(a) NP-hardness of $\mathrm{ERM}_{\mathcal{HS}_n}$

Goal

We show that if there exists a polynomial-time algorithm for ERM over the halfspace class $\mathcal{HS}_n$, then one can solve the *Maximum Feasible Subsystem (Max FS)* problem in polynomial time. Since Max FS is NP-hard, this implies that $\mathrm{ERM}_{\mathcal{HS}_n}$ is NP-hard.

1. Definition of the Max FS Problem

Input:

- $A \in \mathbb{R}^{m\times n}$,
- $b \in \mathbb{R}^m$.

  System of inequalities:

$$\langle A_i, w\rangle > b_i \qquad (i = 1, \dots, m).$$

Since the system may be infeasible, the goal is to find a vector $w \in \mathbb{R}^n$ that satisfies the maximum number of inequalities.

2. Preprocessing Assumptions

- Since all inequalities are strict, we may assume $b_i \neq 0$ for all i. (If $b_i = 0$, it can be replaced by an arbitrarily small $\varepsilon > 0$.)
- The system $Aw > b$ is invariant under positive scaling. Hence, we may assume

$$|b_i| = |b_j| \quad \forall i, j,$$

  and denote this common value by $b > 0$.

3. Construction of the Learning Sample

For each inequality, construct one labeled example:

$$x_i = -\operatorname{sign}(b_i)\, A_i, \qquad y_i = -\operatorname{sign}(b_i), \qquad i = 1, \dots, m.$$

This yields the training sample

$$S = \big((x_1, y_1), \dots, (x_m, y_m)\big).$$

4. Correct Classification by a Halfspace

Define a halfspace classifier by

$$h_{w,b}(x) = \operatorname{sign}(\langle w, x\rangle + b).$$

A data point (x_i, y_i) is correctly classified if and only if

$$h_{w,b}(x_i) = y_i \quad \Longleftrightarrow \quad y_i(\langle w, x_i\rangle + b) > 0.$$

5. Key Equivalence

Substituting the definitions of x_i and y_i, we obtain

$$\begin{aligned} y_i(\langle w, x_i \rangle + b) &= (-\operatorname{sign}(b_i))\big(\langle w, -\operatorname{sign}(b_i)A_i \rangle + b\big) \\ &= \langle w, A_i \rangle - b_i. \end{aligned}$$

Therefore,

$$h_{w,b}(x_i) = y_i \quad \iff \quad \langle w, A_i \rangle - b_i > 0 \quad \iff \quad \langle A_i, w \rangle > b_i.$$

Hence,

$$\boxed{\text{The data point } (x_i, y_i) \text{ is correctly classified} \quad \iff \quad \text{the inequality } \langle A_i, w \rangle > b_i \text{ is satisfied.}}$$

6. Correspondence Between Max FS and ERM

- Each inequality corresponds to one data point.
- Satisfying an inequality corresponds to correctly classifying the data point.

Thus, maximizing the number of satisfied inequalities in Max FS is exactly equivalent to maximizing the number of correctly classified examples in $\mathrm{ERM}_{\mathcal{HS}_n}$.

7. NP-hardness Conclusion

- Max FS is known to be NP-hard.
- Max FS can be reduced to $\mathrm{ERM}_{\mathcal{HS}_n}$ in polynomial time.

Therefore,

$$\boxed{\mathrm{ERM}_{\mathcal{HS}_n} \text{ is NP-hard.}}$$

In particular, if $\mathrm{P} \neq \mathrm{NP}$, there is no polynomial-time algorithm for $\mathrm{ERM}_{\mathcal{HS}_n}$.