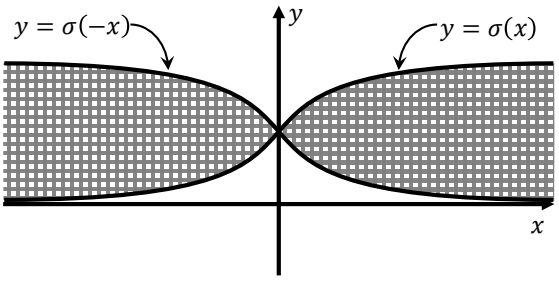
CS 771A: Intro to Machine Learning, IIT Kanpur				**Midsem Exam** (23 Feb 2025)
Name				**40 marks** Page **1** of **4**
Roll No		Dept.		

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. (Sigmoidal Cross) The figure depicts a binary classification task. The bold lines are the curves $y = \sigma(x)$ and $y = \sigma(-x)$ where $\sigma(x) = (1 + \exp(-x))^{-1}$ is the sigmoid and $x \in \mathbb{R}$. Create a feature map $\phi: \mathbb{R}^2 \to \mathbb{R}^D$ for some integer $D > 0$ so that for any 2D vector $\mathbf{z} = (x, y) \in \mathbb{R}^2$, the value of $\text{sign}\left(\mathbf{1}^\top \phi(\mathbf{z})\right)$ is $+1$ if $\mathbf{z}$ is in the cross-hatched region and -1



if $\mathbf{z}$ is in the white region. The D-dimensional all-ones vector is denoted as $\mathbf{1} = (1,1,\ldots,1) \in \mathbb{R}^D$. Write down your feature map in the space below. To create your feature map, you may use common functions such as polynomials, absolute value function, exponential function and even sigmoid function i.e., a feature map of the following kind would be valid (although it may not solve the problem) $\phi(\mathbf{z}) = \left(x, y, \exp(x-y), y^2 - x^2, \sigma(y)\right)$. **No derivation needed.** **(4 marks)**

$$\phi(x,y) = \begin{matrix} \left(-y^2, y\sigma(x), y\sigma(-x), -\sigma(x)\cdot\sigma(-x)\right) \\ \left(-y^2, y, -\sigma(x), (\sigma(x))^2\right) \\ \left(-y^2, y(\sigma(x)+\sigma(-x)), -\sigma(x)\cdot\sigma(-x)\right) \end{matrix}$$

Unshaded region points satisfy $y \geq \max\{\sigma(x), \sigma(-x)\}$ or $y \leq \min\{\sigma(x), \sigma(-x)\}$. In other words, $(y - \sigma(x))(y - \sigma(-x)) \geq 0$. Thus, the decision boundary is $y^2 - y(\sigma(x) + \sigma(-x)) + \sigma(x)\sigma(-x) = 0$
Simplifying this gives multiple solutions as shown on the left.

Q2. (Perpendicular inside a parabola) We say that a 2D vector of the form $\mathbf{v} = (v_1, v_2)$ lies on the parabola $y = x^2$ if $v_2 = v_1^2$. **No derivation needed in any of the three parts.** **(2+2+2 = 6 marks)**

1. Find a non-zero 2D vector $\mathbf{p}$ that lies on the parabola and is also perpendicular to the 2D vector $\left(\frac{1}{2}, \frac{1}{4}\right)$. Write down the value of $\mathbf{p}$.

$\mathbf{p} = (-2,4)$

Let $\mathbf{p} = (t, t^2)$ so that it lies on the parabola. For perpendicularity, we need $\frac{t}{2} + \frac{t^2}{4} = 0$. As we cannot have $t = 0$ as $\mathbf{p} \neq \mathbf{0}$, the other solution is $t = -2$ which gives us $\mathbf{p} = (-2,4)$

2. Given a non-zero 2D vector $\mathbf{q} = (s, t)$ that lies on the parabola (i.e. $t = s^2$ and $\mathbf{q} \neq \mathbf{0}$), find a non-zero 2D vector $\mathbf{p} = (m, n)$ so that $\mathbf{p} \neq \mathbf{0}$ lies on the parabola and is perpendicular to $\mathbf{q}$. Write down the value of m, n (The values of m, n must be in terms of s, t).

$\mathbf{p} = \left(-\frac{1}{s}, \frac{1}{s^2}\right)$ or $\left(-\frac{1}{s}, \frac{1}{t}\right)$

Let $\mathbf{p} = (m, m^2)$. For perpendicularity, we need $ms + m^2 s^2 = 0$ which gives us two solutions $ms = 0$ and $ms = -1$. Since we need non-zero vectors, we cannot have $ms = 0$ which means we must have $m = -\frac{1}{s}$ i.e., $\mathbf{p} = \left(-\frac{1}{s}, \frac{1}{s^2}\right)$ or $\mathbf{p} = \left(-\frac{1}{s}, \frac{1}{t}\right)$.

3. Find two 2D vectors $\mathbf{c}, \mathbf{d}$ that simultaneously satisfy the following properties: $\mathbf{c} \neq \mathbf{0}, \mathbf{d} \neq \mathbf{0}$, both $\mathbf{c}, \mathbf{d}$ lie on the parabola, $\mathbf{c}^\top \mathbf{d} = 0$ and both $\mathbf{c}, \mathbf{d}$ have the same Euclidean length.

$\mathbf{c} = (1,1) \quad \mathbf{d} = (-1,1)$

Part 2 tells us that the vectors must be of the form $\mathbf{c} = (u, u^2), \mathbf{d} = \left(-\frac{1}{u}, \frac{1}{u^2}\right)$. However, $\|\mathbf{c}\|_2^2 = u^2 + u^4$ while $\|\mathbf{d}\|_2^2 = \frac{1}{u^2} + \frac{1}{u^4} = \frac{u^2+u^4}{u^6}$. Thus, to have $\|\mathbf{c}\|_2 = \|\mathbf{d}\|_2$ we need $u^6 = 1$ i.e. $u = 1$ or else $u = -1$. Thus, $\mathbf{c} = (1,1), \mathbf{d} = (-1,1)$ or else $\mathbf{c} = (-1,1), \mathbf{d} = (1,1)$.

Q3 (Mahalanobis Margin) Consider a $D \times D$ square, symmetric matrix A that is invertible and positive definite i.e., $\mathbf{x}^\top A\mathbf{x} > 0$ and $\mathbf{x}^\top A^{-1}\mathbf{x} > 0$ for all non-zero vectors $\mathbf{x} \in \mathbb{R}^D$. Such a matrix gives us a new way of defining distances known as the *Mahalanobis distance* defined as $d_A(\mathbf{x}, \mathbf{y}) \stackrel{\text{def}}{=} \sqrt{(\mathbf{x}-\mathbf{y})^\top A(\mathbf{x}-\mathbf{y})}$ Melbo wishes to use this to define a new notion of margin. Please help Melbo. Hint: *You may solve the parts in any way, by deriving the dual, orthogonal decomposition, etc.*

a. For a hyperplane with non-zero normal vector $\mathbf{w} \in \mathbb{R}^D$, bias $b \in \mathbb{R}$, find the point $\mathbf{x}$ on the hyperplane closest to the origin in Mahalanobis distance by solving the following problem. Find the Mahalanobis distance of this point from the origin. Give brief derivation. (**7 marks**)

$$\min_{\mathbf{x}\in\mathbb{R}^D} \frac{1}{2}\mathbf{x}^\top A\mathbf{x}$$
$$\text{s.t. } \mathbf{w}^\top\mathbf{x} + b = 0$$

First give your final answers below

$$\mathbf{x} = -\left(\frac{b}{\mathbf{w}^\top A^{-1}\mathbf{w}}\right)\cdot A^{-1}\mathbf{w} \qquad d_A(\mathbf{x}, \mathbf{0}) = \frac{|b|}{\sqrt{\mathbf{w}^\top A^{-1}\mathbf{w}}}$$

Then give brief derivation below

The Lagrangian is $\mathcal{L}(\mathbf{x}, \lambda) = \frac{1}{2}\mathbf{x}^\top A\mathbf{x} + \lambda\cdot(\mathbf{w}^\top\mathbf{x} + b)$ and the dual is $\max_\lambda \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda)$. At the optimum, we must have $\frac{\partial\mathcal{L}}{\partial\mathbf{x}} = \mathbf{0}$ that gives $A\mathbf{x} + \lambda\cdot\mathbf{w} = \mathbf{0}$ i.e. $\mathbf{x} = -\lambda\cdot A^{-1}\mathbf{w}$. Hereon, we can use one of two ways to solve the problem – solve the dual or direct substitution.

Solution 1: The dual is $\max_\lambda \lambda b - \frac{\lambda^2}{2}\mathbf{w}^\top A^{-1}\mathbf{w}$ that we can solve to get $\lambda = \frac{b}{\mathbf{w}^\top A^{-1}\mathbf{w}}$

Solution 2: Since $\mathbf{x}$ lies on the hyperplane, we must have $\mathbf{w}^\top(-\lambda\cdot A^{-1}\mathbf{w}) + b = 0$.

Either way gives us the same answer – the closest point is $\mathbf{x} = -\left(\frac{b}{\mathbf{w}^\top A^{-1}\mathbf{w}}\right)\cdot A^{-1}\mathbf{w}$

The distance of the closest point to the origin is $d_A(\mathbf{x}, \mathbf{0}) = \sqrt{\mathbf{x}^\top A\mathbf{x}} = \frac{|b|}{\sqrt{\mathbf{w}^\top A^{-1}\mathbf{w}}}$

For part b below, setting $\frac{\partial\mathcal{L}}{\partial\mathbf{x}} = \mathbf{0}$ gives us $\mathbf{x} = \mathbf{q} - \lambda\cdot A^{-1}\mathbf{w}$ which upon substituting into the equation of the hyperplane (or solving the dual) gives us $\lambda = \frac{\mathbf{w}^\top\mathbf{q}+b}{\mathbf{w}^\top A^{-1}\mathbf{w}}$ i.e. the closest point is $\mathbf{x} = \mathbf{q} - \left(\frac{\mathbf{w}^\top\mathbf{q}+b}{\mathbf{w}^\top A^{-1}\mathbf{w}}\right)\cdot A^{-1}\mathbf{w}$ and $d_A(\mathbf{x}, \mathbf{q}) = \frac{|\mathbf{w}^\top\mathbf{q}+b|}{\sqrt{\mathbf{w}^\top A^{-1}\mathbf{w}}}$. Note: no derivation is needed for part b.

b. Repeat the above exercise for a general point $\mathbf{q} \in \mathbb{R}^D$ i.e., find the point $\mathbf{x}$ on the hyperplane that is closest to $\mathbf{q}$ in Mahalanobis distance by solving the following problem. Also find the Mahalanobis distance of this point from $\mathbf{q}$. Just give final answers – no derivation.(**4 marks**)

$$\min_{\mathbf{x}\in\mathbb{R}^D} \frac{1}{2}(\mathbf{x}-\mathbf{q})^\top A(\mathbf{x}-\mathbf{q})$$
$$\text{s.t. } \mathbf{w}^\top\mathbf{x} + b = 0$$

First give your final answers below

$\mathbf{x} = \mathbf{q} - \left(\frac{\mathbf{w}^\top \mathbf{q}+b}{\mathbf{w}^\top A^{-1}\mathbf{w}}\right) \cdot A^{-1}\mathbf{w}$ $\quad$ $d_A(\mathbf{x}, \mathbf{q}) = \frac{|\mathbf{w}^\top \mathbf{q}+b|}{\sqrt{\mathbf{w}^\top A^{-1}\mathbf{w}}}$

Q4 (Mahalanobis SVM) Melba wants to use the formula for the Mahalanobis distance between a point and a hyperplane to define a new kind of SVM that maximizes the Mahalanobis distance between the decision boundary and any training point. Help Melboo develop this new kind of SVM assuming the data is linearly separable. You only need to write down the primal for the hard SVM (i.e. no slack) and give brief justification. No need to derive the dual or solve the dual. **(4 marks)**

Given a set of data points $(\mathbf{x}^i, y^i)$, maximizing the closest Mahalanobis distance of any data point to the decision boundary while ensuring that classification is correct on every point gives

$$\max_{\mathbf{w},b} \left\{ \min_{i\in[n]} \frac{|\mathbf{w}^\top \mathbf{x}^i + b|}{\sqrt{\mathbf{w}^\top A^{-1}\mathbf{w}}} \right\}$$
$$\text{s.t.} \quad y^i \cdot (\mathbf{w}^\top \mathbf{x}^i + b) \geq 0 \quad \forall i \in [n]$$

Further simplification is not necessary for this question but can be done much the same way the regular SVM is derived. We notice that $|\mathbf{w}^\top \mathbf{x}^i + b| = y^i \cdot (\mathbf{w}^\top \mathbf{x}^i + b)$ due to perfect separation then let $i_0 = \arg\min_{i\in[n]} y^i \cdot (\mathbf{w}^\top \mathbf{x}^i + b)$, define $\epsilon \stackrel{\text{def}}{=} y^{i_0} \cdot (\mathbf{w}^\top \mathbf{x}^{i_0} + b)$ and let $\widetilde{\mathbf{w}} \stackrel{\text{def}}{=} \frac{\mathbf{w}}{\epsilon}, \tilde{b} \stackrel{\text{def}}{=} \frac{b}{\epsilon}$ to get $y^i \cdot (\widetilde{\mathbf{w}}^\top \mathbf{x}^i + \tilde{b}) \geq 1$ for all $i \in [n]$ as well as $\min_{i\in[n]} \frac{y^i \cdot (\widetilde{\mathbf{w}}^\top \mathbf{x}^i + \tilde{b})}{\sqrt{\mathbf{w}^\top A^{-1}\mathbf{w}}} = \frac{1}{\sqrt{\mathbf{w}^\top A^{-1}\mathbf{w}}}$ which gives us the following simplified optimization problem upon squaring and inverting the objective function.

$$\min_{\mathbf{w},b} \frac{1}{2}\mathbf{w}^\top A^{-1}\mathbf{w}$$
$$\text{s.t.} \quad y^i \cdot (\mathbf{w}^\top \mathbf{x}^i + b) \geq 1 \quad \forall i \in [n]$$

Q5. (Decomposable Classifiers) Melbo has learnt the following quadratic classifier to solve a difficult binary classification problem: $\text{sign}(12x^2 + x - 35)$. This classifier takes a real number $x \in \mathbb{R}$ and gives a binary verdict by looking at the sign of the value $12x^2 + x - 35$. Give your answer in the space provided – **no derivations needed in any part**. **(2 + 4 = 6 marks)**

1. What point(s) lie on the decision boundary? Your answer should be a list of real numbers.

$$x = \frac{5}{3}, x = -\frac{7}{4}$$

2. Melbo suspects that the verdict of this quadratic classifier can be expressed as the product of the verdicts of two linear classifiers i.e. there exist real numbers $a, b, p, q \in \mathbb{R}$ such that for every $x \in \mathbb{R}$, we have $\text{sign}(12x^2 + x - 35) = \text{sign}(ax + b) \cdot \text{sign}(px + q)$. Help confirm Melbo's suspicion by giving values of a, b, p, q below.

$a = \boxed{4}, b = \boxed{7}, p = \boxed{3}, q = \boxed{-5}$

We have $12x^2 + x - 35 = (4x + 7)(3x - 5)$ which gives us the solution by using the simple fact that $\text{sign}(ab) = \text{sign}(a) \cdot \text{sign}(b)$.

Q6. (The Melbo Equation) If Melbo has studied diligently for an exam, then with probability 75%, Melbo gets full marks in that exam i.e., with 25% probability, Melbo does not score full marks despite having studied diligently. If Melbo has not studied diligently for an exam, then Melbo can never get full marks in that exam. Now, Melbo is a very busy person with other things to do in life besides studying for exams (such as being the star of YouTube videos). Thus, for any exam, Melbo decides to study diligently for that exam with probability 50% (this choice is independent of what Melbo did for previous exams). It is known that Melbo scored full marks in the CS315 exam but did not score full marks in the CS771 exam. For each of the parts below, give the final answer as well as a brief derivation. **(3 x 3 = 9 marks)**

a. What is the probability that Melbo studied diligently for the CS315 exam?

S: The event that Melbo studied diligently
F: The event that Melbo gets full marks

Since Melbo cannot get full marks without studying diligently, we must have $\mathbb{P}[S \mid F] = 1$. Another way of obtaining the same answer is to use the Bayes rule to get $\mathbb{P}[S \mid F] = \frac{\mathbb{P}[F \mid S] \cdot \mathbb{P}[S]}{\mathbb{P}[F]} = \frac{0.75 \cdot 0.5}{0.375} = 1$. Note that $\mathbb{P}[F]$ is calculated in part c solution below.

b. What is the probability that Melbo studied diligently for the CS771 exam?

Since $\mathbb{P}[F \mid S] = 0.75$, we have $\mathbb{P}[\neg F \mid S] = 0.25$. Applying the Bayes rule gives

$$\mathbb{P}[S \mid \neg F] = \frac{\mathbb{P}[\neg F \mid S] \cdot \mathbb{P}[S]}{\mathbb{P}[\neg F]} = \frac{0.25 \cdot 0.5}{1 - 0.375} = 0.2 = \frac{1}{5}$$

Note that $\mathbb{P}[F]$ is calculated in part c solution below.

c. Melbo has a third exam for CS203 coming up next week. What is the probability that Melbo will score full marks in that exam?

We know that $\mathbb{P}[F \mid S] = 0.75$ and $\mathbb{P}[F \mid \neg S] = 0$ and $\mathbb{P}[S] = 0.5$. The law of total probability tells us that $\mathbb{P}[F] = \mathbb{P}[F \mid S] \cdot \mathbb{P}[S] + \mathbb{P}[F \mid \neg S] \cdot \mathbb{P}[\neg S] = 0.75 \cdot 0.5 = 0.375 = \frac{3}{8}$.