The following information should be so legible that a machine can read it

Roll No : 123456789 **Dept.**: CSE
e.g., 190040001 e.g., CSE

IIT Bombay
CS 217/337: AIML
Quiz 2, 2023-24-II
Date: April 3, 2024

CS 217/337: Artificial Intelligence and Machine Learning

Total: **10 + 8 + 12 = 30 points,** *Duration:* **1 hour, ATTEMPT ALL QUESTIONS**

Instructions:

1. This question-and-answersheet booklet contains a total of **6 sheets** of paper **(12 pages, pages 2 and 12 are blank)**. Please verify.
2. Write your roll number and department on **every side of every sheet** (except the blank sheet) of this booklet. Use only **black/blue ball-point pen**. The first 5 minutes of additional time is given exclusively for this activity.
3. Write final answers neatly with a pen **only in the given boxes**.
4. Use the rough sheets for scratch works / attempts to solution. **Write only the final solution (which may be a sequence of logical arguments) in a precise and succinct manner in the boxes provided**. Do not provide unnecessarily elaborate steps. The space within the boxes is sufficient for the correct and precise answers.
5. Submit your answerscripts to the teaching staff when you leave the exam hall or the time runs out (whichever is earlier). **Your exam will not be graded if you fail to return the paper**.
6. **This is a closed book, notes, internet exam. No communication device, e.g., cellphones, iPad, etc., is allowed**. Keep it switched off in your bag and keep the bag away from you. If anyone is found in possession of such devices during the exam, that answerscript may be disqualified for evaluation and DADAC may be invoked.
7. One A4 assistance sheet (text **only on one side**) and a scientific calculator are allowed for the exam.

The following information should be so legible that a machine can read it

IIT Bombay
CS 217/337: AIML
Quiz 2, 2023-24-II
Date: April 3, 2024

Roll No : e.g., 190040001 | **Dept.**: e.g., CSE

Problem 1 (10 points). Kernels and Support Vector Machine.

(a) Is the following statement true? Every valid *polynomial* kernel K satisfies $K(x_i, x_j) = K(x_j, x_i)$, where x_i and x_j are the feature vectors of the i^{th} and j^{th} examples. If yes, argue in two sentences, if not, provide a counterexample in 2-dimension. Note: marks will be awarded only if both parts are correct.

1 point.

Yes. Polynomial kernel is given by $K(x_i, x_j) = (1 + \phi(x_i)^T \phi(x_j))^d$

where ϕ is a mapping of x_i's to a higher dimension.

Since inner products are symmetric, i.e. $\phi(x_i)^T \phi(x_j) = \phi(x_j)^T \phi(x_i)$

the above statement holds.

(b) Given n feature vectors $x_i, i = 1, \ldots, n$, construct a kernel matrix A, which is an $n \times n$ square matrix, as $A_{i,j} = K(x_i, x_j)$. K is a linear kernel. Tick all the correct alternatives: **1 point.**

- (1) A is positive definite. ☐
- (2) A is negative definite. ☐
- (3) A is positive semidefinite. ☑
- (4) A is negative semidefinite. ☐

(c) Prove all the claims you have ticked in the previous part of this question. *Hint: use only the following definition for positive semidefiniteness of a matrix: $\forall z \neq 0, z^\top Az \geqslant 0$. For negative semidefinite, the inequality reverses and definiteness makes both inequalities strict.* Note: this part may not be checked if the alternative(s) chosen in the previous part is(are) incorrect. **2 points.**

Since K is a linear kernel, $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$

define a matrix $M = [\phi(x_1) \ \phi(x_2) \ \cdots \ \phi(x_n)]$, which has n columns, each corresponding to one transformed data point.

Since $A_{ij} = K(x_i, x_j) = \phi(x_i)^T \phi(x_j) = (M^T M)_{ij}$

$\Rightarrow \quad A = M^T M \quad \text{-- (1)}$

Therefore $\quad z^T A z = z^T M^T M z = \|Mz\|^2 \geqslant 0 \quad \forall z \neq 0$

This proves the claim $\square$

The following information should be so legible that a machine can read it

Roll No : e.g., 190040001   **Dept.**: e.g., CSE

IIT Bombay
CS 217/337: AIML
Quiz 2, 2023-24-II
Date: April 3, 2024

(d) Suppose K_1 and K_2 are two $\mathbb{R}^n \times \mathbb{R}^n$ *linear* kernels, i.e., both of them have feature mappings $\phi_1, \phi_2 : \mathbb{R}^n \to \mathbb{R}^d$ respectively such that $K_i(x,z) = \phi_i(x)^\top \phi_i(z), i = 1, 2$. Let $c > 0$ be a positive constant. Find the feature mapping of a linear kernel $K(x,z) = cK_1(x,z)K_2(x,z)$, given by ϕ, in terms of the original feature mappings ϕ_1 and ϕ_2. Explain each step of the derivation. **2 points.**

Given $K(x,z) = \phi(x)^T \phi(z) = c\ \phi_1(x)^T \phi_1(z) \cdot \phi_2(x)^T \phi_2(z)$

$$= c\left(\sum_{i=1}^{d} \phi_{1i}(x)\, \phi_{1i}(z)\right)\left(\sum_{j=1}^{d} \phi_{2j}(x)\, \phi_{2j}(z)\right)$$

$$= c \sum_{j=1}^{d}\sum_{i=1}^{d}\left[\left(\phi_{1i}(x)\, \phi_{2j}(x)\right) \times \left(\phi_{1i}(z)\, \phi_{2j}(z)\right)\right]$$

define $\phi(x) := \left(\sqrt{c}\ \phi_{1i}(x)\, \phi_{2j}(x)\right)_{\substack{i=1,2,\dots,d \\ j=1,2,\dots,d}}$

hence $\phi : \mathbb{R}^n \to \mathbb{R}^{d^2}$ maps to d^2 dimensional space

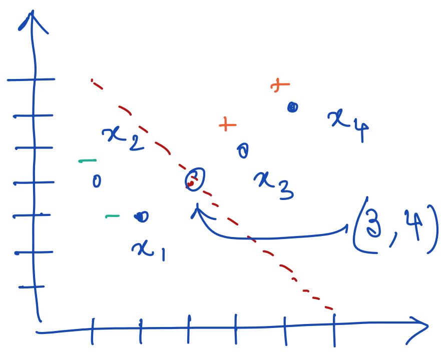
This ϕ gives the desired kernel $K(x,z)$.

(e) Consider training a hard-margin SVM on a tiny dataset of *four* labeled training instances on a 2-d plane:

$$\{((2,3),-1),((1,4),-1),((4,5),+1),((5,6),+1)\}.$$

Each instance i is denoted by $(\mathbf{x}_i, y_i)$ that shows the feature vector and the label respectively. Find (1) the weight vector $\mathbf{w}$ and bias b of the hard-margin SVM, (2) the equation of the decision boundary. Show the relevant steps to derive your answers. **(2 + 1) points.**

Easiest to solve it graphically. Note that x_1, x_3, x_4 points the collinear and the line connecting x_1, x_2 is perpendicular to the line connecting x_1, x_3, x_4.
Hence the optimal decision boundary will pass through the point $(3,4)$ and will be perpendicular to the line joining x_1, x_3, x_4



The following information should be so legible that a machine can read it

Roll No : e.g., 190040001   **Dept.**: e.g., CSE

IIT Bombay
CS 217/337: AIML
Quiz 2, 2023-24-II
Date: April 3, 2024

The equation $w^T x + b$ has to be $(+1)$ at $(4,5)$ and (-1) at $(2,3)$

hence
$$\left.\begin{aligned} 4w_1 + 5w_2 + b &= 1 \\ 2w_1 + 3w_2 + b &= -1 \end{aligned}\right\} \text{(1)}$$

Also, the decision boundary has a slope of -1 since the decision boundary is $w^T x + b = 0$ the ratio of $w_1/w_2 = 1 \Rightarrow w_1 = w_2$.

Plugging this into (1) and solving, we get
$$w_1 = w_2 = \frac{1}{2} \text{ and } b = -7/2 .$$

(f) Write the feature vectors that are the *support vectors* of this SVM. **1 point.**

From the figure (as well as from the equations), it is clear that

$(2,3) \leftarrow x_1$
$(1,4) \leftarrow x_2$
$(4,5) \leftarrow x_3$

are the support vectors of this SVM.

The following information should be so legible that a machine can read it

Roll No : e.g., 190040001 **Dept.**: e.g., CSE

IIT Bombay
CS 217/337: AIML
Quiz 2, 2023-24-II
Date: April 3, 2024

Problem 2 (8 points). **Principal Component Analysis.**

(a) Given *three* data points in 2-d space, $(1,1)$, $(2,2)$, and $(3,3)$, find the first principal component (PC). Remember, PCs are orthonormal. **1 point.**

$$X = \begin{pmatrix} x_1-\mu & x_2-\mu & x_3-\mu \end{pmatrix} = \begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}$$

$$S = \frac{1}{2} XX^T = \frac{1}{2}\begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix}\begin{bmatrix} -1 & -1 \\ 0 & 0 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

The eigenvector corresponding to the largest eigenvalue $(= 2)$ is $\begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$ or the negative of it (Both will be given marks)

(b) If we project the original data points into 1-d space given by the first principal component, what is the variance of the projected data? **1 point.**

$u^T S u = 4/3$.

The following information should be so legible that a machine can read it

IIT Bombay
CS 217/337: AIML
Quiz 2, 2023-24-II
Date: April 3, 2024

Roll No : e.g., 190040001 | **Dept.**: e.g., CSE

(c) Consider n data points given by $x_1, x_2, \ldots, x_n$, where $x_i \in \mathbb{R}^d$. Denote the original covariance matrix of these points as S_{orig}. Suppose, we perform a transformation to the data points by appending a 1 at the end of each data point, i.e., the transformed data points become $z_i = \begin{pmatrix} x_i \\ 1 \end{pmatrix}$. Express the covariance matrix of the transformed data points S_{tran} in terms of S_{orig}. **2 points.**

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad , \quad S_{orig} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T$$

$$\text{Now, } z_i = \begin{pmatrix} x_i \\ 1 \end{pmatrix} \Rightarrow \bar{z} = \begin{pmatrix} \bar{x} \\ 1 \end{pmatrix} ;$$

$$\text{Hence } S_{tran} = \frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})(z_i - \bar{z})^T = \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} x_i - \bar{x} \\ 0 \end{pmatrix}\begin{pmatrix} x_i - \bar{x} \\ 0 \end{pmatrix}^T$$

$$= \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T & 0_{d\times 1} \\ 0_{1\times d} & 0 \end{bmatrix} = \begin{bmatrix} S_{orig} & 0_{d\times 1} \\ 0_{1\times d} & 0 \end{bmatrix}$$

(d) How does the principal components of the transformed dataset look like w.r.t. that of the original dataset? Explain your answer. *Hint: express the new PCs in terms of the old PCs and argue why they should be the PCs of the new data points.* Marks are equally divided between the correct answer and its explanation. **(2 + 2) points.**

Say the PCs of the original dataset, i.e., the eigenvectors of S_{orig} are given by $u_1, u_2, \ldots, u_d$. Consider the vectors $\begin{pmatrix} u_1 \\ 0 \end{pmatrix} \cdots \begin{pmatrix} u_d \\ 0 \end{pmatrix} \underset{e_{d+1}}{\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}}$

It is easy to verify that these are eigenvectors of S_{tran} with the same first d eigenvalues and the last eigenvalue corresponding to e_{d+1} is zero.

Hence, the principal components of the transformed dataset are the same as that of the original with one zero appended at the end with one additional PC with eigenvalue zero.

The following information should be so legible that a machine can read it

Roll No : e.g., 190040001 | **Dept.**: e.g., CSE

IIT Bombay
CS 217/337: AIML
Quiz 2, 2023-24-II
Date: April 3, 2024

Problem 3 (12 points). Nim is a two-player game (see an example image below), in which piles of matchsticks are placed before the players.

The number of piles, i.e., the rows in the image, of the game is finite, and each pile contains a finite number of matchsticks. These numbers can vary for different instances of Nim. Each player, during her turn, chooses exactly one pile, and removes any number of matchsticks from the pile she has selected (she must remove at least one matchstick). The player who removes the last matchstick wins the game. We state (without proof) the following result due to John von Neumann.

Theorem (von Neumann (1928)). In every finite[1] two-player game (with perfect information – where players can observe every board position, i.e., the types of games we discussed in the class) in which the set of outcomes is O = {1 wins, 2 wins, Draw}, one and only one of the following three alternatives holds:

1. Player 1 has a winning strategy.[2]
2. Player 2 has a winning strategy.
3. Each of the two players has a strategy guaranteeing at least a draw.

Now consider the following sequence of exercises and answer the questions. Every question that asks a conclusion and its explanation, the explanation will not be checked if the conclusion is incorrect.

(a) What can you conclude about the existence of a winning strategy (of any player) in Nim from von Neumann's theorem? (never exists/may not exist/always exists) Explain your answer, e.g., why the theorem is applicable and how you reached the conclusion. **(1 + 1) points.**

There always exists a winning strategy for one of the players.
This is a perfect information two-player game, So, the theorem applies.
However, the game has finite number of piles and matchsticks in each pile. Also, at least one matchstick is removed at every turn. So, the game always ends with some player picking the last matchstick. Hence there is

[1] players, number of actions of each player, and the number of rounds – all are finite.
[2] A strategy is a winning strategy for a player if it guarantees a win to that player irrespective of what the other player plays.

The following information should be so legible that a machine can read it

IIT Bombay
CS 217/337: AIML
Quiz 2, 2023-24-II
Date: April 3, 2024

Roll No : ____ e.g., 190040001 **Dept.**: ____ e.g., CSE

---

no draw outcome in this game. Hence, according to the theorem, at least one player must have a winning strategy.

(b) At the beginning of play, list, in a column, the number of matchsticks in each pile, expressed in binary in a right aligned manner. For example, if there are 4 piles containing, respectively, 2, 12, 13, and 21 matchsticks, list:

10		10	
1100	$\longrightarrow$	1100	winning
1101		1101	position
10101	$\xrightarrow{\text{remove 18}}$	00011	

Next, check whether the number of 1s in each column is odd or even. In the above example, counting from the right, in the first and fourth columns the number of 1s is even, while in the second, third, and fifth columns the number of 1s is odd.

A position in the game will be called a **winning position** if the number of 1s in each column is even. The game state depicted above is *not* a winning position.

Starting from any position that is not a winning position, it is possible (possible / not possible) to get to a winning position in one move (that is, by removing matches from a single pile). *Hint: try it out in the given example.* Explain precisely your answer in the response box below. You may use the example above to illustrate, but the explanation should be as general as possible.

(1 + 2) points.

Suppose the game is in a state where the configuration is not a winning position. Then there exists at least one column having odd number of 1's.

- Select the leftmost column with the odd # of 1's, say c^*
- Select a pile which has 1 in that column c^*, say p^*
- Make that 1 to zero and update all the entries in that pile to the right of it such that all columns to the right of c^* now have

The following information should be so legible that a machine can read it

Roll No : e.g., 190040001 | **Dept.**: e.g., CSE

IIT Bombay
CS 217/337: AIML
Quiz 2, 2023-24-II
Date: April 3, 2024

---

even number of 1's.

- This binary number for this pile is smaller than that it had originally. Subtract the updated number from the original. These are the number of sticks to be removed from p^* to bring the state to a winning position.

E.g., remove 18 sticks from the last pile to bring it to a winning position in the given example (see there)

(c) At a winning position, every legal action leads to a non-winning (winning / non-winning) position. Explain precisely your answer in the response box below. **(1 + 2) points.**

In every legal action, the number of matches in exactly one pile decreases. Hence, there must exist a leftmost 1 in that pile that changes to zero after taking that legal action. Hence, the column corresponding to that 1 (which now switched to zero) became a column with odd # of 1's (since the original position was a winning position). This makes the new position non-winning.

(d) Explain why at the end of every instance of Nim, the position of the game will be a winning position. **1 point.**

At the end, all sticks are removes. Hence every pile has zero — which is trivially a winning position.

The following information should be so legible that a machine can read it

IIT Bombay
CS 217/337: AIML
Quiz 2, 2023-24-II
Date: April 3, 2024

Roll No : e.g., 190040001 | **Dept.**: e.g., CSE

(e) Explain how we can identify which player can guarantee a victory for herself (given the initial number of piles of matchsticks and the number of matchsticks in each pile). Describe that player's winning strategy. **(1 + 2) points.**

Case 1 :

If the instance of the game starts with a non-winning position then the first player has a winning strategy. As discussed in part (b), this player can always set the game to a winning position. The other player has no choice but to bring the game to a non winning position. Since the final state of empty will be achieved by player 1, she guarantees a win.

Case 2 :

If the instance of the game starts with a winning position then the second player has a winning strategy. Player 1 will bring the game to a non winning position in the first move. Player 2 then follows the same strategy of player 1 in case 1. This guarantees win to player 2.

END OF QUESTION PAPER. GOOD LUCK!