Note: You are **not** expected to know how to implement these exercises, but to try to implement those yourself following the material. Doing so will set you to success for the upcoming assignments. The first homework assignment will be published later this week.

Problem 3.1: Convexity of Functions

(a) A function $f : \mathbb{R}^n \to \mathbb{R}$, denoted by $f(x) = ||x||$, is called a *norm* if it satisfies the following four properties:

1. $f(x) \geq 0$ for all $x \in \mathbb{R}^n$
2. $f(x) = 0$ only if $x = 0$
3. $f(tx) = |t|f(x)$, for all $x \in \mathbb{R}^n$ and $t \in \mathbb{R}$ (f is *homogeneous* of degree 1)
4. $f(x + y) \leq f(x) + f(y)$, for all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$ (f satisfies triangle inequality)

Show that the norm $f(x) = ||x||$ is a convex function.

(b) Let $f_i : \mathbb{R}^n \to \mathbb{R}$ be convex functions for $i = 1, \ldots, n$, and let $\alpha_i > 0$ be positive scalars for $i = 1, \ldots, n$. Show that the function $g : \mathbb{R}^n \to \mathbb{R}$, defined as

$$g(x) = \sum_{i=1}^{n} \alpha_i f_i(x)$$

is convex.

(c) Let $I = \{1, \ldots, n\}$ be an index set, and let $f_i : \mathbb{R}^n \to \mathbb{R}$ be convex functions for all $i \in I$. Show that the function $g : \mathbb{R}^n \to \mathbb{R}$, defined as

$$g(x) = \max_{i \in I}\{f_i(x)\}$$

is convex.

Problem 3.2: Convexity under Composition

Let $S \subseteq \mathbb{R}^n$ be a nonempty convex set. Let $h : S \to \mathbb{R}$ be a convex function, and let $g : \mathbb{R} \to \mathbb{R}$ be a monotonically non-decreasing convex function over the set $\{h(x) : x \in S\}$. Show that the composition function

$$f(x) = g(h(x))$$

is convex.

Problem 3.3: Convexity of Optimization Problems

(a) Suppose we are given some data that can be separated into two sets in $\mathbb{R}^n$:

$$X = \{x_1, \ldots, x_N\} \quad \text{and} \quad Y = \{y_1, \ldots, y_M\}.$$

We would like to construct a classifier that separates the points in X and Y into two distinct sets based on some features. Ideally, the classifier could then be used to classify future data points to the correct sets.

For example, X could represent spam email, Y regular email, and we would like to train a classifier based on some features, such as word stems appearing in the email. If we can train an accurate enough classifier based on some training data X and Y, we could then use the classifier as an email spam filter to direct incoming emails to either inbox or trash.

In linear classification, we seek an affine function $f(x) = a^\top x - b$ that correctly classifies the points in X and Y, i.e.,

$$a^\top x_i - b > 0, \; i = 1, \ldots, N \qquad a^\top y_i - b < 0, \; i = 1, \ldots, M \tag{1}$$

Geometrically, we seek a hyperplane that separates the points in X and Y. The unknown variables are $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$, and we would like to find the best values for them.

Since there is always a possibility for misclassification, we can introduce a *buffer zone* to trade some of the *robustness* of the classifier to outliers. We can do this by first rewriting the strict inequalities in (1) as

$$a^\top x_i - b \geq 1, \; i = 1, \dots, N \qquad a^\top y_i - b \leq -1, \; i = 1, \dots, M \tag{2}$$

and then relax these constraints by introducing nonnegative variables $u_1, \dots, u_N$ and $v_1, \dots, v_M$, and rewriting (2) as

$$a^\top x_i - b \geq 1 - u_i, \; i = 1, \dots, N \qquad a^\top y_i - b \leq -1 + v_i, \; i = 1, \dots, M \tag{3}$$

We can think of u_i and v_i as measures which compute how much the corresponding points x_i and y_i, respectively, are violated if they are misclassified.

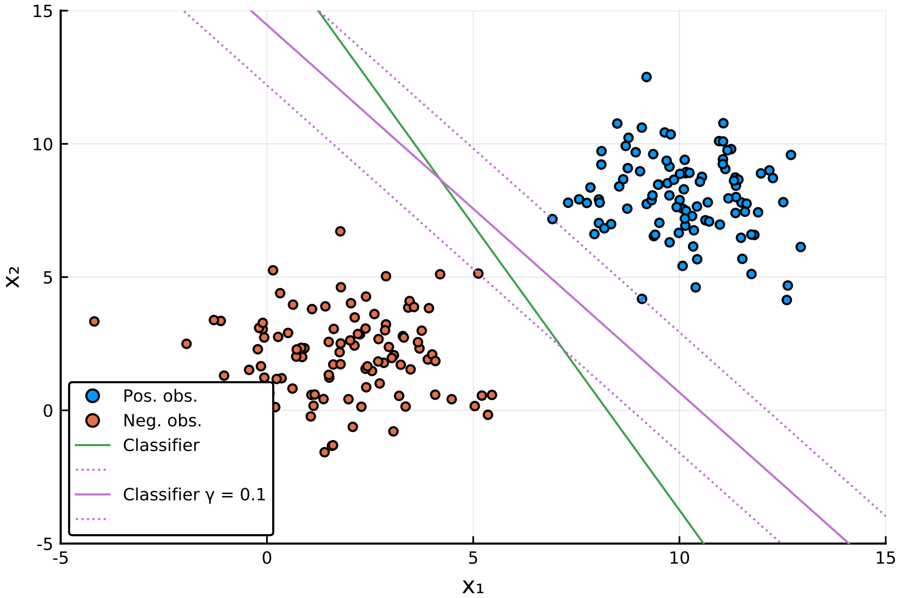

Figure 1: Classification results. Buffer zone corresponds to the area between the dotted pink lines.

Using this information, we can formulate the following *robust* classification problem presented in Lecture 1:

$$\text{min.} \; \sum_{i=1}^{N} u_i + \sum_{i=1}^{M} v_i + \gamma ||a||_2^2 \tag{4}$$

$$\text{subject to: } a^\top x_i - b + u_i \geq \; 1, \qquad i = 1, \dots, N \tag{5}$$

$$a^\top y_i - b - v_i \leq -1, \qquad i = 1, \dots, M \tag{6}$$

$$u_i \geq 0, \qquad i = 1, \dots, N \tag{7}$$

$$v_i \geq 0, \qquad i = 1, \dots, M. \tag{8}$$

The first term in the objective (4) corresponds to the total classification error

$$\sum_{i=1}^{N} u_i + \sum_{i=1}^{M} v_i$$

and second term $||a||_2^2$ is inversely proportional to the width of the *buffer zone* which is equal to $2/||a||_2$ if $u_i = 0$ for all $i = 1, \dots, N$ and $v_i = 0$ for all $i = 1, \dots, m$ (proof). Since we want to minimize the total classification error and to maximize the width of the buffer zone, the objective function (4) achieves both of these goals. The parameter $\gamma > 0$ controls the trade-off between these two goals.

Justify why the problem (4) – (8), also called a *support vector machine*, is a convex optimization problem.

(b) Consider the following portfolio optimization problem with some scalar $\lambda \in [0, 1]$:

$$\begin{aligned} \max_{x}.\ & \lambda(\mu^\top x) - (1-\lambda)x^\top \Sigma x \\ \text{subject to: } & \sum_{i \in N} x_i = 1 \\ & x \geq 0 \end{aligned}$$

which can be equivalently written as

$$\min_{x}.\ -\lambda(\mu^\top x) + (1-\lambda)x^\top \Sigma x \tag{9}$$

$$\text{subject to: } \sum_{i \in N} x_i = 1 \tag{10}$$

$$x \geq 0 \tag{11}$$

The objective (9) is to minimize a weighted sum of negative expected profit $-\mu^\top x$ (which is equal to maximizing the actual expected profit $\mu^\top x$) and the risk $x^\top \Sigma x$ (i.e., portfolio variance). Notice that Σ is a positive semidefinite (covariance) matrix. Justify why the problem (9) – (11) is a convex optimization problem.

(c) Consider the following linear optimization problem with $x \in \mathbb{R}^n$:

$$\min.\ c^\top x \tag{12}$$

$$\text{subject to: } a_i^\top x \leq b_i, \qquad i = 1, \dots, m \tag{13}$$

$$x_i \geq 0, \qquad i = 1, \dots, n \tag{14}$$

in which $c \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$ are fixed, and $a_i \in \mathbb{R}^n$, for all $i = 1, \dots, m$, are known to lie in given ellipsoids:

$$a_i \in \mathcal{E}_i = \{\overline{a}_i + P_i u : ||u||_2 \leq \Gamma_i\}, \tag{15}$$

where

- $\overline{a}_i$ is the nominal (average) value.
- P_i is the characteristic matrix of the ellipsoid $\mathcal{E}_i$.
- Γ_i is the risk-aversion control parameter, or *budget of uncertainty*.

Suppose we want the constraints (13) to be satisfied for all possible values of the parameter vectors $a_i \in \mathcal{E}_i$. This leads to the following *robust* linear optimization problem:

$$\min.\ c^\top x \tag{16}$$

$$\text{subject to: } \max_{a_i \in \mathcal{E}_i} \left\{a_i^\top x\right\} \leq b_i, \qquad i = 1, \dots, m \tag{17}$$

$$x_i \geq 0, \qquad i = 1, \dots, n. \tag{18}$$

By precomputing the lefthand side of constraints (17) as

$$\max_{a_i \in \mathcal{E}_i} \left\{a_i^\top x\right\} = \overline{a}_i^\top x + \max_{u} \left\{u^\top P_i^\top x : ||u||_2 \leq \Gamma_i\right\} = \overline{a}_i^\top x + \Gamma_i ||P_i^\top x||_2, \tag{19}$$

we can finally rewrite the problem (16) – (18) as the *robust* linear optimization problem with *ellipsoidal uncertainty* presented in Lecture 1:

$$\min.\ c^\top x \quad (20)$$

$$\text{subject to: } \overline{a}_i^\top x + \Gamma_i ||P_i^\top x||_2 \leq b_i, \quad i = 1, \dots, m \quad (21)$$

$$x_i \geq 0, \quad i = 1, \dots, n \quad (22)$$

Justify why the problem (20) – (22) is a convex optimization problem.

Note: To see why (19) holds, we can solve $\max_u \left\{ u^\top P_i^\top x : ||u||_2 \leq \Gamma_i \right\}$ simply by writing:

$$u^\top P_i^\top x \leq ||u^\top||_2 ||P_i^\top x||_2 \leq \Gamma_i ||P_i^\top x||_2 \quad (23)$$

as $||u||_2 \leq \Gamma_i$. Thus, we can see that the max value of u that satisfies (23) is obtained at

$$u = \Gamma_i \frac{P_i^\top x}{||P_i^\top x||_2} \quad \text{since then} \quad u^\top P_i^\top x = \Gamma_i \frac{(P_i^\top x)^\top P_i^\top x}{||P_i^\top x||_2} = \Gamma_i \frac{||P_i^\top x||_2^2}{||P_i^\top x||_2} = \Gamma_i ||P_i^\top x||_2$$

See also the following example Figure 2 from Lecture 1.

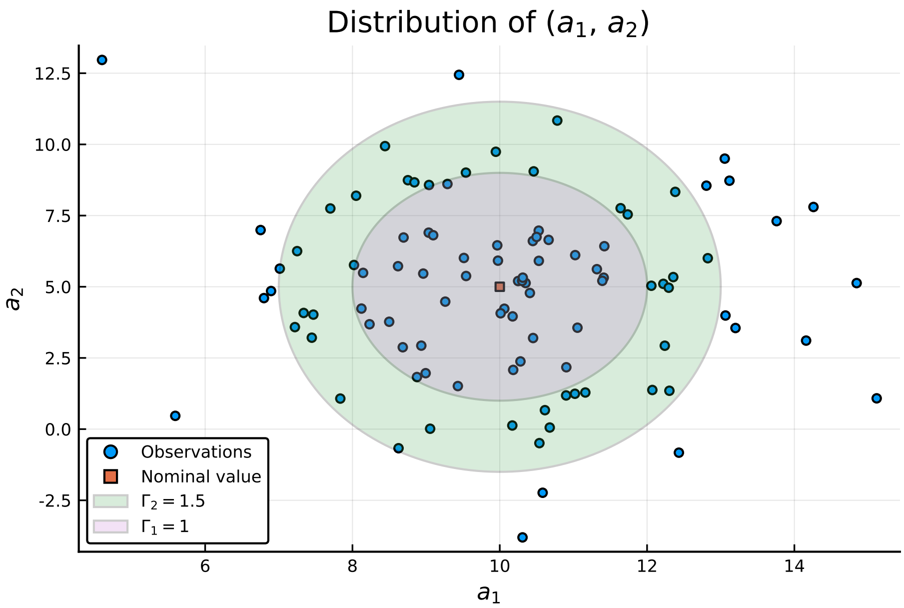

Figure 2: Robust LP example with ellipsoidal uncertainty.