Instructions: All your solutions should be prepared in LaTeX and the PDF and .tex should be submitted to canvas. For each question, the best and correct answers will be selected as sample solutions for the entire class to enjoy. If you prefer that we do not use your solutions, please indicate this clearly on the first page of your assignment.

The programming parts can be written in the programming language of your choice and the code should be submitted alongside your solutions.

1. Convex Sets.   Prove or give a counterexample:

   a. The intersection of convex sets is a convex set.

   b. A half-space is a convex set.

   c. Every polyhedron is a convex set (remember that a polyhedron is the feasible set of a linear program, *i.e.* it is defined by a finite set of linear inequalities)

2. Convex Hulls.   Let us define the *convex hull* of a set X as the smallest (for the partial order defined by inclusion) convex set containing X and denote it by $C(X)$. In other words, there is no other convex set C' such that $X \subseteq C' \subset C(X)$.

   a. Show that $C(X)$ is the intersection of all convex sets containing X that is:

$$C(X) = \bigcap \{C \mid X \subseteq C \text{ and } C \text{ is convex}\}$$

   b. What is the convex hull of a convex set?

   c. Show that when $X = \{\mathbf{x}_1, \dots, \dots \mathbf{x}_n\}$ is finite, then $C(X)$ can also be written as:

$$C(X) = \left\{ \sum_{i=1}^{n} \lambda_i \mathbf{x}_i \mid \lambda_i \geq 0, \ 1 \leq i \leq n \text{ and } \sum_{i=1}^{n} \lambda_i = 1 \right\}$$

   d. What is the convex hull of two points? three points?

3. Strict Convexity of the ℓ_2 norm. Let y be an arbitrary point in $\mathbb{R}^n$ and let us define the function "distance to y" by:
$$d_{\mathbf{y}}(\mathbf{x}) = \|\mathbf{x} - \mathbf{y}\|_2^2, \quad \mathbf{x} \in \mathbb{R}^n$$

Show that the function $d_{\mathbf{y}}$ is strictly convex. Remember that a function f defined over $\mathbb{R}^n$ is strictly convex if and only if for any pair of points $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ with $\mathbf{x} \neq \mathbf{y}$ and any $\lambda \in (0, 1)$:

$$f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{y}) < \lambda f(\mathbf{x}) + (1 - \lambda)f(\mathbf{y})$$

4. Infeasibility and Unboundedness. Discuss the feasibility and boundedness of the following linear programs:

$$
\begin{aligned}
\text{maximize} \quad & 2x_2 + x_3 \\
\text{subject to} \quad & x_1 - x_2 \leq 5 \\
& -2x_1 + x_2 \leq 3 \\
& x_1 - 2x_3 \leq 5 \\
& x_1, x_2, x_3 \geq 0
\end{aligned}
\qquad
\begin{aligned}
\text{minimize} \quad & x + y + z + w \\
\text{subject to} \quad & x + 3y + 2z + 4w \leq 5 \\
& 3x + y + 2z + w \leq 4 \\
& 5x + 3y + 3z + 3w = 9 \\
& x, y, z, w \geq 0
\end{aligned}
$$

5. Linear Classifiers and the Perceptron Algorithm In this problem, we will work on the Iris dataset available at `https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data`. The dataset is a single comma-separated value (CSV) file. The first 4 fields of each line contain the measurements of a sample of Iris flower, the last field is the name of this sample's species of Iris. More information about the dataset can be found at `https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.names`.

The goal of this problem is to construct a classifier to distinguish the *Iris setosa* from the other species of Iris. That is, we want to construct a function f which takes as input the vector $\mathbf{x} \in \mathbb{R}^4$ of measurements of a sample and return $f(\mathbf{x}) \in \{0, 1\}$ such that:

$$
f(\mathbf{x}) = \begin{cases} 1 & \text{if the sample is an Iris setosa} \\ 0 & \text{otherwise} \end{cases}
$$

a. Download the dataset, choose a pair of coordinates (that is, any two of the first four fields) and draw a scatter plot of the dataset for this pair of coordinates. The color of the points should be determined by the Iris species.

b. Are the samples of Iris setosa linearly separable (that is, does there exist a separating hyperplane) from the samples from the other species? Is a linearly separable dataset always separable after having been projected on an arbitrary pair of coordinates? (prove of describe a counter-example)?

c. Explain why finding a separating hyperplane as described in the previous part is sufficient to construct a classifier f as described in the introduction of this problem. A classifier constructed in such a way is called a *linear classifier.*

d. Implement the perceptron algorithm in the programming language of your choice. Run your algorithm on the Iris dataset to find a hyperplane separating the Iris setosa samples from the other samples. Report the weights defining the hyperplane as well as the code you wrote.