This week's homework https://mycourses.aalto.fi/mod/folder/view.php?id=765849 is due no later than **Monday 11.10.2021 23:55**.

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.

Solution.

(a) Let $x, y \in \mathbb{R}^n$ and $\lambda \in [0, 1]$. By the definition of convexity, we have to show that

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y). \tag{1}$$

Using properties 3 and 4, we get

$$||\lambda x + (1 - \lambda)y|| \leq ||\lambda x|| + ||(1 - \lambda)y|| = \lambda||x|| + (1 - \lambda)||y||.$$

which is exactly (1), thus verifying the convexity of f.

(b) Let $x, y \in \mathbb{R}^n$ and $\lambda \in [0, 1]$. Applying the convexity of f, we get

$$\begin{aligned}
g(\lambda x + (1 - \lambda)y) &= \sum_{i=1}^{n} \alpha_i f_i(\lambda x + (1 - \lambda)y) \\
&\leq \sum_{i=1}^{n} \alpha_i (\lambda f_i(x) + (1 - \lambda) f_i(y)) \\
&= \lambda \sum_{i=1}^{n} \alpha_i f_i(x) + (1 - \lambda) \sum_{i=1}^{n} \alpha_i f_i(y)) \\
&= \lambda g(x) + (1 - \lambda) g(y)
\end{aligned}$$

from which we can conclude that g is convex.

(c) The epigraph of g is defined as

$$\mathbf{epi}(g) = \{(x, y) : x \in \mathbb{R}^n,\ y \in \mathbb{R},\ g(x) \leq y\} \subseteq \mathbb{R}^{n+1}.$$

We have $(x, y) \in \mathbf{epi}(g)$ if and only if $g(x) \leq y$, which holds if and only if $f_i(x) \leq y$ for all $i \in I$ (by definition of g). By defining

$$\mathbf{epi}(f_i) = \{(x, y) : x \in \mathbb{R}^n,\ y \in \mathbb{R},\ f_i(x) \leq y\} \subseteq \mathbb{R}^{n+1}$$

we get the following equivalence:

$$(x, y) \in \mathbf{epi}(g) \ \Leftrightarrow\ (x, y) \in \bigcap_{i \in I} \mathbf{epi}(f_i). \tag{2}$$

Recall from Lecture 3 that a function f is convex if and only if its epigraph $\mathbf{epi}(f)$ is convex. Thus, $\mathbf{epi}(f_i)$ is convex for all $i \in I$, and since the intersection of convex sets is also convex (see Exercise 2.1), $\bigcap_{i \in I} \mathbf{epi}(f_i)$ is also convex. Finally, $\mathbf{epi}(g)$ is convex based on (2), thus implying that g 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.

Solution.

Let $x, y \in S$ and $\lambda \in [0, 1]$. Applying the convexity of h and the monotonicity of g, we get

$$\begin{aligned} f(\lambda x + (1 - \lambda)y) &= g\big(h(\lambda x + (1 - \lambda)y)\big) \\ &\leq g\big(\lambda h(x) + (1 - \lambda)h(y)\big) & (3) \\ &\leq \lambda g\big(h(x)\big) + (1 - \lambda)g\big(h(y)\big) & (4) \\ &= \lambda f(x) + (1 - \lambda)f(y) \end{aligned}$$

which implies that $f = g(h(x))$ is convex. The first inequality (3) follows from the convexity of h and from the monotonicity of g. The second inequality (4) follows from the convexity of g.

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.,

$$a^\top x_i - b > 0,\ i = 1, \ldots, N \qquad a^\top y_i - b < 0,\ i = 1, \ldots, M \tag{5}$$

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 (5) as

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

and then relax these constraints by introducing nonnegative variables $u_1, \dots, u_N$ and $v_1, \dots, v_M$, and rewriting (6) as

$$a^\top x_i - b \geq 1 - u_i, \ i = 1, \dots, N \qquad a^\top y_i - b \leq -1 + v_i, \ i = 1, \dots, M \tag{7}$$

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.

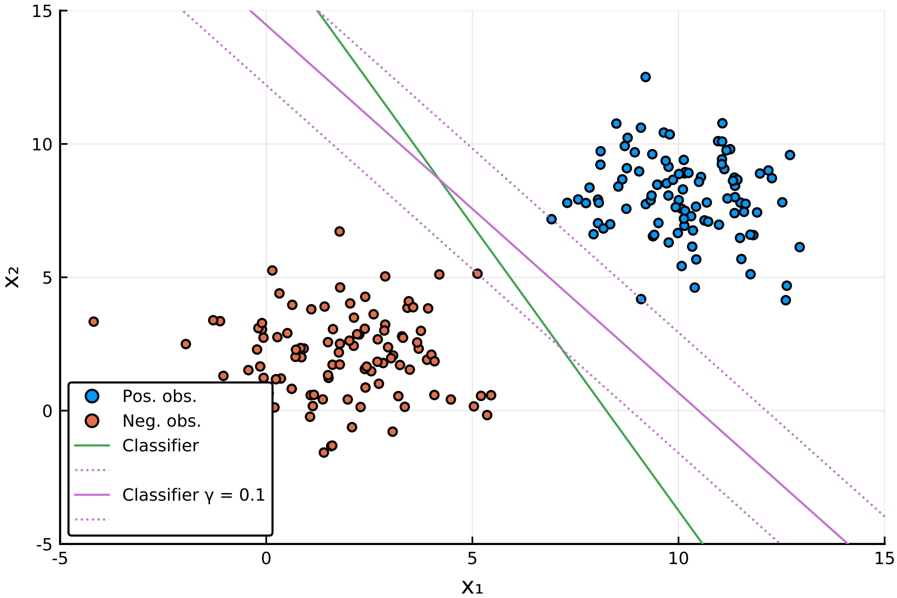

Figure 1: Classification results. Buffer zone corresponds to the area between the dotted pink lines.

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

$$\begin{aligned}
\min. \ & \sum_{i=1}^{N} u_i + \sum_{i=1}^{M} v_i + \gamma ||a||_2^2 && && (8)\\
\text{subject to: } & a^\top x_i - b + u_i \geq \ \ 1, && i = 1, \dots, N && (9)\\
& a^\top y_i - b - v_i \leq -1, && i = 1, \dots, M && (10)\\
& u_i \geq 0, && i = 1, \dots, N && (11)\\
& v_i \geq 0, && i = 1, \dots, M. && (12)
\end{aligned}$$

The first term in the objective (8) 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 (8) achieves both of these goals. The parameter $\gamma > 0$ controls the trade-off between these two goals.

Justify why the problem (8) – (12), 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

$$\min_x. \ -\lambda(\mu^\top x) + (1-\lambda)x^\top \Sigma x \tag{13}$$

$$\text{subject to: } \sum_{i \in N} x_i = 1 \tag{14}$$

$$x \geq 0 \tag{15}$$

The objective (13) 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 (13) – (15) is a convex optimization problem.

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

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

$$\text{subject to: } a_i^\top x \leq b_i, \qquad i = 1, \dots, m \tag{17}$$

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

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:

$$a_i \in \mathcal{E}_i = \{\overline{a}_i + P_i u : ||u||_2 \leq \Gamma_i\}, \tag{19}$$

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 (17) 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:

$$\min. \ c^\top x \tag{20}$$

$$\text{subject to: } \max_{a_i \in \mathcal{E}_i} \left\{a_i^\top x\right\} \leq b_i, \qquad i = 1, \dots, m \tag{21}$$

$$x_i \geq 0, \qquad i = 1, \dots, n. \tag{22}$$

By precomputing the lefthand side of constraints (21) as

$$\max_{a_i \in \mathcal{E}_i} \left\{a_i^\top x\right\} = \overline{a}_i^\top x + \max_u \left\{u^\top P_i^\top x : ||u||_2 \leq \Gamma_i\right\} = \overline{a}_i^\top x + \Gamma_i ||P_i^\top x||_2, \tag{23}$$

we can finally rewrite the problem (20) – (22) as the *robust* linear optimization problem with *ellipsoidal uncertainty* presented in Lecture 1:

$$\begin{aligned} &\text{min. } c^\top x && && (24)\\ &\text{subject to: } \overline{a}_i^\top x + \Gamma_i ||P_i^\top x||_2 \leq b_i, && i = 1, \dots, m && (25)\\ &x_i \geq 0, && i = 1, \dots, n && (26) \end{aligned}$$

Justify why the problem (24) – (26) is a convex optimization problem.

Note: To see why (23) holds, we can solve $\max_u \left\{u^\top P_i^\top x : ||u||_2 \leq \Gamma_i\right\}$ simply by writing:

$$u^\top P_i^\top x \leq ||u^\top||_2 ||P_i^\top x||_2 \leq \Gamma_i ||P_i^\top x||_2 \qquad (27)$$

as $||u||_2 \leq \Gamma_i$. Thus, we can see that the max value of u that satisfies (27) 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.

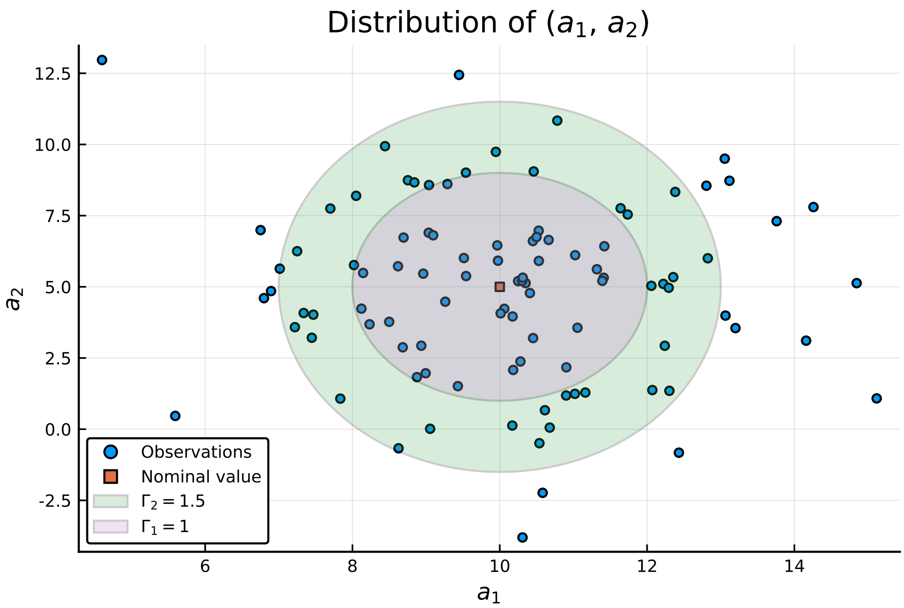

Figure 2: Robust LP example with ellipsoidal uncertainty.

Solution.

(a) To see why the objective function (8) is convex, consider the first term

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

All u_i for $i = 1, \dots, N$ and v_i for $i = 1, \dots, M$ are linear functions which are both convex and concave. The second term $\gamma||a||_2^2$ is a norm which is a convex function (see Exercise 3.1 (a)) multiplied by a positive scalar γ. Multiplying a convex function with a positive scalar is also convex (see Exercise 3.1 (b)), and therefore $\gamma||a||_2^2$ is convex.

The objective (8) is thus a sum of convex functions multiplied by positive scalars so it is convex (see Exercise 3.1 (b)). The constraints (9) – (10) are affine functions which are both convex (also concave), and (11) – (12) are linear functions which are also convex. Together these form a feasible region which is a convex set (see Figure **??** for an example).

The problem (8) – (12) consists of minimizing a convex function over a convex set, so it is a convex optimization problem. Another way to justify why it convex is because the objective function is convex and all inequality constraints are convex functions. But we will learn more about this characterization later in the course.

(b) Here the objective function (13) is a sum of two convex functions $-\mu^\top x$ and $x^\top \Sigma x$ multiplied by positive constants λ and $(1-\lambda)$, respectively. Therefore the objective (13) is a convex function (see Exercise 3.1 (b)). Justification why the individual terms are convex: The first term $-\mu^\top x$ is a linear function and therefore convex, and the second term $x^\top \Sigma x$ is a quadratic function with a positive semidefinite matrix Σ and thus convex.

Also, the constraint (14) is an affine function and thus convex, while the constraint (15) is a linear function and also convex. These constraints form an $|N| - 1$ dimensional simplex which is a convex set.

Thus, the problem (13) – (15) consists of minimizing a convex function over a convex set and is therefore a convex optimization problem. Here, an alternative characterization would be that (13) – (15) is convex, because its objective is convex, all of its equality constraints are *affine* functions, and all of its inequality constraints are convex functions.

(c) In this example, the objective (24) is linear and therefore convex, and the inequality constraints (25) – (26) represent an ellipsoid which is a convex set (see Figure **??** for an example). So here again we're minimizing a convex function over a convex set, wherefore (24) – (26) is a convex optimization problem. The other characterization which we will learn more about later, would be that the problem is convex since the objective is convex and all of its inequalities are convex functions.