CS 771A: Intro to Machine Learning, IIT Kanpur			**Midsem Exam** (17 Sep 2025)
Name			**40 marks**
Roll No		Dept.	Page **1** of **4**

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 will get 0 marks.

Q1 (M-SVM Dual) In a previous discussion, we saw that if Mahalanobis distances are used to derive the SVM objective instead of Euclidean distances, then the SVM changes its form as shown below (bias is hidden inside the model vector). Here $y^i \in \{-1,1\}$, $\mathbf{w}, \mathbf{x}^i \in \mathbb{R}^D$, $A \in \mathbb{R}^{D\times D}$ is a symmetric, invertible, positive definite matrix i.e., $\mathbf{x}^\top A\mathbf{x} > 0$ and $\mathbf{x}^\top A^{-1}\mathbf{x} > 0$ for non-zero vectors $\mathbf{x} \in \mathbb{R}^D$. Derive the Lagrangian dual and show all steps as directed below. **(2 + 2 + 2 = 6 marks)**

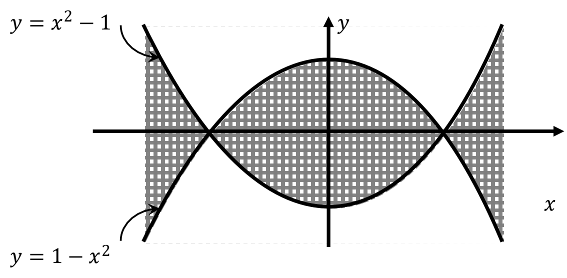
$$\min_{\mathbf{w}} \frac{1}{2}\mathbf{w}^\top A^{-1}\mathbf{w} \quad \text{s.t.} \quad y^i \cdot \left(\mathbf{w}^\top\mathbf{x}^i\right) \geq 1 \quad \forall i \in [n]$$

Write down the Lagrangian by introducing dual variables. No derivation needed.

Give an expression of the model $\mathbf{w}$ in terms of the dual variable. No derivation needed.

Simplify the dual problem (eliminate $\mathbf{w}$) – show major steps. You may use $\frac{\partial \mathbf{x}^\top B\mathbf{x}}{\partial \mathbf{x}} = (B + B^\top)\mathbf{x}$

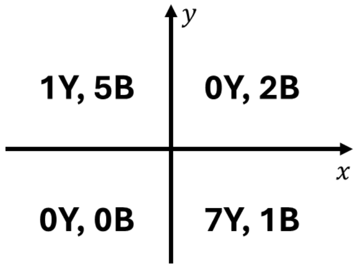
Q2. (Candy Classifier) The figure shows a binary classification task. The bold lines are the curves $y = x^2 - 1$ and $y = 1 - x^2$ where $x \in \mathbb{R}$. Create a feature map $\phi: \mathbb{R}^2 \to \mathbb{R}^D$ for some integer $D > 0$ so that for any 2D vector $\mathbf{z} = (x, y) \in \mathbb{R}^2$, the value of $\text{sign}\left(\mathbf{1}^\top\phi(\mathbf{z})\right)$ is $+1$ if $\mathbf{z}$ is in the cross-hatched region and -1 if $\mathbf{z}$ is in the white region. The D-dimensional all-ones



vector is denoted as $\mathbf{1} = (1,1,\ldots,1) \in \mathbb{R}^D$. Write down your feature map in the space below. To create your map, you may use common functions such as polynomials, absolute value, exponential etc i.e., a feature map such as $\phi(\mathbf{z}) = \left(x, y, \exp(x-y), y^2 - x^2, \sqrt{x}\right)$ would be valid (although it might not solve the problem). **No derivation needed.** **(4 marks)**

$\phi(x, y) =$

Q3. (Axis-aligned DT) Melbo wants to solve a binary classification problem with two labels **Y** and **B** and 2D features using a Decision Tree. We can only ask two kinds of split questions at any non-leaf node. Either we can ask whether $x > 0$ or not. Or else we can ask whether $y > 0$ or not. **Do not use $x < 0$ or $y < 0$ as questions as it will flip your tree.** No data point has either $x = 0$ or $y = 0$. The distribution of data points of the two labels is given on the right e.g. 1 Y and 5 B points have both $x < 0$ and $y > 0$, no point has both $x < 0$ and $y < 0$ etc. Note that the actual coordinates don't matter as we can only ask if a point lies above or below the x-axis, or to the left or right of the y-axis. The tree has 3 levels (one root, its two children and four leaves). Help Melbo construct this DT using entropy minimization. **No derivations needed**. You may use $\log_2 3 = 1.585, \log_2 5 = 2.322, \log_2 7 = 2.807$. **(2+2+2+10+4=20 marks)**



1. What is the entropy at the root note? Write the answer to 2 decimal places. **2 marks**

$H(\text{root}) =$

2. If we split the root using a horizontal split i.e. ask $y > 0$, what is the entropy of this split? (2 decimal places). **Note**: we need the entropy of the split not entropy of any child. **2 marks**

$\mathrm{H}(y > 0) =$

3. If we split the root using a vertical split i.e. ask $x > 0$, what is the entropy of this split? (2 decimal places). **Note**: we need the entropy of the split not entropy of any child. **2 marks**

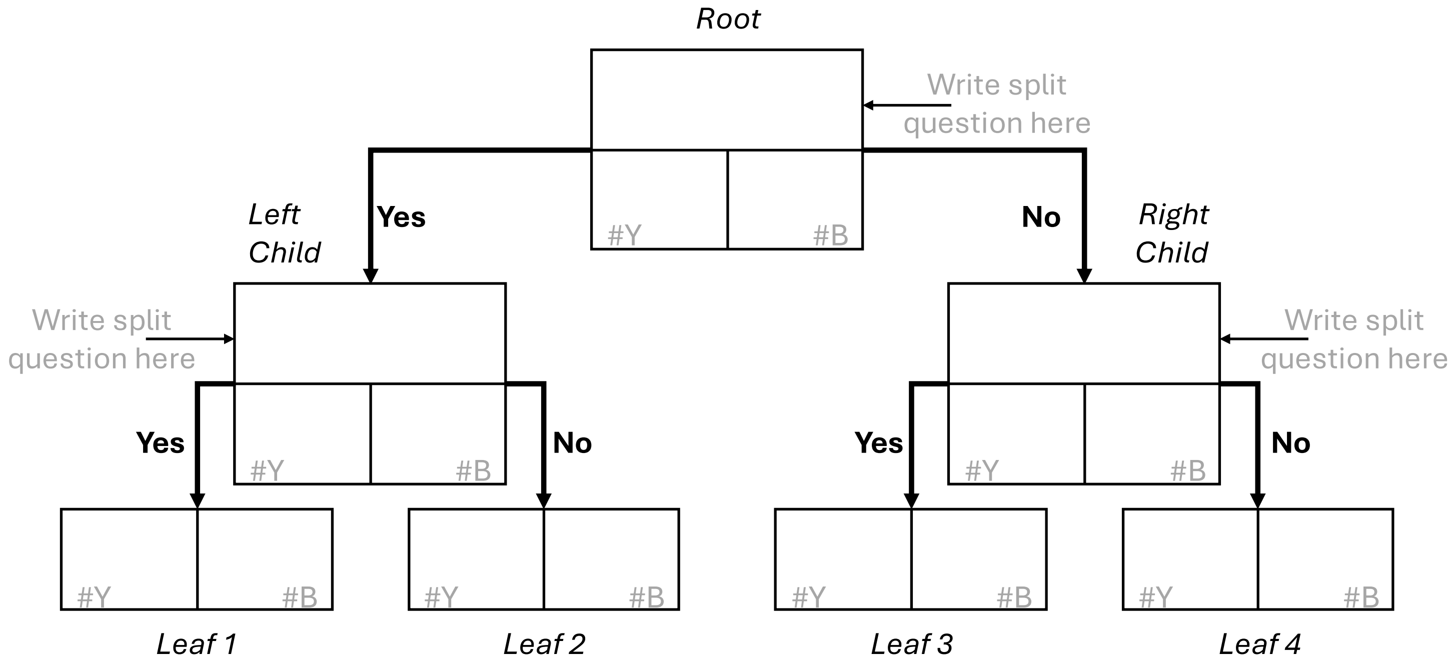
$\mathrm{H}(x > 0) =$

4. In the diagram below, indicate the split question that Melbo should use at the root to get maximum entropy reduction for the root split, as well as the split questions that Melbo should use at the left child and right child of the root. For the root, left child, right child and all 4 leaf nodes, write down, in the space provided in the diagram, how many **Y** and **B** points reached them according to your solution. Pay close attention to which child corresponds to the **Yes** answer to a split question and which child corresponds to the **No** answer to a split question. Mistakes may not get partial marks. **2 x 3 + 1 x 4 = 10 marks**

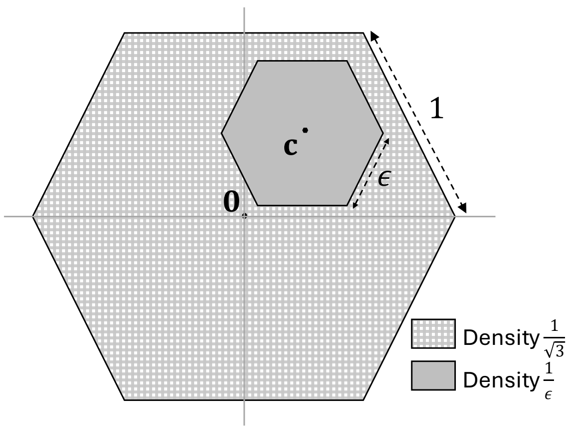
5. In the space below, write down the entropy after leaves have been created (correct to 2 decimal places). **Note**: we need the entropy of the entire set of 4 leaves and not the entropy of the leaves separately. **Show brief derivation.** **1 + 3 = 4 marks**

H(leaves) =

Give brief derivation here



Q4 (Bestagon Distribution) Melbae has a distribution $\mathcal{D}$ with support over 2D vectors lying inside an equilateral hexagon centered at the origin with all side lengths $= 1$. $\mathcal{D}$ is described by two parameters $\mathbf{c} \in \mathbb{R}^2$ and $\epsilon \in [0,1]$. The density for $\mathcal{D}$ is $\frac{1}{\epsilon}$ in the *dense* equilateral hexagon of side length ϵ centered at $\mathbf{c}$. The density is $\frac{1}{\sqrt{3}}$ in the rest of the support. Assume $\mathbf{c}$ lies within an equilateral hexagon of side length $1 - \epsilon$ i.e., the dense region stays inside the support i.e. inside the bigger hexagon.



a. For which values of ϵ will $\mathcal{D}$ be a proper distribution? **Find them and show calculations**.
b. Find out the mean vector $\boldsymbol{\mu} \in \mathbb{R}^2$ of this distribution. **Show calculations**. **(4 + 6 = 10 marks)**

Hint: The mean of the uniform distribution over an equilateral hexagon is its centre. The area of an equilateral triangle of side s *is* $\frac{\sqrt{3}}{4}s^2$.

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

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