CS 771A: Intro to Machine Learning, IIT Kanpur				**Quiz II** (24 Oct 2025)
Name	MELBO			**20 marks**
Roll No	250007	Dept.	AWSM	Page **1** of **2**

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. (Matrix Norm) $A \in \mathbb{R}^{m\times m}$ is a (possibly non-symmetric) square matrix with SVD $A = U\Sigma V^\top$. The diagonal matrix $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_m) \in \mathbb{R}^{m\times m}$ has $\sigma_1 > \sigma_2 > \dots > \sigma_m$ (all singular values are distinct), $U = [\mathbf{u}_1, \dots \mathbf{u}_m], V = [\mathbf{v}_1, \dots \mathbf{v}_m] \in \mathbb{R}^{m\times m}$ are orthonormal. Let $a^* \stackrel{\text{def}}{=} \max_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_2$. Note the constraint $\|\mathbf{x}\|_2 = 1$. Find the value of a^* in terms of $\mathbf{u}_i, \mathbf{v}_i, \sigma_i, i \in [m]$. Show brief derivation. *Hint: try expressing* $\mathbf{x}$ *as a linear combination of the right singular vectors* $\mathbf{v}_1, \dots, \mathbf{v}_m$ *which is always possible as they form a basis over* $\mathbb{R}^m$. *You may also find* $VV^\top = I = V^\top V$ *useful.* (**2+3 = 5 marks**)

$a^* \stackrel{\text{def}}{=} \max_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_2 =$ $\boxed{\sigma_1 \text{ or equivalently, } \max_i \{\sigma_i\}}$

Give brief derivation below

Let $\mathbf{z} \stackrel{\text{def}}{=} V^\top\mathbf{x} \in \mathbb{R}^m$ be representation of $\mathbf{x}$ in the basis of the right singular vectors i.e. $\mathbf{x} = V\mathbf{z}$. Note that $\|\mathbf{z}\|_2^2 = \mathbf{z}^\top\mathbf{z} = \mathbf{x}^\top VV^\top\mathbf{x} = \mathbf{x}^\top\mathbf{x} = \|\mathbf{x}\|_2^2 = 1$. Thus, $\|A\mathbf{x}\|_2 = \|U\Sigma V^\top V\mathbf{z}\|_2 = \|U\Sigma\mathbf{z}\|_2$. This is unsurprising since all that V can do is rotate and swap stuff and not change the length of a vector. By using orthonormality of U similarly, we get $\|U\Sigma\mathbf{z}\|_2 = \|\Sigma\mathbf{z}\|_2$.

Thus, the problem reduces to $\max_{\|\mathbf{z}\|_2=1} \|\Sigma\mathbf{z}\|_2 = \max_{\|\mathbf{z}\|_2=1} \sqrt{\sum_{i\in[m]} \sigma_i^2 z_i^2}$.

Since $\sigma_1 > \sigma_2 > \dots > \sigma_m$, we get $\sum_{i\in[m]} \sigma_i^2 z_i^2 \le \sigma_1^2\left(\sum_{i\in[m]} z_i^2\right) = \sigma_1^2$. This maximum value is achieved when $\mathbf{z} = (1,0,0,\dots,0) = \mathbf{e}_1$ i.e. when $\mathbf{x} = \mathbf{v}_1$, the leading right singular vector. Indeed, we have $\|A\mathbf{x}\|_2 = \|U\Sigma V^\top\mathbf{v}_1\|_2 = \|U\Sigma\mathbf{e}_1\|_2 = \sigma_1 \cdot \|\mathbf{u}_1\|_2 = \sigma_1$.

Q2. (Kernel Smash) $K_1, K_2: \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ are Mercer kernels with $K_i(x,y) = \langle\phi_i(x), \phi_i(y)\rangle$ for any $x, y \in \mathbb{R}$ with the feature maps given below. Design a map $\phi_3: \mathbb{R} \to \mathbb{R}^3$ for a kernel K_3 defined as $K_3(x,y) = K_1(x,y)\cdot K_2(x,y)$. No derivation needed. $\boldsymbol{\phi_3}$ **must not use more than 3 dimensions. If your solution doesn't need 3 dimensions, fill remaining ones with zero.** (**6 marks**)

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

$$\phi_2(x) = (x, x^2) \in \mathbb{R}^2$$

$$\phi_3(x) = \left(\boxed{x}, \boxed{\sqrt{2}}, \boxed{\frac{1}{x}}\right)$$

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+2) = 9 marks)**

1	For a square symmetric matrix $A \in \mathbb{R}^{3\times 3}$ having eigen decomposition $A = VLV^\top$, we always have $\text{trace}(A) = \text{trace}(L)$ where L is the diagonal matrix of eigenvalues.	T

Using the cyclic invariance property of the trace operator, we get $\text{trace}(A) = \text{trace}(VLV^\top) = \text{trace}(V^\top VL) = \text{trace}(L)$ since $V^\top V = I$, as V being the matrix of eigenvectors, is orthonormal.

2	Given any two matrices $A, B \in \mathbb{R}^{3\times 3}$, the largest singular value of $A + B$ is always equal to the sum of the largest singular value of A and largest singular value of B.	F

Consider $A = I$ and $B = -I$ so that $A + B = 0$. The SVD of A is $A = III$ and all singular values are unity. The SVD of B is $B = II(-I)$ or else $B = (-I)II$ and all singular values are still unity. Note that singular values are never negative, so the left or right singular vectors absorb the negative sign. The SVD of the sum is $A + B = I0I$. This presents a counter example: the largest singular value of A is 1, as is for B but the largest singular value of the sum is $0 \neq 1 + 1$.

3	For a kernel $K: \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$, $K(\mathbf{x}, \mathbf{y}) \overset{\text{def}}{=} \exp(\mathbf{x}^\top \mathbf{y})$ and any vectors $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^2$, the matrix $G = [g_{ij}] \in \mathbb{R}^{n\times n}$ with $g_{ij} \overset{\text{def}}{=} K(\mathbf{x}_i, \mathbf{x}_j)$ is always full rank, $\text{rank}(G) = n$.	F

Example 1: If $\mathbf{x}_1 = \mathbf{x}_2$ then the first two rows of G are identical, as are the first two columns of G. This means that G cannot have n linearly independent rows or n independent columns and thus, it must be rank deficient i.e. $\text{rank}(G) < n$.
Example 2: if one of the points is the zero vector i.e., $\mathbf{x}_i = \mathbf{0}$, then the rank collapses to unity.

Side Note: This kernel is called the exponential kernel and it is a close cousin of the Gaussian/RBF kernel. If we normalize the kernel in its RKHS as follows, then we get the RBF kernel

$$\tilde{K}(\mathbf{x}, \mathbf{y}) \overset{\text{def}}{=} \frac{K(\mathbf{x}, \mathbf{y})}{\sqrt{K(\mathbf{x}, \mathbf{x}) \cdot K(\mathbf{y}, \mathbf{y})}}$$

It can be shown that if all points $\mathbf{x}_1, \dots, \mathbf{x}_n$ are non-zero and distinct then this kernel will indeed always be full rank. However, if points come too close to each other, then the smallest eigenvalue starts approaching 0. These results are more involved and beyond the scope of this discussion.