Nome ______________________________

Número ______________________ Curso ______________________

Seja uma base de dados $D = (x^n, y^n)_{n=1}^N$ onde $x^n = (x_1^n, x_2^n) \in \mathbb{R}^2$ e $y^n \in \{-1, 1\}$. Mais especificamente D^- são os eventos (x^n, y^n) tal como $y^n = -1$ enquanto D^+ agrupa os eventos (x^n, y^n) com $y^n = 1$.
Recordamos a notação $\widetilde{x} = (1, x_1, x_2)^T \in \mathbb{R}^3$, $\widetilde{w} = (w_0, w_1, w_2) \in \mathbb{R}^3$. Em particular, uma reta é dada pela sua equação

$$H(x; \widetilde{w}) = w_0 + w_1 x_1 + w_2 x_2 = \widetilde{w}^T \widetilde{x} = 0.$$

O conjunto D é (estritamente) linearmente separado pela reta $H(x; \widetilde{w}) = 0$ se

$$y^n H(x^n; \widetilde{w}) = y^n \widetilde{w}^T \widetilde{x}^n = y^n (w_0 + w_1 x_1^n + w_2 x_2^n) > 0. \qquad (*)$$

Notamos finalmente

$$\mathcal{E}(D) = \{\widetilde{w} \in \mathbb{R}^3, \text{ tal como } \widetilde{w} \text{ separa estritamente os dados de } D\},$$

ou seja, $\widetilde{w}$ satisfaz a relação $(*)$ para todos os eventos.

Parte A.

1) Recordamos a definição do perceptron $\widehat{y}(x; \widetilde{w}) = \text{sgn}(H(x; \widetilde{w}))$. Escrever o algoritmo estocástico do perceptron onde inicializamos o vetor $\widetilde{w}(0)$ com valores aleatórios no intervalo $[0, 1]$.

2) A função custo é dada por

$$E(\widetilde{w}, D) = \frac{1}{N} \sum_{n=1}^{N} \frac{|y^n + \widehat{y}^n|}{2}, \qquad \widehat{y}^n = \text{sgn}(H(x^n; \widetilde{w})).$$

Mostrar que

$$\widetilde{w} \in \mathcal{E}(D) \Leftrightarrow E(\widetilde{w}, D) = 0.$$

3) Mostrar que se $\widetilde{w}$ e $\widetilde{v}$ são elementos de $\mathcal{E}(D)$ então $\forall \theta \in [0, 1]$, temos

$$\widetilde{w}_\theta = \theta \widetilde{w} + (1 - \theta) \widetilde{v} \in \mathcal{E}(D).$$

4) Recordamos a definição do produto interno

$$\widetilde{w} \cdot \widetilde{v} = (w_0, w_1, w_2) \cdot (v_0, v_1, v_2) = w_0 v_0 + w_1 v_1 + w_2 v_2$$

e a norma $\|\widetilde{w}\|^2 = \widetilde{w} \cdot \widetilde{w}$.
Mostrar que

$$\|\theta \widetilde{w} + (1 - \theta) \widetilde{v}\|^2 = \theta^2 \|\widetilde{w}\|^2 + (1 - \theta)^2 \|\widetilde{v}\|^2 + 2\theta(1 - \theta) \widetilde{w} \cdot \widetilde{v}.$$

Deduzir que a norma $\|\widetilde{w}_\theta\|^2$ atinge o seu mínimo em $\overline{\theta} = \dfrac{(\widetilde{v} - \widetilde{w}) \cdot \widetilde{v}}{\|\widetilde{v} - \widetilde{w}\|^2}$.

5) Seja $\widetilde{w}, \widetilde{v} \in \mathcal{E}(D)$. Propor uma estratégia **ótimal** para construir um novo vetor $\widetilde{u} \in \mathcal{E}(D)$ tal que

$$\|\widetilde{u}\| \leq \min(\|\widetilde{w}\|^2, \|\widetilde{v}\|^2).$$

Parte B.

O objetivo desta parte é, dado um elemento $\widetilde{w} \in \mathcal{E}(D)$, normalizar a condição de margem que consiste em garantir a propriedade: $\forall n = 1, \cdots, N$

$$y^n H(x^n, \widetilde{v}) \geq 1 \text{ com } y^p H(x^p, \widetilde{v}) = 1, y^q H(x^q, \widetilde{v}) = 1, \qquad (\mathcal{P})$$

onde $x^p \in D^+$ e $x^q \in D^-$.

1) Seja $\widetilde{w} \in \mathcal{E}$. Mostrar que existe um evento $(x^p, y^p) \in D^+$ e um evento $(x^q, y^q) \in D^-$ tal que

$$\begin{aligned} 0 &< H(x^p; \widetilde{w}) \leq H(x^n; \widetilde{w}), \qquad \forall (x^n, y^n) \in D^+, \\ 0 &> H(x^q; \widetilde{w}) \geq H(x^n; \widetilde{w}), \qquad \forall (x^n, y^n) \in D^-. \end{aligned}$$

2) Mostrar que para qualquer $\mu \in \mathbb{R}$ e $\lambda > 0$ temos

$$\mu + \lambda(w_1 x_1^p + w_2 x_2^p) \leq \mu + \lambda(w_1 x_1^n + w_2 x_2^n), \quad \forall (x^n, y^n) \in D^+.$$

Do mesmo modo, mostrar que

$$\mu + \lambda(w_1 x_1^q + w_2 x_2^q) \geq \mu + \lambda(w_1 x_1^n + w_2 x_2^n), \quad \forall (x^n, y^n) \in D^-.$$

3) Deduzir que existem $\overline{\lambda}$ e $\overline{\mu}$ tal que

$$\mu + \lambda(w_1 x_1^p + w_2 x_2^p) = 1, \qquad \mu + \lambda(w_1 x_1^q + w_2 x_2^q) = -1$$

Provar que os valores dados para $\overline{\lambda}$ e $\overline{\mu}$ são

$$\overline{\lambda} = \frac{2}{w_1(x_1^p - x_1^q) + w_2(x_2^p - x_2^q)}, \qquad \overline{\mu} = -\frac{w_1(x_1^p + x_1^q) + w_2(x_2^p + x_2^q)}{w_1(x_1^p - x_1^q) + w_2(x_2^p - x_2^q)}.$$

4) Seja $\widetilde{v} = (\overline{\mu}, \overline{\lambda} w_1, \overline{\lambda} w_2)$. Justificar que temos a propriedade $(\mathcal{P})$. O vetor $\widetilde{v}$ chama-se o vetor normalizado de $\widetilde{w}$

Parte laboratorial.

O objetivo é implementar um algoritmo que determina uma aproximação $\widetilde{w} \in \mathcal{E}(D)$ de norma mínima que satisfaça a propriedade $(\mathcal{P})$. Para o efeito, as tarefas a realizar são as seguintes.

1. Implementar uma mudaça do script perceptron com uma inicialização aleatória do vetor $\widetilde{w}(0)$. O procedimento iterativo acaba quando a função erro $E(\widetilde{w}, D) = 0$.
2. Implementar a função de normalização que transforma o vetor $\widetilde{w} \in \mathcal{E}(D)$ num vetor normalizado $\widetilde{v}$ como indicado no ponto 4) da parte B.
3. Implementar uma função que contrói aleatoriamente um vetore $\widetilde{v} \in \mathcal{E}(D)$ normalizado (*i.e.* que satisfaz a propriedade $(\mathcal{P})$).
4. Implementar um script que determina um conjunto de 10 vetores normalizados e selecciona o de norma mínima.