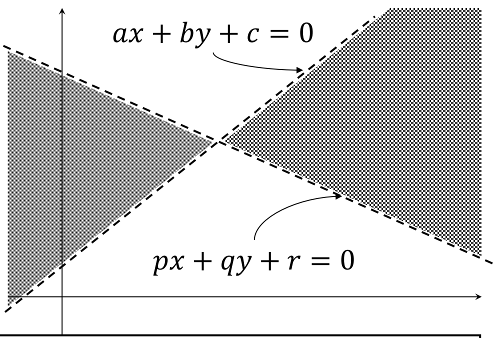
CS 771A: Intro to Machine Learning, IIT Kanpur				**Midsem Exam** (22 Feb 2024)
Name				**40 marks**
Roll No		Dept.		Page **1** of **4**

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.

Q1a. (Another X marks the split) Given lines $ax + by + c = 0$ and $px + qy + r = 0$, create a feature map $\phi: \mathbb{R}^2 \to \mathbb{R}^D$ for $D > 0$ and a classifier $\mathbf{W} \in \mathbb{R}^D$ so that for any $\mathbf{z} = (x, y) \in \mathbb{R}^2$, we have $\text{sign}\left(\mathbf{W}^\top \phi(\mathbf{z})\right) = \text{sign}(ax + by + c) \cdot \text{sign}(px + qy + r)$. Your map ϕ must **not depend on** a, b, c, p, q, r but your classifier $\mathbf{W}$ may depend on a, b, c, p, q, r. No derivation needed – just give the final map and classifier below. **(2 + 2 = 4 marks)**



$\phi(x, y) =$

$\mathbf{W} =$

Q1b. (… except when it doesn't) In the figure, the two lines divide $\mathbb{R}^2$ into two regions labeled $+1$ and two labeled -1. Melbo warns us that this may not always be true and there may exist cases where the label is same almost everywhere. Assume $\text{sign}(0) = 0$. Find values of a, b, c, p, q, r so that $\text{sign}(ax + by + c) \cdot \text{sign}(px + qy + r) = +1$ or 0 for every $\mathbf{z} = (x, y) \in \mathbb{R}^2$. Next, find values of a, b, c, p, q, r so that $\text{sign}(ax + by + c) \cdot \text{sign}(px + qy + r) = -1$ or 0 for every $\mathbf{z} \in \mathbb{R}^2$. Note that all six values i.e., a, b, c, p, q, r in your responses must be non-zero. Moreover, you must ensure that $a \neq p, b \neq q, c \neq r$ in your responses. No derivations needed. **(2 + 2 = 4 marks)**

Your answer for the case when label is $+1$ or 0 everywhere

$a =$ ☐ $b =$ ☐ $c =$ ☐ $p =$ ☐ $q =$ ☐ $r =$ ☐

Your answer for the case when label is -1 or 0 everywhere

$a =$ ☐ $b =$ ☐ $c =$ ☐ $p =$ ☐ $q =$ ☐ $r =$ ☐

Q3. (Split subgradients) For 2D points $\mathbf{z} = (x, y) \in \mathbb{R}^2$, let $g(\mathbf{z}) = |x - 1|$ and $h(\mathbf{z}) = |y - 2|$.

a. Find out the subdifferentials $\partial g(\mathbf{z})$ and $\partial h(\mathbf{z})$. Show brief derivation for any one function.
b. Define $f(\mathbf{z}) \stackrel{\text{def}}{=} g(\mathbf{z}) + h(\mathbf{z})$. Find $\partial f(\mathbf{z})$ showing brief derivation. (*Hint: use sum rule*).

Note that even though g, h depend only on one variable each, they are still technically multivariate functions. Thus, both $\partial g(\mathbf{z}), \partial h(\mathbf{z})$ will be sets of vectors and not sets of scalars. **(4+4=8 marks)**

Answer to part a.

Answer to part b.

Q4 (Balanced Budget) Melbo was asked to help the finance ministry create India's budget. India has n citizens classified as rich ($y_i = -1$) or poor ($y_i = +1$). For each citizen $i \in [n]$, an amount x_i is to be decided. $x_i < 0$ means that ₹ $-x_i$ was given to citizen i as subsidy. $x_i > 0$ means a tax of ₹ x_i was demanded from that citizen. Amounts can be fractional (e.g. $x_i = 0.3414$). A few conditions must be kept in mind while deciding the allocation (as elections are coming up 😊)

$$\min_{\mathbf{x} \in \mathbb{R}^n} \frac{1}{2} \|\mathbf{x} - \mathbf{a}\|_2^2 + \mathbf{a}^\top \mathbf{x}$$

$x_i \leq +1$ for all $i \in [n]$

$-1 \leq x_i$ for all $i \in [n]$

$$\mathbf{x}^\top \mathbf{1} = D$$

$$\mathbf{x}^\top \mathbf{y} = 0$$

a. The govt wants to claim it is neither anti-poor nor anti-rich by ensuring $\sum_{i\in[n]} x_i y_i = 0$.
b. Subsidy/tax is capped at ₹ 1 i.e. $-1 \le x_i \le 1$ to avoid allegations of tax terrorism or *revdi.*
c. Despite all of the above, the govt must ensure a fiscal deficit of ₹ D i.e., $\sum_{i\in[n]} x_i = D$.

The ministry has an allocation $\mathbf{a} \in \mathbb{R}^n$ which it likes but which violates the conditions. Help Melbo find a valid allocation most aligned to $\mathbf{a}$ by solving the optimization problem in the figure. We introduce dual variables α_j for the constraints $x_i \le 1$, β_i for $-1 \le x_i$ (we collect these dual variables as vectors $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^n$), λ for $\mathbf{x}^\top \mathbf{1} = D$ and δ for $\boldsymbol{x}^\top \boldsymbol{y} = 0$. **(4+2+3+7+8=24 marks)**

a. Fill in the circle indicating the correct constraint for the dual variables. *(4x1 marks)*

$\alpha_i \le 0$ ○	$\beta_i \le 0$ ○	$\lambda \le 0$ ○	$\delta \le 0$ ○
$\alpha_i \ge 0$ ○	$\beta_i \ge 0$ ○	$\lambda \ge 0$ ○	$\delta \ge 0$ ○
No constraint ○	No constraint ○	No constraint ○	No constraint ○

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

c. To simplify the dual $\max_{\boldsymbol{\alpha},\boldsymbol{\beta},\lambda,\delta} \left\{\min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda, \delta)\right\}$, solve $\min_{\mathbf{x}} \mathcal{L}$ to get an expression for $\mathbf{x}$ in terms of $\boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda, \delta$ and constants such as $\mathbf{a}, \mathbf{y}, \mathbf{1}$ etc. No need for derivation. *(3 marks)*

d. Show us the simplified dual you get. Ignore constant terms e.g. $\|\mathbf{y}\|_2, \|\mathbf{a}\|_2$ etc. Note that we have turned the dual into a min problem by negating the objective. If a certain dual variable has no constraints, leave that box blank or write "No constraint". *(7x1 marks)*

$$\min_{\boldsymbol{\alpha},\boldsymbol{\beta},\lambda,\delta} \frac{1}{2}\left\| \qquad \right\|_2^2 - (\qquad)^\top \mathbf{1} - \lambda \cdot (\qquad)$$

s.t.

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

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

⟸ Write constraint for λ here.

⟸ Write constraint for δ here.

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

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

Show brief steps. You may use the QUIN trick or shorthand notation to save space. *(2+2+2+2 marks)*

Show steps to find δ

Show steps to find λ

Show steps to find $\boldsymbol{\beta}$

Show steps to find $\boldsymbol{\alpha}$