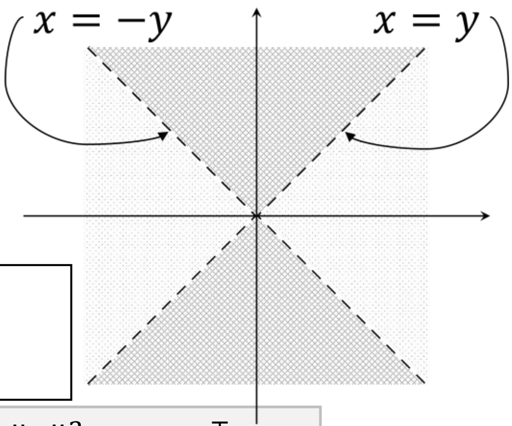
Instructions:
1. This question paper contains 2 pages (4 sides of paper). Please verify.
2. Write your name, roll number, department in **block letters** with **ink** on **each page**.
3. Write your final answers neatly **with a blue/black pen**. Pencil marks may get smudged.
4. Don't overwrite/scratch 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+2) = 12 marks)**

1	For $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^2$ s.t. $\|\mathbf{x}\|_2 = \|\mathbf{y}\|_2 = \sqrt{2}$, $\|\mathbf{z}\|_2 = 1$ and $\mathbf{x}^\top \mathbf{y} \geq \mathbf{x}^\top \mathbf{z}$, we always have $\|\mathbf{x} - \mathbf{y}\|_2^2 \leq \|\mathbf{x} - \mathbf{z}\|_2^2$. Give a brief proof if True else give a counter example if False.	
2	Let $f, g: \mathbb{R} \to \mathbb{R}$ be two distinct, non-constant, convex functions i.e., $f \neq g$ and it is not the case that for some $c, d \in \mathbb{R}$, $f(x) = c, g(x) = d$ for all $x \in \mathbb{R}$. Then $h: \mathbb{R} \to \mathbb{R}$ defined as $h(x) \stackrel{\text{def}}{=} f(x)/g(x)$ can never be convex. Give a brief proof if True else if False, give a counter example using two distinct non-constant, convex functions. It is okay to give a counter example where h has isolated, removable discontinuities.	
3	X is a discrete random variable that takes value -1 with probability p and 1 with probability $1 - p$. The value of p at which X has maximum entropy is the same as the value of p at which X has maximum variance.	

4	Y is a Boolean random variable $\mathbb{P}[Y=1] = 1/(1+\exp(-t))$. Then Y's entropy is maximized as $t \to \infty$. Justify your answer by giving brief calculations.	

Q2. (X marks the split) Create a feature map $\phi\colon \mathbb{R}^2 \to \mathbb{R}^D$ for some $D > 0$ so that for any $\mathbf{z} = (x, y) \in \mathbb{R}^2$, $\operatorname{sign}\left(\mathbf{1}^\top \phi(\mathbf{z})\right)$ 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). $\mathbf{1} = (1,1,\ldots,1) \in \mathbb{R}^D$ is the D-dimensional all-ones vector. The dashed lines in the fig are $x = y$ and $x = -y$. No derivation needed – just give the final map below. **(3 marks)**



$\phi(x, y) =$

Q3. (Maximum stretch) Consider the optimization problem $\min_{\mathbf{x}\in\mathbb{R}^3} \frac{1}{2}\|\mathbf{x}\|_2^2 \quad \text{s.t.} \quad \mathbf{c}^\top\mathbf{x} \geq p$ which has a single constraint and $\mathbf{c} \in \mathbb{R}^3$ is a constant vector and $p \in \mathbb{R}$ is a real constant. **(3+2 = 5 marks)**

(a) Give brief derivation solving the problem for $\mathbf{c} = (1,2,3)$ and $p = 7$. Write the value of $\mathbf{x}$ at which the optimum is achieved. (*Hint: try orthogonal decomposition or some other trick*)

(b) Give brief derivation solving the problem for $\mathbf{c} = (-1,-2,-3)$ and $p = -7$. Write the value of $\mathbf{x}$ at which the optimum is achieved.

Q4 (Elastic-net regression) Given n pts $(\mathbf{x}^i, y^i)$ $\mathbf{x}^i \in \mathbb{R}^d, y^i \in \mathbb{R}$, we wish to solve $\min_{\mathbf{w}\in\mathbb{R}^d} \frac{1}{2}\|\mathbf{w}\|_2^2 + \|\mathbf{w}\|_1 + \frac{1}{2}\sum_{i\in[n]}(y^i - \mathbf{w}^\top\mathbf{x}^i)^2$. To create its dual, we introduce variables $\mathbf{z} = [z_1, \dots, z_d] \in \mathbb{R}^d$ and $\mathbf{r} = [r_1, \dots, r_n] \in \mathbb{R}^n$ to give us the constrained problem in the box on the right. Note that $\mathbf{1} \in \mathbb{R}^d$ is the all-ones vector.

$$\min_{\substack{\mathbf{w},\mathbf{z}\in\mathbb{R}^d \\ \mathbf{r}\in\mathbb{R}^n}} \frac{1}{2}\|\mathbf{w}\|_2^2 + \mathbf{z}^\top\mathbf{1} + \frac{1}{2}\|\boldsymbol{r}\|_2^2 \quad \text{s.t.}$$
$$w_j - z_j \leq 0 \text{ for all } j \in [d]$$
$$-w_j - z_j \leq 0 \text{ for all } j \in [d]$$
$$y^i - \mathbf{w}^\top\mathbf{x}^i - r_i = 0 \text{ for all } i \in [n]$$

We introduce dual variables α_j for the constraints $w_j - z_j \leq 0$, β_j for $-w_j - z_j \leq 0$ and λ_i for $y^i - \mathbf{w}^\top\mathbf{x}^i - r_i = 0$. For simplicity, we collect the dual variables as vectors $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^d$ and $\boldsymbol{\lambda} \in \mathbb{R}^n$. For each part, give your answers in the space demarcated for that part. **(3+2+6+5+4=20 marks)**

a. \`Fill in the circle indicating the correct constraint for the dual variables $\alpha_j, \beta_j, \lambda_i$. *(3x1 marks)*

$\alpha_j \leq 0$ ○	$\beta_j \leq 0$ ○	$\lambda_i \leq 0$ ○
$\alpha_j \geq 0$ ○	$\beta_j \geq 0$ ○	$\lambda_i \geq 0$ ○
No constraint on α_j ○	No constraint on β_j ○	No constraint on λ_i ○

b. Write down the Lagrangian $\mathcal{L}(\mathbf{w}, \mathbf{z}, \mathbf{r}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda})$ – no derivation needed. *(2 marks)*

c. The dual problem is $\max_{\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\lambda}} \left\{\min_{\mathbf{w},\mathbf{z},\mathbf{r}} \mathcal{L}(\mathbf{w}, \mathbf{z}, \mathbf{r}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\lambda})\right\}$. To simplify it, solve the 3 inner problems $\min_{\mathbf{w}} \mathcal{L}$, $\min_{\mathbf{z}} \mathcal{L}$ and $\min_{\mathbf{r}} \mathcal{L}$. In each case, give brief derivation and write the expression you get while solving the inner problem (e.g., in CSVM $\min_{\mathbf{w}}\mathcal{L}$ gives $\mathbf{w} = \sum_i \alpha_i y^i \mathbf{x}^i$). *(3x(1+1) marks)*

Expression + derivation for $\min_{\mathbf{w}} \mathcal{L}$.

Expression + derivation for $\min_{\mathbf{z}} \mathcal{L}$.

Expression + derivation for $\min_{\mathbf{r}} \mathcal{L}$.

d. Use the expressions obtained above and eliminate $\boldsymbol{\beta}$. Fill in the 5 blank boxes below to show us the simplified dual you get. $X \in \mathbb{R}^{n \times d}$ is the feature matrix with the ith row being $\mathbf{x}^i$. We have turned the max dual problem into a min problem by negating the objective. *(5x1 marks)*

$$\min_{\substack{\boldsymbol{\alpha} \in \mathbb{R}^d \\ \boldsymbol{\lambda} \in \mathbb{R}^n}} \frac{1}{2} \left\| X^\top(\quad\quad) + (\quad\quad) \right\|_2^2 + \frac{1}{2} \|\boldsymbol{\lambda}\|_2^2 - \boldsymbol{\lambda}^\top(\quad\quad)$$

s.t.

⟸ Write constraint for $\boldsymbol{\alpha}$ here.

⟸ Write constraint for $\boldsymbol{\lambda}$ here.

e. For the simplified dual obtained above, let us perform block coordinate minimization.

1. For any fixed value of $\boldsymbol{\alpha} \in \mathbb{R}^{\boldsymbol{d}}$, obtain the optimal value of $\boldsymbol{\lambda} \in \mathbb{R}^{\boldsymbol{n}}$.
2. For any fixed value of $\boldsymbol{\lambda} \in \mathbb{R}^{\boldsymbol{n}}$, obtain the optimal value of $\boldsymbol{\alpha} \in \mathbb{R}^{\boldsymbol{d}}$.

Note: the optimal value for a variable must satisfy its constraints (if any). Show brief calculations. You may use the QUIN trick and invent shorthand notation to save space e.g., $\mathbf{m} \stackrel{\text{def}}{=} X\boldsymbol{\alpha}$.*(2+2 marks)*