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 $y^n \in \{-1, 1\}$. Recordamos a notação $\widetilde{x} = (1, x_1, x_2)^T \in \mathbb{R}^3$. O objetivo é realizar classificadores combinandos dois perceptrons

$$\widehat{y}_a(x; \widetilde{a}) = \text{sgn}(\widetilde{a}^T \widetilde{x}), \quad \widehat{y}_b(x; \widetilde{b}) = \text{sgn}(\widetilde{b}^T \widetilde{x}),$$

onde $\widetilde{a}, \widetilde{b} \in \mathbb{R}^3$. Notamos por d_a e d_b as retas $\widetilde{a}^T \widetilde{x} = 0$ e $\widetilde{b}^T \widetilde{x} = 0$ que nos supomos não serem colineares com o ponto de intersecção X.

Parte A.
O primeiro classificador combinado que consideramos nesta parte é apenas o produto

$$\widehat{y}(x; \widetilde{a}, \widetilde{b}) = \widehat{y}_a(x; \widetilde{a}) \widehat{y}_b(x; \widetilde{b}).$$

1) Seja a base de dados XOR caracterizada pelos pontos $((0,0), -1)$, $((1,0), 1)$, $((0,1), 1)$, $((1,1), -1)$. Determinar $\widetilde{a}$ e $\widetilde{b}$ de modo que o classificador $\widehat{y}(x; \widetilde{a}, \widetilde{b})$ separe perfeitamente os dados, seja $\widehat{y}(x^n; \widetilde{a}, \widetilde{b}) = y^n$, $n = 1, \cdots, 4$.
2) Mostar que para qualquer $x \neq X$ e $\lambda \neq 0$.

$$\widehat{y}_a(x; \widetilde{a}) = \text{sgn}(\lambda) \widehat{y}_a\Big(\lambda(x - X); \widetilde{a}\Big), \quad \widehat{y}_b(x; \widetilde{b}) = \text{sgn}(\lambda) \widehat{y}_b\Big(\lambda(x - X); \widetilde{b}\Big).$$

Deduzir que

$$\widehat{y}(x; \widetilde{a}, \widetilde{b}) = \widehat{y}\Big(\lambda(x - X); \widetilde{a}, \widetilde{b}\Big)$$

3) Concluir que o classificador $\widehat{y}(x; \widetilde{a}, \widetilde{b})$ separa linearmente duas zonas de $\mathbb{R}^2$

$$P^- = \{x \in \mathbb{R}^2;\ \widehat{y}(x; \widetilde{a}, \widetilde{b}) = -1\}, \quad P^+ = \{x \in \mathbb{R}^2;\ \widehat{y}(x; \widetilde{a}, \widetilde{b}) = +1\},$$

que são simétricas em ordem ao ponto X, limitadas pelas retas d_a e d_b. Exibir com um exemplo.
4) Propor dois vetores $\widetilde{a}$ e $\widetilde{b}$ para classificar perfeitamente a base de dados AND caracterizada pelos pontos $((0,0), -1)$, $((1,0), -1)$, $((0,1), -1)$, $((1,1); 1)$.
5) Propor um conjunto de 5 pontos com label, ou seja (x^n, y^n), $n = 1, \cdots, 4$, que não possamos classificar perfeitamente com este classificador.

Parte B.
Nesta parte, supomos que $y^n \in \{-1, 0, 1\}$ e consideramos um classificador com a expressão

$$\widetilde{z}(x; \widetilde{a}, \widetilde{b}) = \frac{1}{2}\Big(\widehat{y}_a(x; \widetilde{a}) + \widehat{y}_b(x; \widetilde{b})\Big).$$

1) Mostar que $\widetilde{z}(x; \widetilde{a}, \widetilde{b}) \in \{-1, 0, 1\}$. Propor dois vetores $\widetilde{a}, \widetilde{b}$ tais como o predictor $\widetilde{z}(x; \widetilde{a}, \widetilde{b})$ classifica perfeitamente o conjunto de pontos: $((0,0), 0)$, $((1,1), 0)$, $((0,1), -1)$, $((1,0), +1)$.
2) Caracterizar geometricamente os conjuntos

$$P^- = \{x \in \mathbb{R}^2; \widehat{z}(x; \widetilde{a}, \widetilde{b}) = -1\}, \quad P^0 = \{x \in \mathbb{R}^2; \widehat{z}(x; \widetilde{a}, \widetilde{b}) = 0\}, \quad P^+ = \{x \in \mathbb{R}^2; \widehat{z}(x; \widetilde{a}, \widetilde{b}) = +1\}$$

em relação às duas retas e ao ponto X. Visualizar P^-,P^0,P^+ quando $\widetilde{a} = (1, 0, 1)$ e $\widetilde{b} = (2, -1, 1)$.
3) Determinar se o conjunto $((-1,0), 0)$, $((-1/2,0), -1)$, $((1/2,0), 1)$, $((1,0), 0)$ pode ser perfeitamente separado com este predictor.

Parte C.

De novo supomos que $y^n \in \{-1, 1\}$ e definimos o classificador como

$$\widehat{p}(x; \widetilde{a}, \widetilde{b}, \widetilde{c}) = \operatorname{sgn}\Big(c_0 + c_1 \widehat{y}_a(x; \widetilde{a}) + c_2 \widehat{y}_b(x; \widetilde{b})\Big),$$

onde $\widetilde{c} = (c_1, c_2, c_3)^T \in \mathbb{R}^3$.

1) Mostrar que o classificador envolve apenas as 4 situações seguintes:

- caso (1) $p_1 = \operatorname{sgn}(c_0 - c_1 - c_2)$;
- caso (2) $p_2 = \operatorname{sgn}(c_0 + c_1 - c_2)$;
- caso (3) $p_3 = \operatorname{sgn}(c_0 - c_1 + c_2)$;
- caso (4) $p_4 = \operatorname{sgn}(c_0 + c_1 + c_2)$;

e notamos por P^1, P^2, P^3 e P^4 os quatro subconjuntos correspondentes aos quatro casos a caracterizar em função das retas d_a, d_b e X.

2) Determinar $\widetilde{c}$ tal como, $p_4 = 1$ e $p_1 = p_2 = p_3 = -1$.

3) Propor um classificador (ou seja determinar os vetores $\widetilde{a}$, $\widetilde{b}$ e $\widetilde{c}$) tal como o conjunto de dados seguinte seja perfeitamente classificado: $((-1, -1), -1)$, $((0, 0), -1)$, $((1, 0), 1)$, $((0, 1), 1)$, $((1, 1), 1)$.

4) Consideramos o caso particular de duas retas colineares $\widetilde{a} = (1, 1, 1)$ e $\widetilde{b} = (-1, 1, 1)$. Determinar as zonas onde $\widehat{y}_a$ e $\widehat{y}_b$ são positivas ou negativas. Notamos Δ a faixa situada entre as duas retas. Determinar $\widetilde{c}$ tal que $\widehat{p}(x; \widetilde{a}, \widetilde{b}, \widetilde{c})$ seja igual a 1 dentro de Δ e -1 fora de Δ.

Correção

Parte A.

1) Seja $\widetilde{a} = (-1, 2, 0)$ and $\widetilde{b} = (1, 0, -2)$. Temos $\widehat{y}_a(x) = 1$ se $x_1 > 1/2$ e $\widehat{y}_a(x) = -1$ se $x_1 < 1/2$ enquanto $\widehat{y}_b(x) = -1$ se $x_2 > 1/2$ e $\widehat{y}_b(x) = 1$ se $x_2 < 1/2$. Logo $\widehat{y}$ separa a base de dados XOR.

2) Temos

$$\widetilde{a}^T\widetilde{x} = \widetilde{a}^T(\widetilde{x} - \widetilde{X}) = \frac{1}{\lambda}\widetilde{a}^T\lambda(\widetilde{x} - \widetilde{X})$$

Logo

$$\widehat{y}_a(x;\widetilde{a}) = \mathrm{sgn}\left(\frac{1}{\lambda}\widetilde{a}^T\lambda(\widetilde{x} - \widetilde{X})\right) = \mathrm{sgn}(\lambda)\mathrm{sgn}\left(\widetilde{a}^T\lambda(\widetilde{x} - \widetilde{X})\right) = \mathrm{sgn}(\lambda)\widehat{y}_a\Big(\lambda(x - X);\widetilde{a}\Big).$$

Temos a mesma relação para $\widehat{y}_b(x;\widetilde{b})$ e deduzimos que

$$\widehat{y}(x;\widetilde{a},\widetilde{b}) = \mathrm{sgn}(\lambda)\widehat{y}_a\Big(\lambda(x - X);\widetilde{a}\Big)\mathrm{sgn}(\lambda)\widehat{y}_b\Big(\lambda(x - X);\widetilde{b}\Big) = \widehat{y}_a\Big(\lambda(x - X);\widetilde{a}\Big)\widehat{y}_b\Big(\lambda(x - X);\widetilde{b}\Big) = \widehat{y}\Big(\lambda(x - X)\widetilde{a},\widetilde{b}\Big)$$

3) Com $\lambda = 1$ e $\lambda = -1$ deduzimos

$$\widehat{y}\Big((x - X);\widetilde{a},\widetilde{b}\Big) = \widehat{y}\Big((X - x);\widetilde{a},\widetilde{b}\Big)$$

logo a função é simetrica em ordem ao ponto X e então a propriedade é verdadeira para os conjuntos associados P^+ e P^-. O ponto 1) representa um bom exemplo.

4) Separamos a base de dados AND com os vetores $\widetilde{a} = (-3/2, 1, 1)$ and $\widetilde{b} = (1, 0, 0)$.

5) A propriedade de simetria relativemente a um ponto implica que não podemos separa pelo menos 4 pontos alinhados que tem alternativemente -1 e $+1$. Por exemplo,$((0,0),+1)$ e $((-2,1),-1)$, $((-1,1),+1)$, $((+1,1),-1)$, $((2,1),+1)$.

Parte B.

1) Como os valores de $\widehat{y}_a$ e $\widehat{y}_b$ são $\{-1, 1\}$, a soma retorna $\{-2, 0, 2\}$ logo $\widetilde{z} \in \{-1, 0, 1\}$. Usando $\widetilde{a} = \{-1, 0, 2\}$ e $\widetilde{b} = \{+1, -2, 0\}$, temos um classificador que separa os quatro indicados.

2) Seja X o ponto de intersecção das duas retas. P^- e P^+ são dois cones simetricos de topo X situados entre d_a e d_b enquanto P^0 é a união dos dois outros cones.

3) Como $(2, -1, 1)^T(1, -1/2, 0) > 0$ deduzimos que $\widehat{y}_b((1, -1/2);\widetilde{b}) = 1$ logo $\widetilde{z}((-1/2, 0);\widetilde{a},\widetilde{b}) \neq -1$ e não podemos classificar este conjunto.

Parte C.

1) Como os valores de $\widehat{y}_a$ e $\widehat{y}_b$ são $\{-1, 1\}$, temos obviamente os quatro casos.

2) $p_4 > 0$ e $p_1, p_2, p_3 < 0$ implica que

$$c_0 - c_1 - c_2 < 0, \quad c_0 + c_1 - c_2 < 0, \quad c_0 - c_1 + c_2 < 0, \quad c_0 + c_1 + c_2 > 0.$$

De $c_0 + c_1 - c_2 < 0$ e $c_0 - c_1 + c_2 < 0$, deduzimos que $c_0 < 0$. Do outro lado, temos $c_0 - c_1 - c_2 < 0$ e $-c_0 - c_1 - c_2 < 0$ implica $0 < c_1 + c_2$. Aranjamos uma solução simetrica escolhando $c_0 = -1$ e $c_1 = c_2 = 1$.

3) $\widetilde{a} = (-1, 2, 0)$, $\widetilde{b} = (-1, 0, 2)$ e $\widetilde{c} = (1, 1, 1)$.

4) $\mathbb{R}^2$ é dividido em 3 partes. Notamos $P^- = \{1 + x_1 + x_2 < 0\}$, $P^+ = \{-1 + x_1 + x_2 > 0\}$, e $\Delta = \{1 + x_1 + x_2 > 0 \wedge -1 + x_1 + x_2 < 0\}$ a faixa situada entre os dois semi-planos. Temos

$$\begin{aligned}
\widetilde{y}_a(x;\widetilde{a}) = -1,\ \widetilde{y}_b(x;\widetilde{b}) = -1,\ \text{se } x \in P^- \\
\widetilde{y}_a(x;\widetilde{a}) = +1,\ \widetilde{y}_b(x;\widetilde{b}) = -1,\ \text{se } x \in \Delta \\
\widetilde{y}_a(x;\widetilde{a}) = +1,\ \widetilde{y}_b(x;\widetilde{b}) = +1,\ \text{se } x \in P^+
\end{aligned}$$

Logo $\widetilde{c} = (-1, 1, -1)$ é a resposta certa.