

CS 771A: Intro to Machine Learning, IIT Kanpur				Midsem Exam (18 Jun 2023)	
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 (Optimal DT) Melbo has a multiclass problem with three classes $+$, \times , \square . There are 16 datapoints in total, each with a 2D feature vector (x, y) . x, y can take value 0 or 1. The table below describes each data point. All 16 points are at the root of a decision tree. Melbo wishes to learn a decision stump based on the entropy reduction principle to split this node into two children. Help Melbo finish this task. *Hint: take logs to base 2 so no need for calculator 😊.* **(8 x 0.5 = 4 marks)**

SNo	Class	(x, y)	SNo	Class	(x, y)	SNo	Class	(x, y)	SNo	Class	(x, y)
1	+	(0,1)	5	+	(0,1)	9	\times	(1,0)	13	\square	(1,0)
2	+	(1,1)	6	+	(0,1)	10	\times	(1,0)	14	\square	(0,0)
3	+	(0,1)	7	+	(1,1)	11	\times	(0,0)	15	\square	(1,0)
4	+	(1,1)	8	+	(1,1)	12	\times	(0,0)	16	\square	(0,0)

What is the entropy of the root node?

What is the entropy of the two child nodes (give answers for the two nodes separately) if the split is done using the x feature ($x = 0$ becomes left child, $x = 1$ becomes right child)?

What is the reduction in entropy (i.e., $H_{\text{root}} - H_{\text{children}}$) if the split is done using the x feature as described above?

What is the entropy of the two child nodes (give answers for the two nodes separately) if the split is done using the y feature ($y = 0$ becomes left child, $y = 1$ becomes right child)?

What is the reduction in entropy (i.e., $H_{\text{root}} - H_{\text{children}}$) if the split is done using the y feature as described above?

To get the most entropy reduction, should we split using x feature or y feature?

Q2. Write T or F for True/False in the box. Also, give justification. **(4 x (1+3) = 16 marks)**

1	Recall that $\ \mathbf{v}\ _0$ is the number of non-zero coordinates of the vector \mathbf{v} . Then for any two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$, we always have $\ \mathbf{a} + \mathbf{b}\ _0 \leq \ \mathbf{a}\ _0 + \ \mathbf{b}\ _0$. If true, give a brief proof, else give a counterexample of two 3D vectors that violate this inequality. Show brief calculations in either case.	
---	--	--

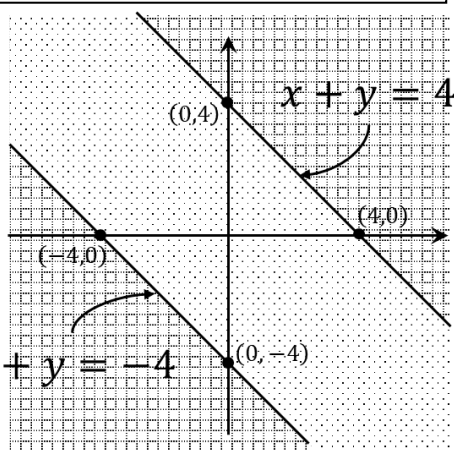
2	The function $N(p) \stackrel{\text{def}}{=} p \ln p$ is convex over the interval $p \in (0, \infty)$. If true, give a proof via chord definition or tangent definition or second derivative definition of convex functions. If false, give a counter example that violates any one definition.	
3	The optimum for $\underset{\mathbf{w} \in \mathbb{R}^d}{\text{argmin}} \ \mathbf{w} - \mathbf{w}_0\ _2^2 + \ X\mathbf{w} - \mathbf{y}\ _2^2$ is always achieved at \mathbf{w}_0 . Justify by deriving the optimum. $X \in \mathbb{R}^{n \times d}$, $\mathbf{y} \in \mathbb{R}^n$, $\mathbf{w}_0 \in \mathbb{R}^d$ are all constants.	
4	For any two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$, we always have $\mathbf{u}^\top \mathbf{v} \leq (\ \mathbf{u}\ _2^2 + \ \mathbf{v}\ _2^2)/2$. If true, give a brief proof, else give a counter example and calculations with two 3D vectors.	

Q3 (Absolute Tilt) Consider the optimization problem $\min_{x \in \mathbb{R}} f(x)$ with objective $f: \mathbb{R} \rightarrow \mathbb{R}$ defined as $f(x) \stackrel{\text{def}}{=} |x| + a \cdot x$ where $a \in \mathbb{R}$ is a constant (maybe pos/neg/zero). Find the point x^* at which the optimum is achieved and $f(x^*)$. **Note:** x^* and $f(x^*)$ depend on a . Both $x^*, f(x^*)$ can be ∞ or $-\infty$ for certain cases. You must tell us for each possible case, where is the optimum achieved i.e., x^* and what is $f(x^*)$. E.g., you might say that case 1 is $a < 1$, in which case we get $f(x^*) = 1$ at $x^* = 0.5$, and case 2 is $a \geq 1$, in which case we get $f(x^*) = -1$ at $x^* = \infty$. You may use at most 3 cases to describe your solution. If you don't need those many cases, leave cases blank. Give brief derivations. *Hint: you should not have to derive the dual to solve this problem.* **(8 marks)**

Case No.	Case Condition (write condition such as $a < 1$ or $a = 1$ etc).	Point x^* where opt. is reached for this case	Optimal objective value $f(x^*)$ for this case
1			
2			
3			

Give brief derivation below.

Q4. (Parallel Classifier) Create a feature map $\phi: \mathbb{R}^2 \rightarrow \mathbb{R}^D$ for some $D > 0$ so that for any $\mathbf{z} = (x, y) \in \mathbb{R}^2$, $\text{sign}(\mathbf{1}^\top \phi(\mathbf{z}))$ takes value +1 if \mathbf{z} is in the dark cross-hatched region and -1 if \mathbf{z} is in the light dotted region (see fig). E.g., $(0,0)$ is labelled -1 while the points $(2,5)$ and $(-6,1)$ are both labelled +1. The lines in the figure are $x + y = 4$ and $x + y = -4$. We don't care what values are taken on points lying on these two lines (as these are the decision boundaries). $\mathbf{1} = (1, 1, \dots, 1) \in \mathbb{R}^D$ is the all-ones vector. No need for derivation – just give the final map below. **(5 marks)**



$$\phi(x, y) =$$

Q5 (CM to the rescue) Consider the following problem where $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^d$ and $\lambda \in \mathbb{R}$ are constants and $\lambda > 0$. Design a coordinate minimization algorithm (choose coordinates cyclically) to solve the primal. Give brief calculations on how you will create a simplified unidimensional problem for a chosen coordinate $i \in [d]$ and then show how to get the optimal value of x_i . **(7 marks)**

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{R}^d} \quad & \frac{1}{2} \|\mathbf{x}\|_2^2 + \mathbf{a}^\top \mathbf{x} \\ \text{s. t.} \quad & \mathbf{b}^\top \mathbf{x} \leq \lambda \\ & \mathbf{c}^\top \mathbf{x} \leq \lambda \end{aligned}$$