DSAI 512 Fall 2025 HW #1

Yiğit Ateş - 2025776009
Institute for Data Science and Artificial Intelligence, Boğaziçi University

I. Problem 1

A. 1.a

The initial decision boundary is defined where $\mathbf{w}^T\mathbf{x} = 0$. Expanding this equation:

$$w_0 + w_1x_1 + w_2x_2 = 0 \tag{1}$$

Solving for x_2:

$$x_2 = -\frac{w_1}{w_2}x_1 - \frac{w_0}{w_2} \tag{2}$$

Comparing with the line equation $x_2 = ax_1 + b$:

$$a = -\frac{w_1}{w_2}, \quad b = -\frac{w_0}{w_2} \tag{3}$$

B. 1.b

Given $\mathbf{w} = [2, 1, 4]^T$ and $-\mathbf{w} = [-2, -1, -4]^T$, the decision boundary for both $\mathbf{w}$ and $-\mathbf{w}$ is:

$$x_2 = -0.25x_1 - 0.5 \tag{4}$$

Both vectors yield the same decision boundary. Two verification points on this line are $(0, -0.5)$ and $(-2, 0)$. **The two decision boundaries coincide**; however, **the decision regions differ** because:

$$\text{For } \mathbf{w} : \text{sign}(\mathbf{w}^T\mathbf{x}) \tag{5}$$

$$\text{For } -\mathbf{w} : \text{sign}((-\mathbf{w})^T\mathbf{x}) = -\text{sign}(\mathbf{w}^T\mathbf{x}) \tag{6}$$

The two weight vectors produce opposite classifications on each side of the boundary as it is illustrated in Figure 1.

II. Problem 2

A. 2.a

$$w^* = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix}, \quad X = \begin{bmatrix} 1 & -1 & 3 \\ 1 & -2 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 1 \\ 1 & -1 & -1 \end{bmatrix}, \quad y = \begin{bmatrix} -1 \\ -1 \\ -1 \\ 1 \\ 1 \end{bmatrix}$$

Computing functional margins $y_n(w^{*T}x_n)$:

$$Xw^* = \begin{bmatrix} 1 & -1 & 3 \\ 1 & -2 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 1 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} -5 \\ -2 \\ -1 \\ 2 \\ 3 \end{bmatrix}$$

$$y \odot (Xw^*) = \begin{bmatrix} 5 \\ 2 \\ 1 \\ 2 \\ 3 \end{bmatrix}$$

Therefore: $\rho = \min_n y_n(w^{*T}x_n) = 1$

Since w^* perfectly separates all data points, we have $y_n(\mathbf{w}^{*T}\mathbf{x}_n) > 0$ for all n, which guarantees $\rho > 0$.

From the calculations, ρ measures how confidently the algorithm separates the data. Each value $y_n(\mathbf{w}^{*T}\mathbf{x}_n)$ tells us how far point x_n is from the decision boundary. The smaller this value, the closer the point is to the decision boundary.

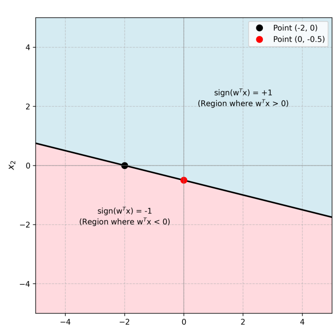

Fig. 1. The decision boundary $x_2 = -0.25x_1 - 0.5$ with its corresponding decision regions for $\mathbf{w} = [2, 1, 4]^T$. For the vector $-\mathbf{w}$, the decision regions are swapped.

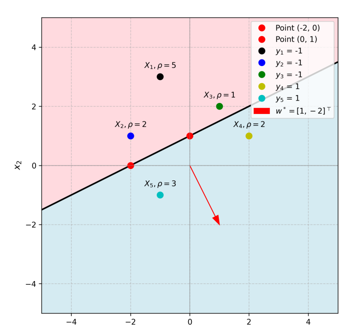

Fig. 2. Decision boundary for $w^* = [2, 1, -2]^\top$ with the 5 data points. The normal vector w^* is shown in red, and the point x_3 achieving the minimum margin $\rho = 1$ is marked in dark green.

B. 2.b

We state the PLA update rule, which is applied when the point $x(t-1)$ (with label $y(t-1)$) is misclassified by $w(t-1)$:

$$w(t) = w(t-1) + y(t-1)x(t-1) \tag{7}$$

From the definition of ρ in part 2.a, we are given

$$\rho = \min_{1 \leq n \leq N} y_n(w^{*T}x_n) \tag{8}$$

Since w^* perfectly separates the data, $\rho > 0$. This definition means that for any data point n, including the misclassified point $x(t-1)$, the following must be true:

$$y(t-1)(x(t-1)w^*) \geq \rho \tag{9}$$

Substituting the update rule from equation (7) for $w(t)$.

$$\begin{align} w(t)^T w^* &= (w(t-1) + y(t-1)x(t-1))^T w^* \tag{10} \\ &= w(t-1)^T w^* + (y(t-1)x(t-1))^T w^* \tag{11} \\ &= w(t-1)^T w^* + y(t-1)(x(t-1)^T w^*) \tag{12} \end{align}$$

We can now form an inequality using our fact from equation (9). We replace the term $y(t-1)(x(t-1)^T w^*)$ with ρ. Since we know from equation (9) that $y(t-1)(x(t-1)^T w^*) \geq \rho$, this substitution changes the $=$ to $\geq$:

$$w(t)^T w^* \geq w(t-1)^T w^* + \rho \tag{13}$$

This completes the proof. We will now use the result from equation (13) to prove $w(t)^T w^* \geq t\rho$ by induction. The problem states $w(0) = 0$.

$$w(0)^T w^* = 0^T w^* = 0 \tag{14}$$

We check this against our formula: $0 \geq (0) \cdot \rho$, which simplifies to $0 \geq 0$. Assume the statement is true for step $t-1$:

$$w(t-1)^T w^* \geq (t-1)\rho \tag{15}$$

We must now show the statement holds for step t. We start with the inequality we proved in equation (13). We can substitute $(t-1)\rho$ for $w(t-1)^T w^*$:

$$\begin{aligned} w(t)^T w^* &\geq w(t-1)^T w^* + \rho \\ &\geq ((t-1)\rho) + \rho \\ &\geq t\rho - \rho + \rho \\ &\geq t\rho \end{aligned} \tag{16}$$

Proven.

C. 2.c

We begin with the PLA update rule from equation (7):

$$w(t) = w(t-1) + y(t-1)x(t-1) \tag{17}$$

We take the squared norm of both sides:

$$||w(t)||^2 = ||w(t-1) + y(t-1)x(t-1)||^2 \tag{18}$$

We expand the right hand side:

$$\begin{aligned} ||w(t)||^2 = &||w(t-1)||^2 + \dots \\ &2(w(t-1))^T (y(t-1)x(t-1)) + \dots \\ &||y(t-1)x(t-1)||^2 \end{aligned} \tag{19}$$

We simplify the middle and last terms. Since $y(t-1)$ is a scalar, we can pull it out.

$$\begin{align} \text{Middle term} &= 2y(t-1)(w(t-1)^T x(t-1)) \tag{20} \\ \text{Last term} &= (y(t-1))^2 ||x(t-1)||^2 \tag{21} \end{align}$$

Since $y(t-1)$ is either $+1$ or -1, $(y(t-1))^2 = 1$. The last term simplifies to $||x(t-1)||^2$. Substituting these back into equation (19) yields:

$$\begin{aligned} ||w(t)||^2 = &||w(t-1)||^2 + \dots \\ &2y(t-1)(w(t-1)^T x(t-1)) + ||x(t-1)||^2 \end{aligned} \tag{22}$$

The problem states $x(t-1)$ was misclassified by $w(t-1)$. By definition, so their product is non-positive:

$$y(t-1)(w(t-1)^T x(t-1)) \leq 0 \tag{23}$$

Therefore, the middle term $2y(t-1)(w(t-1)^T x(t-1))$ is also non-positive. If we remove this term from the right hand side, the expression can only become larger or stay the same. This allows us to form the inequality:

$$||w(t)||^2 \leq ||w(t-1)||^2 + ||x(t-1)||^2 \tag{24}$$

Proven.

D. 2.d

From part 2.c, we have the inequality for each update t:

$$||w(t)||^2 \leq ||w(t-1)||^2 + ||x(t-1)||^2 \tag{25}$$

From the problem definition, we let $R = \max_n ||x_n||$. This implies $||x(t-1)||^2 \leq R^2$ for any misclassified point $x(t-1)$. We can substitute this into the inequality:

$$||w(t)||^2 \leq ||w(t-1)||^2 + R^2 \tag{26}$$

We apply this inequality:

$$\begin{aligned} ||w(t)||^2 &\leq ||w(t-1)||^2 + R^2 \\ &\leq (||w(t-2)||^2 + R^2) + R^2 = ||w(t-2)||^2 + 2R^2 \\ &\leq (||w(t-3)||^2 + R^2) + 2R^2 = ||w(t-3)||^2 + 3R^2 \\ &\dots \\ &\leq ||w(0)||^2 + tR^2 \end{aligned} \tag{27}$$

Since $w(0) = 0$, $||w(0)||^2 = 0$. The expression simplifies to:

$$||w(t)||^2 \leq tR^2 \tag{28}$$

Proven.

E. 2.e

We combine the results from parts 2.b and 2.d.

1) From 2.b: $w(t)^T w^* \geq t\rho$
2) From 2.d: $||w(t)||^2 \leq tR^2$, which implies $||w(t)|| \leq \sqrt{t}R$

We start with the expression for the cosine of the angle between $w(t)$ and w^*:

$$\frac{w(t)^T w^*}{||w(t)|| ||w^*||} \tag{29}$$

We substitute the result from 2.b into the numerator. Since $w(t)^T w^* \geq t\rho$, we establish a lower bound:

$$\frac{w(t)^T w^*}{||w(t)||||w^*||} \geq \frac{t\rho}{||w(t)||||w^*||} \tag{30}$$

Next, we substitute the result from 2.d into the denominator. Since $||w(t)|| \leq \sqrt{t}R$, it follows that $\frac{1}{||w(t)||} \geq \frac{1}{\sqrt{t}R}$. Applying this:

$$\begin{aligned}\frac{t\rho}{||w(t)||||w^*||} &\geq \frac{t\rho}{(\sqrt{t}R)||w^*||} \\ &= \frac{\sqrt{t}\sqrt{t}\rho}{\sqrt{t}R||w^*||} \\ &= \sqrt{t}\frac{\rho}{R||w^*||} \end{aligned} \tag{31}$$

Combining these steps, we arrive at the final inequality:

$$\frac{w(t)^T w^*}{||w(t)||||w^*||} \geq \sqrt{t}\frac{\rho}{R||w^*||} \tag{32}$$

Proven.

F. 2.f

From part 2.e, we derived the inequality:

$$\frac{w(t)^T w^*}{||w(t)||||w^*||} \geq \sqrt{t}\frac{\rho}{R||w^*||} \tag{33}$$

The left hand side is the cosine of the angle between $w(t)$ and w^* which can not be greater than 1.

$$1 \geq \frac{w(t)^T w^*}{||w(t)||||w^*||} \tag{34}$$

We combine these two inequalities:

$$1 \geq \sqrt{t}\frac{\rho}{R||w^*||} \tag{35}$$

We now solve for t to find the upper bound on the number of updates.

$$\begin{aligned}\frac{R||w^*||}{\rho} &\geq \sqrt{t} \\ \left(\frac{R||w^*||}{\rho}\right)^2 &\geq t\end{aligned} \tag{36}$$

This concludes the proof that the PLA converges in a finite number of steps, t, bounded by:

$$t \leq \frac{R^2||w^*||^2}{\rho^2} \tag{37}$$

Proven.

III. Code Repository

Codes, data visualizations and handwritten proofs will be available at *this GitHub repository* after the deadline.

IV. Acknowledgements

Gemini 2.5 Pro was used to assist in converting handwritten mathematical notation to LaTeX format, correcting mathematical wording, and optimizing spacing in Problem 2.