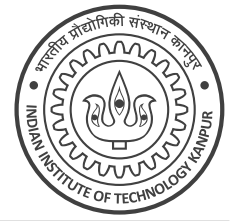


Instructions:

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 (Total Confusion) Melbu learnt a linear model to solve a binary classification problem with two classes $-1, 1$ as $\text{sign}(\mathbf{w}^\top \mathbf{x} + b)$ with $\mathbf{w} \in \mathbb{R}^{100}$ and $b \in \mathbb{R}$. The classifier was evaluated on 10 test data points (\mathbf{x}^i, y^i) that gave the confusion matrix on the right. y is

	$\hat{y} = 1$	$\hat{y} = -1$
$y = 1$	1	1
$y = -1$	7	1

the true label of a test point and \hat{y} is the label predicted by the classifier. The entries in the matrix show how many points of a given class were classified in a certain way by the classifier (e.g. 7 points whose true label was $y = -1$ were (mis)predicted as $\hat{y} = 1$). Calculate the following quantities for the classifier based on its test performance (no derivations needed) **(4 x 0.5 + 2 + 2 = 6 marks)**

Accuracy $\mathbb{P}[\hat{y} = y]$		False discovery rate $\mathbb{P}[y \neq 1 \hat{y} = 1]$	
False Omission Rate $\mathbb{P}[y = 1 \hat{y} = -1]$		Neg. predictive value $\mathbb{P}[\hat{y} = y \hat{y} = -1]$	

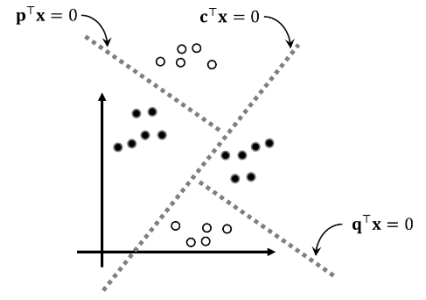
The table below shows the classifier's prediction scores $\mathbf{w}^\top \mathbf{x}^i + b$ on the 10 test points. Melbu wants to change the model parameters \mathbf{w}, b to improve the test accuracy on these 10 points. Retraining the model or changing the training algorithm is not allowed. The test feature vectors $\mathbf{x}^i \in \mathbb{R}^{100}, i \in [10]$ are not available either. All we are allowed to do is make simple changes directly to the model parameters \mathbf{w}, b learnt by Melbu (e.g. scale or shift them). Help Melbu achieve this goal. What is the best test accuracy you get after the modifications? Briefly justify.

$\mathbf{w}^\top \mathbf{x}^i + b$	-3	-1	1	3	5	7	9	11	13	15
true label y	-1	+1	-1	-1	-1	-1	-1	-1	-1	+1

Test accuracy of modified classifier:

Give details of modifications below

Q2 (Probabilistic DT) Melbo wants to solve a binary classification problem using two classification models $\mathbf{p}, \mathbf{q} \in \mathbb{R}^d$. A classifier $\mathbf{c} \in \mathbb{R}^d$ decides which model to use at test time (see figure). For a point $\mathbf{x} \in \mathbb{R}^d$, if $\mathbf{c}^\top \mathbf{x} \geq 0$, Melbo will predict $\text{sign}(\mathbf{p}^\top \mathbf{x})$. If $\mathbf{c}^\top \mathbf{x} < 0$, predict $\text{sign}(\mathbf{q}^\top \mathbf{x})$. Bias terms are hidden inside the models. Note that this is simply a decision tree with one root and two leaves. **(4 x 4 = 16 marks)**



Melbo has train data $(\mathbf{x}^i, y^i), i \in [N]$ with $\mathbf{x}^i \in \mathbb{R}^d, y^i \in \mathbb{R}$ but doesn't know which model, \mathbf{p} or \mathbf{q} , should handle which point, so Melba advises using latent variables. For each point $i \in [N]$, Melbo uses latent variables $z^i \in \{-1, +1\}$, a naïve prior $\mathbb{P}[z^i | \mathbf{p}, \mathbf{q}, \mathbf{c}] = 0.5$ and (conditional) likelihood functions $\mathbb{P}[z^i | \mathbf{x}^i, \mathbf{p}, \mathbf{q}, \mathbf{c}] = \sigma(z^i \cdot \mathbf{c}^\top \mathbf{x}^i)$ and $\mathbb{P}[y^i | z^i = +1, \mathbf{x}^i, \mathbf{p}, \mathbf{q}, \mathbf{c}] = \sigma(y^i \cdot \mathbf{p}^\top \mathbf{x}^i)$ and $\mathbb{P}[y^i | z^i = -1, \mathbf{x}^i, \mathbf{p}, \mathbf{q}, \mathbf{c}] = \sigma(y^i \cdot \mathbf{q}^\top \mathbf{x}^i)$, where $\sigma(t) \stackrel{\text{def}}{=} \frac{1}{1 + \exp(-t)}$ is the sigmoid function.

Derive an expression for total likelihood $\mathbb{P}[y^i | \mathbf{x}^i, \mathbf{p}, \mathbf{q}, \mathbf{c}]$ in terms of $y^i, \mathbf{x}^i, \mathbf{p}, \mathbf{q}, \mathbf{c}$ (no z^i allowed).

As exact MLE is hard, Melbo instead tries to solve $\arg\max_{\mathbf{p}, \mathbf{q}} \arg\max_{\{z^i\}} \arg\max_{\mathbf{c}} \{\mathcal{L}(\mathbf{p}, \mathbf{q}, \mathbf{c}, \{z^i\})\}$ with $\mathcal{L}(\mathbf{p}, \mathbf{q}, \mathbf{c}, \{z^i\}) \stackrel{\text{def}}{=} \sum_{i \in [N]} \ln(\mathbb{P}[y^i | z^i, \mathbf{x}^i, \mathbf{p}, \mathbf{q}, \mathbf{c}]) + \sum_{i \in [N]} \ln(\mathbb{P}[z^i | \mathbf{x}^i, \mathbf{p}, \mathbf{q}, \mathbf{c}])$ using alternating optimization. **You are free to use simple operations like least squares, logistic regression directly.**

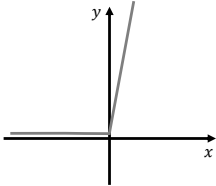
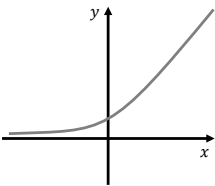
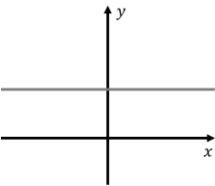
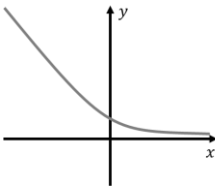
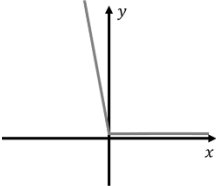
Step 1: Freeze $\mathbf{c}, \{z^i\}$ and give brief derivation on how to find $\arg\max_{\mathbf{p}, \mathbf{q}} \mathcal{L}(\mathbf{p}, \mathbf{q}, \mathbf{c}, \{z^i\})$.

CS 771A: Intro to Machine Learning, IIT Kanpur			Endsem Exam (18 Nov 2025)	
Name				40 marks Page 3 of 4
Roll No		Dept.		

Step 2: Freeze $\mathbf{p}, \mathbf{q}, \mathbf{c}$ and give brief derivation on how to find $\operatorname{argmax}_{\{z^i\}} \mathcal{L}(\mathbf{p}, \mathbf{q}, \mathbf{c}, \{z^i\})$.

Step 3: Freeze $\mathbf{p}, \mathbf{q}, \{z^i\}$ and give brief derivation on how to find $\operatorname{argmax}_{\mathbf{c}} \mathcal{L}(\mathbf{p}, \mathbf{q}, \mathbf{c}, \{z^i\})$.

Q3 (Rapidly Rising ReLUs) The ReLU activation becomes more expressive if used with a sharpness parameter B as $\rho(x; B) \stackrel{\text{def}}{=} \ln((1 + \exp(-B \cdot x)))$. For each of the following five curves, select the value of B that best generates that curve. **Shade only one circle in each part.** (5 x 1 = 5 marks)

 <p>(a)</p> <div> <input type="radio"/> $B \rightarrow -\infty$ <input type="radio"/> $B = -1$ <input type="radio"/> $B = 0$ <input type="radio"/> $B = +1$ <input type="radio"/> $B \rightarrow +\infty$ </div>	 <p>(b)</p> <div> <input type="radio"/> $B \rightarrow -\infty$ <input type="radio"/> $B = -1$ <input type="radio"/> $B = 0$ <input type="radio"/> $B = +1$ <input type="radio"/> $B \rightarrow +\infty$ </div>	 <p>(c)</p> <div> <input type="radio"/> $B \rightarrow -\infty$ <input type="radio"/> $B = -1$ <input type="radio"/> $B = 0$ <input type="radio"/> $B = +1$ <input type="radio"/> $B \rightarrow +\infty$ </div>	 <p>(d)</p> <div> <input type="radio"/> $B \rightarrow -\infty$ <input type="radio"/> $B = -1$ <input type="radio"/> $B = 0$ <input type="radio"/> $B = +1$ <input type="radio"/> $B \rightarrow +\infty$ </div>	 <p>(e)</p> <div> <input type="radio"/> $B \rightarrow -\infty$ <input type="radio"/> $B = -1$ <input type="radio"/> $B = 0$ <input type="radio"/> $B = +1$ <input type="radio"/> $B \rightarrow +\infty$ </div>
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Q4 (The Perils of Pollution) Melba is studying the effect of factory output on pollution levels. It is known that if the factory output is p and the pollution probe is at distance q from the factory, then the pollution level measured by the probe is p/q . Melba conducted experiments with 4 output levels l, m, n, o and for each experiment, Melba placed the

	l	m	n	o
d	12	x_{dm}	x_{dn}	x_{do}
e	30	40.5	x_{en}	x_{eo}
f	x_{fl}	x_{fm}	1.25	x_{fo}
g	x_{gl}	x_{gm}	x_{gn}	4.5

probe at 4 distances d, e, f, g from the factory, thus getting 16 readings. Melba recorded the readings in the above matrix but by mistake, Melba spilled coffee on the spreadsheet causing some of the entries (labeled x_{fl}, x_{gn} etc in gray) to get erased. Melba is in panic as not only did a lot of data get erased, but the values l, m, n, o, d, e, f, g used to conduct experiments are also gone. Help find these values so that Melba can repeat the experiment. It is known that l, m, n, o, d, e, f, g are all positive integers, $d + e + f + g = 50$ and $f < g$. Melba also recalls that if we arrange the readings in the first column as a vector $\mathbf{v} = [12, 30, x_{fl}, x_{gl}] \in \mathbb{R}^4$, then $\|\mathbf{v}\|_2^2 = 1169$, $\|\mathbf{v}\|_1 = 57$. Give brief derivation on how you obtained l, m, n, o, d, e, f, g . (8 + 5 = 13 marks)

$l =$

$d =$

$m =$

$e =$

$n =$

$f =$

$o =$

$g =$

Give brief derivation here