Task 1 - Example Prototype classifier

The goal of this task is to compute and visualize a prototype classifier for a very simple two-class classification problem. Consider the following data points:

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} 4 \\ 2 \end{bmatrix}, \mathbf{x}_4 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$\mathbf{x}_1$ and $\mathbf{x}_2$ belong to class -1, while $\mathbf{x}_3$ and $\mathbf{x}_4$ belong to class $+1$.

1. Compute the class means $\mathbf{w}_{-1}$ and $\mathbf{w}_{+1}$.
2. Compute the classification boundary $\mathbf{w}^\top \mathbf{x} - \beta = 0$ of the prototype classifier. Remember the following formulas:
$$\mathbf{w} = \mathbf{w}_{+1} - \mathbf{w}_{-1}$$
$$\beta = \frac{1}{2}(\mathbf{w}_{+1}^\top \mathbf{w}_{+1} - \mathbf{w}_{-1}^\top \mathbf{w}_{-1})$$
3. For each point, compute the assigned class label $\text{sign}(\mathbf{w}^\top \mathbf{x} - \beta)$. Are all points correctly classified?
4. Sketch the data points, their class means $\mathbf{w}_{-1}$ and $\mathbf{w}_{+1}$, the normal vector $\mathbf{w}$, and the classification boundary in the x_1-x_2 space.

Task 2 - The linear classification boundary

Consider a linear classification boundary $\mathbf{w}^\top \mathbf{x} - \beta = 0$. Draw a sketch in 2D to visualize the classification boundary and answer the following questions:

1. Suppose $\beta = 0$ and $\|\mathbf{w}\| = 1$. How large is the distance of a point $\mathbf{z}$ to the classification boundary?
2. How large is the distance of a point $\mathbf{z}$ to the classification boundary if $\|\mathbf{w}\| = 1$ but $\beta \neq 0$?
3. How large is the distance of a point $\mathbf{z}$ to the classification boundary for arbitrary β and $\mathbf{w}$?

Task 3 - Convergence of the perceptron

Suppose we have N points $\mathbf{x}_1, \ldots, \mathbf{x}_N \in \mathbb{R}^D$ with class labels $y_1, \ldots, y_N \in \{-1, +1\}$, and that the data set is linear separable. In this exercice we want to prove that the perceptron algorithm converges to a separating hyperplane in a finite number of steps.

As in the lecture, we denote a hyperplane by $\mathbf{w}^\top \mathbf{x} = 0$. Linear separability implies the existence of a $\mathbf{w}^{\text{sep}} \in \mathbb{R}^D$ such that for all $i \in \{1, \ldots, N\}$:

$$(\mathbf{w}^{\text{sep}})^\top \mathbf{x}_i y_i \geq \|x_i\|^2 \tag{1}$$

(You can see this as follows: Linear separabilty implies the existance of a $\tilde{\mathbf{w}}$ such that all data points are correctly classified, i.e. $\text{sign}(\tilde{\mathbf{w}}^\top \mathbf{x}_i) = y_i$. Hence $\forall i$ $(\tilde{\mathbf{w}}^\top \mathbf{x}_i) y_i \geq \epsilon$ for some $\epsilon > 0$. Rescaling of $\tilde{\mathbf{w}}$ yields $\mathbf{w}^{sep}$.)

Given a current $\mathbf{w}^{\text{old}} \in \mathbb{R}^D$, the perceptron algorithm identifies a point $\mathbf{x}_m$ that is misclassified, and produces the update rule $\mathbf{w}^{\text{new}} = \mathbf{w}^{\text{old}} + \mathbf{x}_m y_m$. Using Equation (1), show that

$$\|\mathbf{w}^{\text{new}} - \mathbf{w}^{\text{sep}}\|^2 \leq \|\mathbf{w}^{\text{old}} - \mathbf{w}^{\text{sep}}\|^2 - \|\mathbf{x}_m\|^2. \tag{2}$$

This implies that the perceptron algorithm converges to a separating hyperplane in a finite number of steps.