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

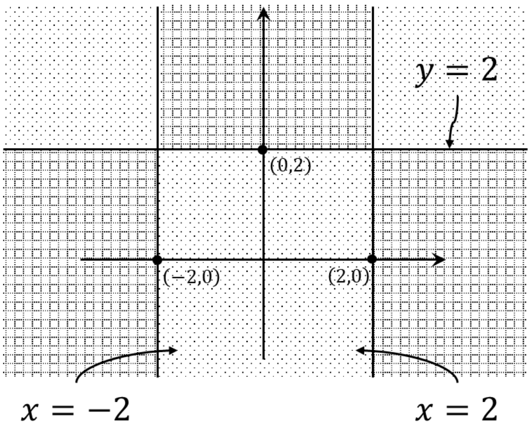
1	All stationary points of the function $f(x) \stackrel{\text{def}}{=} x^3 - x^5$ are either local/global minima or local/global maxima. Justify your answer using first and second derivative tests.	

2	Let $f: \mathbb{R} \to \mathbb{R}$ be a convex differentiable function. Let $g \stackrel{\text{def}}{=} -f$ i.e., $g(x) = -f(x)$ for all $x \in \mathbb{R}$. Then g can never be convex. Give either a proof or a counter example.	

3	The optimum for $\underset{x \in \mathbb{R}}{\arg\min} \exp(\lvert x - x_0 \rvert) + (x - x_0)^2$ is always x_0. Justify by deriving the optimum. Note that $x_0 \in \mathbb{R}$ is a constant. (*Hint*: using calculus may be tricky)	

4	The dot product of two **Boolean** vectors $\mathbf{u}, \mathbf{v} \in \{0,1\}^3$ cannot be zero unless one of them is the zero vector. If true, give a brief proof, else give a counter example.

Q2. (Chessboard Classifier) Create a feature map $\phi: \mathbb{R}^2 \to \mathbb{R}^D$ for some $D > 0$ so that for any $\mathbf{z} = (x, y) \in \mathbb{R}^2$, $\text{sign}\left(\mathbf{1}^\top \phi(\mathbf{z})\right)$ 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., $(0,0), (3,3), (-3,3)$ are all labelled $+1$ while $(-3,0), (0,3), (3,0)$ are all labelled -1. The lines in the figure are $x = 2$, $x = -2$ and $y = 2$. We don't care what label is given to points lying on the three lines (these are the decision boundaries). $\mathbf{1} = (1,1, \dots ,1) \in \mathbb{R}^D$ is the all-ones vector. No need for derivation – give only the final map below.**(5 marks)**



$\phi(x, y) =$

Q3 (Optimal Checkerboard DT) Melbo has received data for the problem in Q2. There are 10 datapoints (given in the table), each with a 2D feature vector (x, y). All 10 points are at the root of a decision tree. Melbo wants to learn a decision stump based on the entropy reduction principle to split the root into two children. Only 3 decision stumps are allowed which ask the questions $(x \leq -2?), (x \leq 2?)$ and $(y \leq 2?)$. **All logs are to base 2, assume $\log_2 3 = 1.58, \log_2 5 = 2.32$ Give your answers correct to at least 2 decimal places**. **(11 x 1 = 11 marks)**

S.	Class	(x,y)	S.	Class	(x,y)	S.	Class	(x,y)	S.	Class	(x,y)	S.	Class	(x,y)
1	$-$	$(-3,0)$	3	$+$	$(1,1)$	5	$+$	$(-1,1)$	7	$-$	$(1,5)$	9	$-$	$(-1,5)$
2	$+$	$(3,3)$	4	$+$	$(1,-1)$	6	$+$	$(-1,-1)$	8	$-$	$(1,3)$	10	$-$	$(-1,3)$

What is the entropy of the root node?

What is the entropy of the two child nodes (answers for the two nodes separately) if split is done using the question $x \leq -2?$ i.e., $x \leq -2$ becomes the left child, $x > -2$ becomes right child)?

What is the reduction in entropy (i.e., $H_{\text{root}} - H_{\text{children}}$) if the split is done using the question $x \leq -2?$ as described above?

What is the entropy of the two child nodes (answers for the two nodes separately) if split is done using the question $x \leq 2$? i.e., $x \leq 2$ becomes the left child, $x > 2$ becomes right child)?

What is the reduction in entropy (i.e., $H_{\text{root}} - H_{\text{children}}$) if the split is done using the question $x \leq 2$? as described above?

What is the entropy of the two child nodes (answers for the two nodes separately) if split is done using the question $y \leq 2$? i.e., $y \leq 2$ becomes the left child, $y > 2$ becomes right child)?

What is the reduction in entropy (i.e., $H_{\text{root}} - H_{\text{children}}$) if the split is done using the question $y \leq 2$? as described above?

To get the most entropy reduction, which decision stump should we use?

Q4 (Tables are turned). A curious type of regularization is *Morozov regularization* which turns the loss function into a constraint (btw, SVMs & ridge regression use *Tikhonov regularization* instead). Consider the following regression problem where $X \in \mathbb{R}^{N \times d}$ gives us d-dimensional features for N data points and $\mathbf{y} \in \mathbb{R}^N$ gives the labels. Give 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 w_i. Assume $\|\mathbf{y}\|_2^2 \leq 1$ so that the constraint set is not empty (e.g., $\mathbf{w} = \mathbf{0}$ satisfies the constraint). Feel free to define shorthand notation to simplify your answer. **(8 marks)**

$$\begin{aligned} \min_{\mathbf{w} \in \mathbb{R}^d} \quad & \frac{1}{2}\|\mathbf{w}\|_2^2 \\ \text{s.t.} \quad & \|X\mathbf{w} - \mathbf{y}\|_2^2 \leq 1 \end{aligned}$$