CS 771A: Intro to Machine Learning, IIT Kanpur				**Quiz II** (01 Apr 2025)
Name	MELBO			**20 marks** Page **1** of **2**
Roll No	24007	Dept.	AWSM	

Instructions:

1. This question paper contains 1 page (2 sides of paper). Please verify.
2. Write your name, roll number, department above in **block letters neatly with ink**.
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 – such cases may get straight 0 marks.
5. Do not rush to fill in answers. You have enough time to solve this quiz.

Q1. (Tracing stuff) Let $\mathbf{u}_1, \dots, \mathbf{u}_{100} \in \mathbb{R}^{100}$ be the left singular vectors of a full-rank matrix $B \in \mathbb{R}^{100\times100}$. Let $\mathbf{p}, \mathbf{q}$ be random vectors with support $\{\mathbf{u}_1, \dots, \mathbf{u}_{100}\}$. Find the following quantities (no derivations). *Hint 1*: linearity of expectation. *Hint 2*: trace is linear but rank is not.(**1 x 5 = 5 marks**)

2.1	$\mathbb{E}[\text{trace}(\mathbf{p}\mathbf{q}^\top)]$ if $\mathbf{p}, \mathbf{q}$ are chosen uniformly but without replacement i.e. $\mathbf{p}$ is chosen uniformly randomly from $\{\mathbf{u}_1, \dots, \mathbf{u}_{100}\}$ and then $\mathbf{q}$ is chosen uniformly randomly from $\{\mathbf{u}_1, \dots, \mathbf{u}_{100}\}\backslash\{\mathbf{p}\}$.	0 as $\mathbb{E}[0] = 0$ $\text{trace}(\mathbf{p}\mathbf{q}^\top) = 0$ as $\mathbf{p} \perp \mathbf{q}$
2.2	$\text{trace}(\mathbb{E}[\mathbf{p}\mathbf{q}^\top])$ if $\mathbf{p}, \mathbf{q}$ are chosen uniformly but without replacement i.e. $\mathbf{p}$ is chosen uniformly randomly from $\{\mathbf{u}_1, \dots, \mathbf{u}_{100}\}$ and then $\mathbf{q}$ is chosen uniformly randomly from $\{\mathbf{u}_1, \dots, \mathbf{u}_{100}\}\backslash\{\mathbf{p}\}$.	0 because $\mathbb{E}[\text{trace}(\mathbf{p}\mathbf{q}^\top)] = \text{trace}(\mathbb{E}[\mathbf{p}\mathbf{q}^\top])$
2.3	$\text{trace}(\mathbb{E}[\mathbf{p}\mathbf{q}^\top])$ if $\mathbf{p}, \mathbf{q}$ are chosen uniformly but with replacement i.e. $\mathbf{p}$ is chosen uniformly randomly from $\{\mathbf{u}_1, \dots, \mathbf{u}_{100}\}$ and then $\mathbf{q}$ is chosen uniformly randomly from $\{\mathbf{u}_1, \dots, \mathbf{u}_{100}\}$ independently of $\mathbf{p}$.	0.01 as once $\mathbf{p}$ is chosen, w.p. 0.01 we get $\mathbf{q} = \mathbf{p}$
2.4	$\text{rank}(\mathbb{E}[\mathbf{p}\mathbf{p}^\top])$ if $\mathbf{p}$ is chosen uniformly randomly from $\{\mathbf{u}_1, \dots, \mathbf{u}_{100}\}$	100 as $\mathbb{E}[\mathbf{p}\mathbf{p}^\top] = 0.01 \cdot UU^\top = 0.01 \cdot I$
2.5	$\mathbb{E}[\text{rank}(\mathbf{p}\mathbf{p}^\top)]$ if $\mathbf{p}$ is chosen uniformly randomly from $\{\mathbf{u}_1, \dots, \mathbf{u}_{100}\}$	1 as $\mathbb{E}[1] = 1$ $\text{rank}(\mathbf{p}\mathbf{p}^\top) = 1$ for every $\mathbf{p}$

Q2. (Kernel Smash) $K_1, K_2 \colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ 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 the feature maps given below. Design a map $\phi_3 \colon \mathbb{R} \to \mathbb{R}^9$ for kernel K_3 s.t. $K_3(x, y) = \left(K_1(x, y) + K_2(x, y)\right)^2$ for all $x, y \in \mathbb{R}$. No derivation needed. **Note that ϕ_3 must not use more than 9 dimensions. If your solution does not require 9 dimensions then fill the rest of the dimensions with zero.** **(9 marks)**

$$\phi_1(x) = \left(\frac{1}{x}, x\right), \phi_2(x) = \left(\frac{1}{x^2}, x^2\right)$$

$$\phi_3(x) = \left(\boxed{x^{-4}}, \boxed{\sqrt{2} \cdot x^{-3}}, \boxed{x^{-2}}, \boxed{\sqrt{2} \cdot x^{-1}}, \boxed{2}, \boxed{\sqrt{2} \cdot x}, \boxed{x^2}, \boxed{\sqrt{2} \cdot x^3}, \boxed{x^4}\right)$$

Q3. (**True-False**) Write **T** or **F** for True/False in the **box on the right** and a **brief justification** in the space below (brief proof if **T**, counterexample if **F**). A square matrix is termed *diagonal* if all of its off-diagonal entries are zero (its diagonal entries can be zero/-ve/+ve). (**3 x (1+1) = 6 marks**)

1	A diagonal matrix $A \in \mathbb{R}^{3\times 3}$ with all diagonal entries being non-zero must always have rank exactly equal to 3.	T

Let $A = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3]$ where $\mathbf{v}_i = a_i \cdot \mathbf{e}_i \in \mathbb{R}^3$, $a_i \neq 0$ and $\mathbf{e}_i$ are the standard basis vectors. We claim that the three columns of A are linearly independent of each other. To see this, suppose for some non-zero constants p_i, we have $\sum_{i\in[3]} p_i \cdot \mathbf{v}_i = \mathbf{0}$. This means that $\sum_{i\in[3]}(p_i \cdot a_i) \cdot \mathbf{e}_i = \mathbf{0}$. However, the standard basis vectors are independent of each other which means $p_i \cdot a_i = 0$. Since we know that $a_i \neq 0$ this means that we must have $p_i = 0$. This means that $\mathbf{v}_i$ are independent of each other and thus A has column rank (and hence rank) equal to 3.

2	It is possible for a diagonal matrix $B \in \mathbb{R}^{3\times 3}$ to be positive semi-definite if it has two strictly positive diagonal entries and one strictly negative diagonal entry.	F

Let $B = [\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3]$ where $\mathbf{u}_i = b_i \cdot \mathbf{e}_i \in \mathbb{R}^3$, $\mathbf{e}_i$ are the standard basis vectors and $b_1, b_2 > 0$ but $b_3 < 0$. Consider the vector $\mathbf{x} \stackrel{\text{def}}{=} [0,0,c]$. We have $\mathbf{x}^\top B\mathbf{x} = c^2 b_3 < 0$. This means B is not positive semi-definite.

3	For any Mercer kernel $K: \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ and any vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^2$, the matrix $G = [g_{ij}] \in \mathbb{R}^{n\times n}$ defined as $g_{ij} \stackrel{\text{def}}{=} K(\mathbf{x}_i, \mathbf{x}_j)$ is always positive semi-definite.	T

Let $\phi: \mathbb{R}^2 \to \mathcal{H}$ be the feature map for the kernel K. For any vector $\mathbf{v} \in \mathbb{R}^n$, we have

$$\mathbf{v}^\top G\mathbf{v} = \sum_{\substack{i\in[n]\\ j\in[n]}} v_i v_j g_{ij} = \sum_{\substack{i\in[n]\\ j\in[n]}} v_i v_j \phi(\mathbf{x}_i)^\top \phi(\mathbf{x}_j) = \left(\sum_{i\in[n]} v_i \phi(\mathbf{x}_i)\right)^\top \left(\sum_{j\in[n]} v_j \phi(\mathbf{x}_j)\right)$$

The last expression is just $\|\mathbf{p}\|_{\mathcal{H}}^2 > 0$ where $\mathbf{p} = \sum_{i\in[n]} v_i \phi(\mathbf{x}_i) \in \mathcal{H}$. This means that G is positive semi-definite.