Name:

Roll No.: **Dept.**:

IIT Kanpur
CS771A (IML)
Quiz-2
Date: Sept 12, 2023

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

1. Duration is 45 minutes. Please write your name, roll number, department on **all pages**.
2. Write your answers clearly in the provided box. Keep your answer precise and concise.

Section 1 (Short/medium-length answer questions: 30 marks). .

1. Is a 3D sphere with a hole inside it a convex set? Briefly justify your answer. **(2 marks)**

2. Briefly explain (in at most 2-3 sentences) how AdaGrad defines a different learning rate in each dimensions (it is okay if you can't write the exact expression; explaining in words is fine). **(2 marks)**

3. Briefly explain (in at most 2-3 sentences) why mini-batch gradient descent is better than stochastic gradient descent which uses a single training example in each iteration. **(2 marks)**

4. We have an optimization problem in which we have a non-negativity constraint on a $D \times 1$ variable of interest $\boldsymbol{w}$. From a computational cost point of view, which of the following two approaches will be better and why: (1) projected gradient descent, and (2) Lagrangian based optimization? **(2 marks)**

5. Gaussian (RBF) kernel is defined as $k(\boldsymbol{x}_n, \boldsymbol{x}_m) = \exp(-\gamma||\boldsymbol{x}_n - \boldsymbol{x}_m||^2)$ where $\gamma > 0$ is the bandwidth parameter. Suppose we want each feature of the original inputs $\boldsymbol{x}_n, \boldsymbol{x}_m \in \mathbb{R}^D$ to have a different bandwidth parameter. Write down the expression for this variant of the RBF kernel. **(2 marks)**

Name:

Roll No.: **Dept.**:

IIT Kanpur
CS771A (IML)
Quiz-2
Date: Sept 12, 2023

6. What advantage does a linear SVM classifier offer over a Perceptron linear classifier? Can you modify the Perceptron algorithm to have a similar effect? If yes, how? If no, why not? **(2 marks)**

7. Consider two inputs $\boldsymbol{x} = [x_1, x_2]$ and $\boldsymbol{z} = [z_1, z_2]$. For the quadratic kernel defined as $k(\boldsymbol{x}, \boldsymbol{z}) = (1+\boldsymbol{x}^\top \boldsymbol{z})^2$, write down the expression of the feature mapping $\phi(\boldsymbol{x})$. **(2 marks)**

8. Consider the learning with prototypes (LwP) model. Suppose we want to replace the Euclidean distance used in standard LwP by a kernelized distance using a kernel function k. For this kernelized version of LwP, write down the expression for the distance of a test point $\boldsymbol{x}_*$ from the mean of any class c where $c = 1, 2, \ldots, K$ and K is the total number of classes. The expression must only be in terms of the kernel function k and the training and test inputs. **(6 marks)**

Name:

Roll No.: **Dept.**:

IIT Kanpur
CS771A (IML)
Quiz-2
Date: Sept 12, 2023

9. For a binary classification problem on the same training data, will of these will have a larger margin: (1) Hard-margin linear SVM, (2) Soft-margin SVM? Briefly justify your answer. **(2 marks)**

10. Why is kernel SVM slower than linear SVM at test time? **(2 marks)**

11. **(MCQ)** For which of the following problems, the objectives function is convex? (1) Hard-margin linear SVM, (2) Soft-margin linear SVM, (3) Hard-margin kernel SVM, (4) Soft-margin kernel SVM. **(2 marks)**

12. **(MCQ)** Which of the following is/are true regarding the Newton's method for minimizing a function: (1) Its per iteration cost is higher than gradient descent, (2) It always gives the global minima, (3) It uses a linear approximation of the function being minimized, (4) It usually takes fewer iterations to converge as compared to gradient descent. **(2 marks)**

13. **(MCQ)** Which of the following statements about optimization methods is/are true: (1) Co-ordinate descent does not require gradients, (2) Newton's method does not require learning rate, (3) Subgradient descent automatically handles constraints (e.g., nonnegativity) on the variables being optimized over, (4) How quickly (in terms of number of iterations) gradient descent converges for convex functions depends on the learning rate. **(2 marks)**

Name:

Roll No.: **Dept.**:

IIT Kanpur
CS771A (IML)
Quiz-2
Date: Sept 12, 2023