Name:

Roll No.: **Dept.**:

IIT Kanpur
CS771A (IML)
Quiz-1
Date: August 23, 2023

Instructions: *Total:* **30 marks**

1. Please write your name, roll number, department on **all pages** of this question paper.
2. Write your answers clearly in the provided box. Keep your answer precise and concise.

Section 1 (15 very short answer questions: 15 x 2 = 30 marks). .

1. For the Learning with Prototypes (LwP) model with Euclidean distances, suppose the means of the positive and negative class are denoted by $\boldsymbol{\mu}_+$ and $\boldsymbol{\mu}_-$, respectively. A new test input $\boldsymbol{x}_*$ is classified as positive if (write down the expression):

2. Write down the expressions for the loss functions for least squares linear regression and ridge regression. Assume training data to be denoted as $\{\boldsymbol{x}_n, y_n\}_{n=1}^N$ and the weight vector to be denoted as $\boldsymbol{w}$.

3. Briefly explain (in at most 1-2 sentences) overfitting in terms of the training and test error of a machine learning model.

4. Consider a linear regression model $y_n = wx_n + b$ with scalar inputs, scalar weight w, and assume the bias term b as well. Also assume that the inputs x_n have already been centered (i.e., by subtracting off their mean from the original inputs) so their mean is 0. Assuming you are given N input-response pairs $\{x_n, y_n\}_{n=1}^N$, write down the expression for the bias term b (please do show the basic steps).

5. Consider two simple decision stumps D_1 and D_2. The first leaf of D_1 has 200 positive and 0 negative examples, and its second leaf has 0 positive and 200 negatives. The first leaf of D_2 has 100 positives and 100 negatives, and its second leaf also has 100 positives and 100 negatives. Which is these two decision stumps has a higher information gain and why? Explain only using words in at most 1-2 sentences.

Name:

Roll No.: **Dept.**:

IIT Kanpur
CS771A (IML)
Quiz-1
Date: August 23, 2023

6. For a multi-class classification model (e.g., softmax classification), given the learned weight vectors $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots,$ and a test input $\boldsymbol{x}_*$, how would you predict the class $\boldsymbol{x}_*$ belongs to? How would you predict the probability of $\boldsymbol{x}_*$ belonging to any class $i \in \{1, 2, \ldots, K\}$?

7. What is the advantage of using gradient descent to learn the weight vector of a linear/ridge regression model instead of learning it using the closed form solution?

8. The cross entropy loss on an example $(\boldsymbol{x}_n, y_n)$ where $y_n \in \{0, 1\}$ is defined as $\ell(y_n, \boldsymbol{x}_n, \boldsymbol{w}) = -[y_n \log \mu_n + (1 - y_n)\log(1 - \mu_n)]$ where $\mu_n = \sigma(\boldsymbol{w}^\top \boldsymbol{x}_n)$ is the model's predicted probability of the label being 1. Briefly explain why this loss makes sense for a binary classification problem?

9. Consider learning a decision tree, given some training data, where each input has D binary-valued features. Let's assume that we will not test any feature that has been tested at one of the previous levels (but we can possibly test a feature at multiple nodes at the same level). How many information gain calculations would be needed to construct the full decision tree (i.e., assuming no pruning)? Just give the basic expression; no need to try simplifying it too much to get a more compact expression.

10. In what situation, an ϵ-ball nearest neighbors method may fail to make a prediction at test time?

Name:

Roll No.: **Dept.**:

IIT Kanpur
CS771A (IML)
Quiz-1
Date: August 23, 2023

11. Briefly explain the difference between multi-class classification and multi-label classification. You may use an example to explain.

12. Rank (fastest to slowest) the following methods in terms of their prediction speed (i.e., how long they take to predict the label of a new test input) for a binary classification problem: LwP, K-nearest neighbors, decision tree, logistic regression. If some of these take roughly equal time, you may say so.

13. Can we use a linear regression model to learn a nonlinear regression function? Briefly explain your answer.

14. **(MCQ)** Which of these classifiers learns a linear separator (select all options that you think are correct)? (1) LwP when using Euclidean distance, (2) K-nearest neighbors classifier for K=1 and any distance function, (3) K-nearest neighbors classifier for any value of K and Euclidean distance, (4) logistic regression.

15. **(MCQ)** Which of these can be used for regression (select all options that you think are correct)? (1) Decision Tree, (2) K-nearest neighbor, (3) Learning with Prototypes, (4) Logistic Regression.

Name:

Roll No.: **Dept.**:

IIT Kanpur
CS771A (IML)
Quiz-1
Date: August 23, 2023