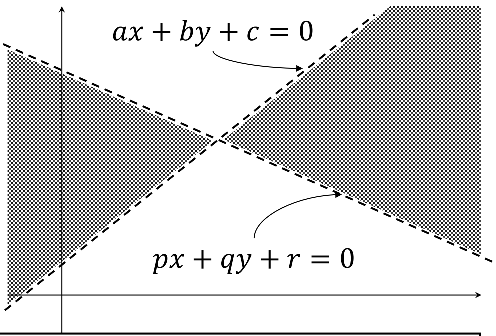
| CS 771A: Intro to Machine Learning, IIT Kanpur | | | | **Midsem Exam** (22 Feb 2024) |
|---|---|---|---|---|
| Name | MELBO | | | **40 marks** |
| Roll No | 20007 | Dept. | AWSM | Page **1** of **4** |

**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 $\phi$ 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)**




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

$\mathbf{W} = [ap, aq + bp, bq, ar + cp, br + cq, cr]$

**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

$a = 1 \quad b = 2 \quad c = 3 \quad p = 2 \quad q = 4 \quad r = 6$

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

$a = 1 \quad b = 2 \quad c = 3 \quad p = -1 \quad q = -2 \quad r = -3$

**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.
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.

Define $g(\mathbf{z}) \stackrel{\text{def}}{=} p(q(\mathbf{z}))$ with $q(\mathbf{z}) = x - 1$ and $p(t) = |t|$. Note that $\nabla q(\mathbf{z}) = (1,0)$ as $q$ is differentiable and $\frac{\partial q}{\partial y} = 0$. Note that We get $\partial p(t) = \begin{cases} \{-1\} & t < 0 \\ [-1,1] & t = 0 \\ \{+1\} & t > 0 \end{cases}$

This gives us $\partial g(\mathbf{z}) = \begin{cases} \{(-1,0)\} & x < 1 \\ \{(u,0): u \in [-1,1]\} & x = 1 \\ \{(1,0)\} & x > 1 \end{cases}$

Similarly, we get $\partial h(\mathbf{z}) = \begin{cases} \{(0,-1)\} & y < 2 \\ \{(0,v): v \in [-1,1]\} & y = 2 \\ \{(0,1)\} & y > 2 \end{cases}$

Answer to part b.

Using the sum rule gives us $\partial f(\mathbf{z}) = \begin{cases} \{(-1,-1)\} & x < 1 & y < 2 \\ \{(u,-1): u \in [-1,1]\} & x = 1 & y < 2 \\ \{(1,-1)\} & x > 1 & y < 2 \\ \{(-1,v): \in v[-1,1]\} & x < 1 & y = 2 \\ \{(u,v): u,v \in [-1,1]\} & x = 1 & y = 2 \\ \{(1,v): v \in [-1,1]\} & x > 1 & y = 2 \\ \{(-1,1)\} & x < 1 & y > 2 \\ \{(u,1): u \in [-1,1]\} & x = 1 & y > 2 \\ \{(1,1)\} & x > 1 & y > 2 \end{cases}$

We can get a more compact representation for $\partial f(\mathbf{z})$ by using the max rule. For this, we first rewrite the function as $f(\mathbf{z}) = \max_{\mathbf{s} \in \{-1,1\}^2} s_1(x-1) + s_2(y-2) = \max_{\mathbf{s} \in \{-1,1\}^2} \mathbf{s}^\top(\mathbf{z} - \mathbf{c})$ where $\mathbf{c} = (1,2) \in \mathbb{R}^2$. The max rule then tells us that the subdifferential can be found by taking convex combinations of the subgradient at all the winners i.e. all $\mathbf{s}$ such that $\mathbf{s}^\top(\mathbf{z} - \mathbf{c}) = f(\mathbf{z})$

$$\partial f(\mathbf{z}) = \{\mathbf{g} : \|\mathbf{g}\|_\infty \leq 1, \mathbf{g}^\top(\mathbf{z} - \mathbf{c}) = f(\mathbf{z})\}$$

where $\|\mathbf{g}\|_\infty = \max_i |g_i|$ is the so-called *sup* norm.

**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 😊)

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

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. $-1 \leq x_i \leq 1$ 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 $\alpha_j$ for the constraints $x_i \leq 1$, $\beta_i$ for $-1 \leq x_i$ (we collect these dual variables as vectors $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^n$), $\lambda$ for $\mathbf{x}^\top \mathbf{1} = D$ and $\delta$ 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)*

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \lambda, \delta) = \frac{1}{2}\|\mathbf{x}-\mathbf{a}\|_2^2 + \mathbf{a}^\top\mathbf{x} + \boldsymbol{\alpha}^\top(\mathbf{x}-\mathbf{1}) - \boldsymbol{\beta}^\top(\mathbf{x}+\mathbf{1}) + \lambda\cdot(\mathbf{x}^\top\mathbf{1} - D) + \delta\cdot\mathbf{x}^\top\mathbf{y}$$

The sign for the last two terms can be + or − as they correspond to equality constraints.

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)*

Setting $\frac{\partial \mathcal{L}}{\partial \mathbf{x}} = \mathbf{0}$ gives us $(\mathbf{x}-\mathbf{a}) + \mathbf{a} + \boldsymbol{\alpha} - \boldsymbol{\beta} + \lambda\cdot\mathbf{1} + \delta\cdot\mathbf{y} = 0$ which gives us

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

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)*

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

s.t.

| | |
|---|---|
| $\alpha_i \geq 0$ for all $i \in [n]$ | ⟸ Write constraint for $\boldsymbol{\alpha}$ here. |
| $\beta_i \geq 0$ for all $i \in [n]$ | ⟸ Write constraint for $\boldsymbol{\beta}$ here. |
| | ⟸ Write constraint for $\lambda$ here. |
| | ⟸ Write constraint for $\delta$ 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 $\delta$

The problem marginalized over $\delta$ becomes $\min_{\delta} \frac{1}{2}\|\mathbf{p} - \delta \cdot \mathbf{y}\|_2^2$ where $\mathbf{p} = (\boldsymbol{\beta} - \boldsymbol{\alpha}) - \lambda \cdot \mathbf{1}$

Using first order optimality gives us $(\mathbf{p} - \delta \cdot \mathbf{y})^\top(-\mathbf{y}) = 0$ i.e., $\delta = \mathbf{p}^\top \mathbf{y}/\|\mathbf{y}\|_2^2 = \frac{1}{n}\mathbf{p}^\top \mathbf{y}$ where the last step follows as $y_i \in \{-1,1\}$

Show steps to find $\lambda$

The problem over $\lambda$ becomes $\min_{\lambda} \frac{1}{2}\|\mathbf{q} - \lambda \cdot \mathbf{1}\|_2^2 + \lambda \cdot D$ where $\mathbf{q} = (\boldsymbol{\beta} - \boldsymbol{\alpha}) - \delta \cdot \mathbf{y}$

Using first order optimality gives us $(\mathbf{q} - \lambda \cdot \mathbf{1})^\top(-\mathbf{1}) + D = 0$ i.e., $\lambda = \frac{1}{n}(\mathbf{q}^\top \mathbf{1} - D)$

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

The problem over $\boldsymbol{\beta}$ becomes $\min_{\boldsymbol{\beta} \geq 0} \frac{1}{2}\|\boldsymbol{\beta} - \mathbf{r}\|_2^2 + \boldsymbol{\beta}^\top \mathbf{1}$ where $\mathbf{r} = \boldsymbol{\alpha} + \lambda \cdot \mathbf{1} + \delta \cdot \mathbf{y}$. Note that in general, the QUIN trick is guaranteed to work only on unidimensional problems. To be safe, we focus on a coordinate $\beta_i$ giving us $\min_{\beta_i \geq 0} \frac{1}{2}(\beta_i - r_i)^2 + \beta_i$. Applying the QUIN trick then gives us $\beta_i = \max\{r_i - 1, 0\}$. Thus, $\boldsymbol{\beta} = \max\{\mathbf{r} - \mathbf{1}, \mathbf{0}\}$ where $\max\{\cdot\}$ is applied coordinate-wise.

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

The problem over $\boldsymbol{\alpha}$ becomes $\min_{\boldsymbol{\alpha} \geq 0} \frac{1}{2}\|\mathbf{s} - \boldsymbol{\alpha}\|_2^2 + \boldsymbol{\alpha}^\top \mathbf{1}$ where $\mathbf{s} = \boldsymbol{\beta} - (\lambda \cdot \mathbf{1} + \delta \cdot \mathbf{y})$. As before, we focus on a coordinate $\alpha_i$ giving us $\min_{\alpha_i \geq 0} \frac{1}{2}(s_i - \alpha_i)^2 + \alpha_i$. Applying the QUIN trick now gives us $\alpha_i = \max\{s_i - 1, 0\}$ which means that $\boldsymbol{\alpha} = \max\{\mathbf{s} - \mathbf{1}, \mathbf{0}\}$.