Mid Sem

Graded

Student

PATHE NEVISH ASHOK

Total Points

36 / 40 pts

Question 1

Another X marks the split

4 / 4 pts

- ✔ **+ 2 pts** _A valid map ϕ that depends only on x, y apart from universal constants such as $1, 2$ etc, possibly with minor mistakes

- **− 1 pt** Minor mistakes such as forgetting a constant term

- ✔ **+ 2 pts** A valid classifier $\mathbf{W}$ that does not depend on x, y at all and depends only on a, b, c, p, q, r and universal constants such as $1, 2$ etc, possibly with minor mistakes.

- **− 1 pt** Minor mistakes e.g. if the classifier fails to reproduce desired result at isolated locations e.g. on a line or at a point etc.

- **+ 0 pts** Totally wrong or else unanswered

Question 2

... except when it doesn't

4 / 4 pts

- ✔ **+ 2 pts** All values of a, b, c, p, q, r, are given (no blanks) that are non-zero, satisfy $a \neq p, b \neq q, c \neq r$ and produce the label 1 or 0 everywhere. This will be the case if $(p, q, r) = t \cdot (a, b, c)$ for some $t > 0, t \neq 1$. No partial marking.

- ✔ **+ 2 pts** All values of a, b, c, p, q, r, are given (no blanks) that are non-zero, satisfy $a \neq p, b \neq q, c \neq r$ and produce the label -1 or 0 everywhere. This will be the case if $(p, q, r) = t \cdot (a, b, c)$ for some $t < 0$. No partial marking.

- **+ 0 pts** Totally wrong or else unanswered

Question 3

Split subgradients

7 / 8 pts

3.1 part a

4 / 4 pts

- ✔ **+ 1 pt** Valid subdifferential for g possibly with minor mistakes
- ✔ **+ 1 pt** Valid subdifferential for h possibly with minor mistakes
- **− 1 pt** Minor mistakes in subdifferentials e.g. missing case or wrong sign.
- ✔ **+ 2 pts** Valid derivation for the subdifferential of at least one function with possibly minor errors
- **− 1 pt** Minor errors in derivation e.g. wrong sign.
- **+ 0 pts** Totally wrong or else unanswered

3.2 part b

3 / 4 pts

- ✔ **+ 2 pts** Valid subdifferential for f either in closed form or else in case-by-case form, possibly with minor mistakes
- ✔ **+ 2 pts** Valid derivation of the subdifferential using either the sum rule or the max rule
- ✔ **− 1 pt** Minor mistakes in the subdifferential e.g. missing cases if giving case-by-case form, wrong sign etc.
- **− 1 pt** Minor mistakes in derivation.
- **+ 0 pts** Totally wrong or else unanswered

Question 4

Balanced budget **21** / 24 pts

4.1 part a **4** / 4 pts

+ **0 pts** Give marks directly out of 4 using points adjustment. 1 mark per correct answer. Give marks even if there is overwriting if the final answer is clearly visible (since some student answers required modification due to verbal clarification given during the exam). No partial marking.

+ **0 pts** Completely wrong or else unanswered

+ **4 pts** Point adjustment

4.2 part b **2** / 2 pts

✓ + **2 pts** Valid expression for Lagrangian, possibly with minor mistakes.

− **1 pt** Minor mistakes e.g. wrong sign. Note that the sign for terms corresponding to equality constraints can be positive or negative e.g., both $+\delta \cdot \mathbf{x}^\top \mathbf{y}$ and $-\delta \cdot \mathbf{x}^\top \mathbf{y}$ are admissible.

+ **0 pts** Completely wrong or else unanswered

4.3 part c **3** / 3 pts

✓ + **3 pts** Valid expression for $\mathbf{x}$ in terms of dual variables and constants, possibly with minor mistakes. No derivation needed.

− **1 pt** Minor mistakes e.g. wrong sign

+ **0 pts** Completely wrong or else unanswered

4.4 part d **7** / 7 pts

✓ + **0 pts** Give marks directly out of 7 using points adjustment. 1 mark for each correct answer. Deduct 0.5 marks for an answer where expression is correct but sign is wrong e.g., writing $\alpha + \beta$ in the second blank in the objective instead of $-\alpha - \beta$. There should be no constraints over λ, δ.

+ **0 pts** Completely wrong or else unanswered

+ **7 pts** Point adjustment

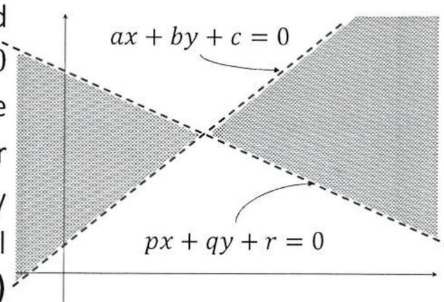
4.5 **part e** **5** / 8 pts

- ✔ **+ 2 pts** Valid rule to update δ with derivation, possibly with minor mistakes
- ✔ **+ 2 pts** Valid rule to update λ with derivation, possibly with minor mistakes
- ✔ **− 1 pt** Minor mistakes in λ update rule or derivation.
- ✔ **+ 2 pts** Valid rule to update β with derivation, possibly with minor mistakes
- ✔ **− 1 pt** Minor mistakes in β update rule or derivation.
- ✔ **+ 2 pts** Valid rule to update α with derivation, possibly with minor mistakes
- ✔ **− 1 pt** Minor mistakes in α update rule or derivation.
- **+ 0 pts** Completely wrong or else unanswered

CS 771A: Intro to Machine Learning, IIT Kanpur				**Midsem Exam** (22 Feb 2024)
Name	PATHE NEVISH ASHOK			**40 marks**
Roll No	220757	Dept.	CSE	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 $\text{sign}(\mathbf{W}^\top \phi(\mathbf{z})) = \text{sign}(ax + by + c) \cdot \text{sign}(px + qy + r)$. 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)**



$\phi(x, y) =$ $[x^2, xy, y^2, x, y, 1]$

$\mathbf{W} =$ $[ap, (aq+bp), bq, (cp+ra), (cq+rb), cr]$

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

$a = \boxed{1}$ $b = \boxed{1}$ $c = \boxed{1}$ $p = \boxed{2}$ $q = \boxed{2}$ $r = \boxed{2}$

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

$a = \boxed{1}$ $b = \boxed{2}$ $c = \boxed{3}$ $p = \boxed{-1}$ $q = \boxed{-2}$ $r = \boxed{-3}$

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

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.

$$g(z) = |x-1| = \max\{x-1, 1-x\}$$

for $x < 1$, $g(z) = 1-x$; $x > 1$, $g(z) = x-1$; $x=1$; $g(z)=0$

$$\therefore \partial g(z) = \begin{cases} \{(-1,0)\} & \text{if } x<1 \\ \{\lambda u+(1-\lambda)v\} & \text{if } x=0 \\ \{(1,0)\} & \text{if } x>1 \end{cases}$$

where $u = (-1,0)$, $v = (1,0)$ $\Rightarrow \{(1-2\lambda, 0)\}$ for $\lambda \in [0,1]$

Similarly for $\partial h(z)$

$$= \begin{cases} \{(0,-1)\} & \text{if } y<2 \\ \{(0,1-2\mu)\} & \text{if } y=2 \quad (\mu \in [0,1]) \\ \{(0,1)\} & \text{if } y>2 \end{cases}$$

Answer to part b.

$$f(z) = g(z) + h(z).$$

sum rule $\Rightarrow$ $\partial f(z) = \partial g(z) + \partial h(z)$

$$= \{g' + h' : g' \in \partial g(z),\ h' \in \partial h(z)\}$$

$$= \{(-1,0), (1,0), (1-2\lambda, 0)\} + \{(0,-1), (0,1), (0,1-2\mu)\}$$

$$= \left\{\begin{matrix} (-1,-1), & (-1,1), & (-1,1-2\mu), \\ (1,-1), & (1,1), & (1,1-2\mu), \\ (1-2\lambda,-1), & (1-2\lambda,1), & (1-2\lambda,1-2\mu) \end{matrix}\right\} = \boxed{\{(1-2\lambda, 1-2\mu)\}}$$

$\boxed{\text{where } \lambda \in [0,1],\ \mu \in [0,1]}$

(multiple vectors).

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

$$\min_{\mathbf{x}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{x}-\mathbf{a}\|_2^2 + \mathbf{a}^\top\mathbf{x}$$
$$x_i \le +1 \text{ for all } i \in [n]$$
$$-1 \le x_i \text{ for all } i \in [n]$$
$$\mathbf{x}^\top\mathbf{1} = D$$
$$\mathbf{x}^\top\mathbf{y} = 0$$

the allocation (as elections are coming up 😊)

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. $-1 \leq x_i \leq 1$ 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 $x_i \leq 1$, β_i for $-1 \leq x_i$ (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)*

$\alpha_i \leq 0$ ○	$\beta_i \leq 0$ ○	$\lambda \leq 0$ ○	$\delta \leq 0$ ○
$\alpha_i \geq 0$ ●	$\beta_i \geq 0$ ●	$\lambda \geq 0$ ○	$\delta \geq 0$ ○
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)*

$$\frac{1}{2}\|x - a\|_2^2 + a^T x + \alpha^T(x - \mathbf{1}) - \beta^T(x + \mathbf{1}) + \lambda(x^T \mathbf{1} - D) + \delta(x^T y)$$

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)*

$$x = -\alpha + \beta - \lambda \cdot \mathbf{1} \mp \delta \cdot y$$

negative (-ve sign) $\lambda, \delta \in \mathbb{R}$

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)*

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

s.t.

$\alpha \geqslant 0$ (all coordinates $\geqslant 0$)	⟸ Write constraint for $\boldsymbol{\alpha}$ here.
$\beta \geqslant 0$ (all coordit $\geqslant 0$)	⟸ Write constraint for $\boldsymbol{\beta}$ here.
No constraint	⟸ Write constraint for λ here.

$-\left(\left(\frac{1}{2}\|x\|^2\right) - a^T x + \|a\|^2 + a^T x - x^T x - \alpha^T 1 - \beta^T 1 - \lambda D\right)$

NO constraint

⟸ 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

1. for minimizing wrt $\delta \rightarrow \min \|-\alpha+\beta-\lambda\cdot 1-\delta\cdot y\|^2$

$$= \min \|A-\delta y\|^2 = \min \left(\|A\|^2 - 2\delta\cdot y^T A + \delta^2 \|y\|^2\right)$$

use QUIN $\Rightarrow$ $\delta = \frac{y^T A}{\|y\|^2}$ (wh $A = -\alpha+\beta-\lambda\cdot 1$)

Show steps to find λ

$$\min\left(\frac{\|B-\lambda\cdot 1\|^2}{2} + \lambda D\right) = \min \frac{\|B\|^2 - 2\lambda B^T 1 + \lambda^2}{2} + \lambda D$$

$$= \min -\lambda(B^T 1 - D) + \lambda^2$$

$$\therefore \lambda = \frac{B^T 1 - D}{2} \quad \text{where } B = -\alpha+\beta-\delta\cdot y$$

Show steps to find β

$$\min \frac{\|C+\beta\|^2}{2} + \beta^T 1 \rightarrow \text{FOO} \quad (C+\beta)+\beta = 0$$

$$\Rightarrow \beta = -\frac{C}{2} \quad \text{where } C = -\alpha-\lambda\cdot 1-\delta\cdot y$$

Show steps to find α

$$\min \frac{\|D-\alpha\|^2}{2} + \alpha^T 1 \rightarrow \text{FOO.}$$

$$(\alpha - D) + \alpha = 0$$

$$\Rightarrow \alpha = \frac{D}{2} \quad \text{when } D = \beta-\lambda\cdot 1-\delta\cdot y$$