of derivatives for $\frac{\partial \mathbb{L}}{\partial \theta^{(1)}}$.

2. Perform a hierarchical clustering of the one-dimensional set of points {1, 4, 9, 16, 25, 36, 49, 64, 81} assuming clusters are represented by their centroid, and at each step the clusters with the closest centroids are merged. Draw the dendrogram illustrating the merging steps.

3. Given the following confusion matrix for a model, find its accuracy and the balanced accuracy values.

		Predictions		
		Class 0	**Class 1**	**Class 2**
Ground Truth	**Class 0**	**4**	**1**	**0**
	Class 1	**0**	**23**	**4**
	Class 2	**10**	**3**	**40**

4. Cluster the following eight points (with (x, y) representing locations) into three clusters:

   $A_1(2, 10), A_2(2, 5), A_3(8, 4), A_4(5, 8), A_5(7, 5), A_6(6, 4), A_7(1, 2), A_8(4, 9)$

   Initialize the cluster centers as: $A_1(2, 10), A_4(5, 8)$ and $A_7(1, 2)$.

   The distance function between two points $a = (x_1, y_1)$ and $b = (x_2, y_2)$ is defined as-

   $D(a, b) = |x_2 – x_1| + |y_2 – y_1|$

   Use K-Means Algorithm to find the three cluster centers after the second iteration.

5. How can the impurity of a split be evaluated in a decision tree? Show that the computation complexity for evalauting the split is $O(KN)$, where K are the number of categories and N are number of points in a dataset.

6. Show that the overall complexity of Decision Tree construction for all the splits is naively $O(DKN^2)$, where D are the number of dimensions. How can it be improved?

   Refer: `https://www.youtube.com/watch?v=0LB1cy2sCXc`

7. Given the dataset of whether John plays tennis or not (slide 33 of DT), usign ID3 algorithm, construct the complete decision tree. If we decide to use CART, then how can we evaluate a split based on attribute Wind, assuming Weak = 0, Strong = 1, and threshold of 0.5.

8. What is the Kullback-Liebler distance and why is it useful?

9. If we have n data points and d features, we store nd values in total. We can use principal component analysis to store an approximate version of this dataset in fewer values overall. If we use the first q principal components of this data, how many values do we need to approximate the original demeaned dataset? Justify your answer. (Ans: $qd + qn$)

10. True/False: For k-means clustering, the number of clusters k should be that which minimizes the loss function.

11. Give one advantage of hierarchical clustering over K-means clustering, and one advantage of K-means clustering over hierarchical clustering.

12. Suppose that after solving a soft margin SVM problem we obtain that the best separating hyperplane is $w^T x + b = 0$ for $w = [1, 2]$ and $b = 3$. Consider the following points $x_1 = [2, 1]$, $x_2 = [0.5, 1.5], x_3 = [1.75, 0.5]$. What are the labels (+1 or 1) assigned by our model to the three points?

13. Consider a data matrix $X \in \mathbb{R}^{n \times d}$. What is the smallest upper bound on $rank(X)$ which holds for every X? $rank(X) \leq$ ________

14. What are committee machines? Explain with an example.

15. What is Crowding Problem in t-SNE? Explain.

16. What is bagging? If we increase the number of classifers used in bagging, what is the effect on the overall variance of the ensemble model.

17. Explain the working of Random Forests? How can RFs help reduce the variance of the model?

18. (Advanced) Mathematically show that Bagging reduces variance.

19. (Advanced) How can KD-Trees be used for decision tree construction? Do they always find the tree that is balanced. (Refer: Kilian Weinberger - Cornell Univ).

20. (Advanced - Out of Scope) Find the SVD of A, $U\Sigma V^T$, where $A = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}$

    Refer: `https://www.d.umn.edu/~mhampton/m4326svd_example.pdf`

21. (Advanced) Mathematically show that the PCA minimizes the squared projection error when $X \in \mathbb{R}^{m \times n}$ is projected on a k dimensional subspace where $k < n$.

    Refer slides 3 to 6: `https://www.cs.toronto.edu/~mren/teach/csc411_19s/lec/lec12_matt.pdf`

    OR

    Derive the objective function for PCA.

22. (Advanced) How does Spectral Clustering work? Mention some applications of spectral clustering where they may be used.

23. (Advanced) Mathematically derive the optimization objective for SVMs.

24. (Advanced) What is the main assumption in Naive Bayes? What do you mean by Optimal Bayes Classifier? Explain.