CS 771A: Intro to Machine Learning, IIT Kanpur				**Endsem Exam** (16 July 2024)
Name				**40 marks** Page **1** of **4**
Roll No		Dept.		

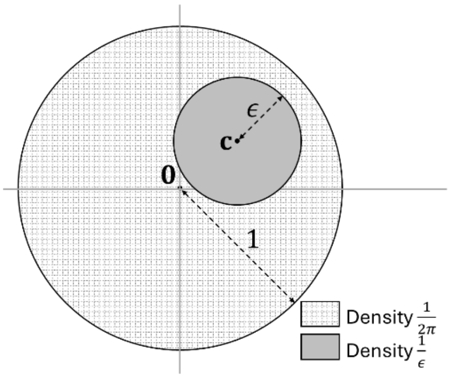
Instructions:

1. This question paper contains 2 pages (4 sides of paper). Please verify.
2. Write your name, roll number, department in **block letters** with **ink** on **each page**.
3. Write your final answers neatly **with a blue/black pen**. Pencil marks may get smudged.
4. Don't overwrite/scratch answers especially in MCQ – ambiguous cases may get 0 marks.

Q1. (**True-False**) Write **T** or **F** for True/False (write **only in the box on the right-hand side**). You must also give a brief justification for your reply in the space provided below.**(3 x (1+2) = 9 marks)**

1	EM run on data $\mathbf{x}_1, \ldots, \mathbf{x}_N \in \mathbb{R}^2$ s.t. $\|\mathbf{x}_i\|_2 \leq 2$ for all $i \in [N]$ to learn mixture of two Gaussians $\mathcal{N}(\boldsymbol{\mu}_k, I), k \in [2]$ will always ensure that the means satisfy $\|\boldsymbol{\mu}_k\|_2 \leq 2$.	
2	The difference of two Mercer kernels can never be Mercer. If True, give a proof. If False, construct two Mercer kernels K_1, K_2 with maps ϕ_1, ϕ_2 s.t. the difference $K_3 \stackrel{\text{def}}{=} K_1 - K_2$ is also a Mercer kernel with map ϕ_3. Give maps ϕ_1, ϕ_2, ϕ_3 explicitly.	
3	For convex differentiable $f: \mathbb{R} \to \mathbb{R}$, if $f\left(\frac{x+y}{2}\right) > 1$ for some $x, y \in \mathbb{R}$, then we must have $\max\{f(x), f(y)\} > 1$. Justify either using a proof or counter example.	

Q2 (Almost Uniform) Melbo is constructing a distribution $\mathcal{D}$ with support over 2D vectors of length up to 1 i.e. $\{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\|_2 \leq 1\}$. $\mathcal{D}$ has two parameters $\mathbf{c} \in \mathbb{R}^2, \epsilon \in [0,1]$ and assigns a *high* density $\frac{1}{\epsilon}$ in a "dense ball" of radius ϵ centered at $\mathbf{c}$ i.e., in $\{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\|_2 \leq 1, \|\mathbf{x} - \mathbf{c}\|_2 \leq \epsilon\}$ and a *low* density of $\frac{1}{2\pi}$ in the rest of the support i.e., in $\{\mathbf{x} \in \mathbb{R}^2 : \|\mathbf{x}\|_2 \leq 1, \|\mathbf{x} - \mathbf{c}\|_2 > \epsilon\}$. We have $\|\mathbf{c}\|_2 \leq 1 - \epsilon$ i.e., the dense ball stays within the support.



a. For which values of ϵ will $\mathcal{D}$ be a proper distribution? Find them and show calculations. You may find the fact that $\pi - \sqrt{\pi^2 - 2} \in [0,1]$ and $\pi - \sqrt{\pi^2 - 1} \in [0,1]$ to be useful.
b. Find out the mean vector $\boldsymbol{\mu} \in \mathbb{R}^2$ of this distribution. Show calculations. **(5 + 7 = 12 marks)**

Hint: the mean of a uniform distribution over a circle is its centre.

Find value(s) of ϵ for which $\mathcal{D}$ is a proper distribution.

Find out the mean vector of the distribution $\mathcal{D}$.

Q3 (Positive Linear Regression) We have data features $\mathbf{x}_1, \ldots, \mathbf{x}_N \in \mathbb{R}^D$ and labels $y_1, \ldots, y_N \in \mathbb{R}$ stylized as $X \in \mathbb{R}^{N \times D}, \mathbf{y} \in \mathbb{R}^N$. We wish to fit a linear model with positive coefficients:

$$\underset{\mathbf{w} \in \mathbb{R}^D,}{\operatorname{argmin}} \frac{1}{2} \|X\mathbf{w} - \mathbf{y}\|_2^2 \text{ s.t.} \, w_j \geq 0 \text{ for all } j \in [D]$$

1. Write the Lagrangian for this problem by introducing dual variables (no derivation needed).
2. Simplify the dual problem (eliminate $\mathbf{w}$) – show major steps. Assume $X^\top X$ is invertible.
3. Give a coordinate descent/ascent algorithm to solve the dual. **(2 + 4 + 6 = 12 marks)**

Write down the Lagrangian here (you will need to introduce dual variables and give them names)

Derive and simplify the dual. Show major calculations steps.

Give a coordinate descent/ascent algorithm to solve the dual problem.

Q4. (Kernel Smash) $K_1, K_2, K_3: \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ are Mercer kernels i.e., for any $x, y \in \mathbb{R}$, we have $K_i(x, y) = \langle \phi_i(x), \phi_i(y) \rangle$ with $\phi_1(x) = (1, x), \phi_2(x) = (x, x^2), \phi_3(x) = (x^2, x^4, x^6)$. Design a map $\phi_4: \mathbb{R} \to \mathbb{R}^7$ for kernel K_4 s.t. $K_4(x, y) = \left(K_1(x, y) - K_2(x, y)\right)^2 + 3K_3(x, y)$ for all $x, y \in \mathbb{R}$. No derivation needed. **Note that ϕ_4 must not use more than 7 dimensions. If your solution does not require 7 dimensions then leave the rest of the dimensions blank or fill with zero. (7 marks)**

$$\phi_4(x) =$$

$$(\boxed{}, \boxed{}, \boxed{}, \boxed{}, \boxed{}, \boxed{}, \boxed{})$$