Homework 1

CMU 10-725: Optimization for Machine Learning

OUT: Tuesday, Jan 20th, 2026

DUE: Tuesday, February 3rd, 2026, 11:59pm

START HERE: Instructions

- **Collaboration policy:** Collaboration on solving the homework is allowed, after you have thought about the problems on your own. To remind you, many questions in this HW have solutions that are very easy to find online (and many are from previous versions of this course). It is also OK to get clarification (but not solutions) from books or online resources, again after you have thought about the problems on your own. There are two requirements: first, cite your collaborators fully and completely (e.g., "Jane explained to me what is asked in Question 2.1"). Second, write your solution *independently*: close the book and all of your notes, and send collaborators out of the room, so that the solution comes from you only.
- **Submitting your work:**
  - **Gradescope:** For the written problems such as short answer, multiple choice, derivations, proofs, or plots, we will be using the Gradescope. The best way to format your homework is by using the Latex template released in the handout and writing your solutions in Latex. However, submissions can be handwritten onto the template, but should be labeled and clearly legible. If your writing is not legible, you will not be awarded marks.
  Regrade requests can be made after the homework grades are released, however this gives the TA the opportunity to regrade your entire paper, meaning if additional mistakes are found then points will be deducted.
  - **Programming:** You should submit all code used to solve the programming aspect of the homework to the corresponding 'Programming' submission slot on Gradescope. If you do not do this, you will not get any credit for any of the programming section irrespective of the plots and values submitted to the 'Written' submission slot.

1 Convex sets - Michael

In this problem you want to prove that some commonly used sets are convex.

1.1 The polytope (5 points)

A d-dimensional polytope $\mathcal{P}$ is a set in $\mathbb{R}^d$, defined as the set of points $x \in \mathbb{R}^d$ satisfying the following constraints: For an integer $m > 0$, for m vectors $a_1, \cdots, a_m \in \mathbb{R}^d$ and m values $b_i \in \mathbb{R}$:

$$\forall i \in [m] : \langle a_i, x\rangle \leq b_i \tag{1}$$

Show that $\mathcal{P}$ is a convex set.

Use this to show that the ℓ_1 ball: $\{x \in \mathbb{R}^d \mid \|x\|_1 \leq 1\}$ is a convex set.

Solution To show $\mathcal{P}$ is convex, let $x, y \in \mathcal{P}$ then both x and y satisfies (1). Consider their convex combination $X := \theta x + (1-\theta)y$ for $\theta \in [0,1]$ then by linearity and $X \in \mathbb{R}$ for any $i \in [m]$

$$\langle a_i, \theta x + (1-\theta)y\rangle = \theta\langle a_i, x\rangle + (1-\theta)\langle a_i, y\rangle \leq \theta b_i + (1-\theta)b_i = b_i.$$

So $X \in \mathcal{P}$ and this holds for all θ so $\mathcal{P}$ is a convex set.

To show that the ℓ_1 ball is a convex set. Let $m = 2^d$ and let a_i be the sign vectors $\{-1,1\}^d$ and

$$\langle s, x\rangle = \sum_{k=1}^{d} s_k x_k.$$

Claim: $\langle x, y\rangle$ is an inner product.

1. Conjugate symmetry holds as $x_k y_k = y_k x_k$ and since both are real then we have conjugate symmetry
2. Linearity holds because for $a, b \in \mathbb{R}$

$$\langle ax+by, z\rangle = \sum_{k=1}^{d}(ax+by)_k z_k = \sum_{k=1}^{d} ax_k + by_k z_k = a\sum_{k=1}^{d} x_k z + b\sum_{k=1}^{d} y_k z = a\langle x, z\rangle + b\langle y, z\rangle.$$

3. $\langle x, x\rangle > 0$ if $x \neq 0$ as $\sum_{k=1}^{d} x_k^2 > 0$ as $x_k^2 \geq 0$ and $x_k^2 > 0$ if $x_k > 0$ and by assumption $x_k > 0$ for some k.

We know $|x_k| = \max\{-x_k, x_k\}$, so there exist some sign vector that always chooses the maximum so the condition, $||x||_1 \leq 1$ is equivalent to satisfying

$$\langle s_k, x\rangle \leq 1 \text{ for all } k = 1, 2, ..., 2^d.$$

Since ℓ_1 ball can be represented as $\mathcal{P}$ polytope then it is a convex set.

1.2 The unit ball (5 points)

A d-dimensional unit ball $\mathcal{B}$ is a set in $\mathbb{R}^d$, defined as the set of points $x \in \mathbb{R}^d$ satisfying the following constraints:

$$\sum_{i\in[d]} x_i^2 \le 1$$

Where x_i is the i-th coordinate of x.

Show that $\mathcal{B}$ is a convex set.

(Hint: For any $a, b \in \mathbb{R}$, $2ab \le a^2 + b^2$.)

Solution Let $x, y \in \mathcal{B}$ then x, y satisfying the constraint and for $\theta \in [0,1]$, we want to show the convex combination $\theta x + (1-\theta)y$ satisfy

$$\sum_{i\in[d]} (\theta x_i + (1-\theta)y_i)^2 \le 1.$$

If we expand then we get

$$\begin{aligned}
\sum_{i\in[d]} (\theta x_i + (1-\theta)y_i)^2 &= \sum_{i\in[d]} \theta^2 x_i^2 + 2\theta(1-\theta)x_i y_i + (1-\theta)^2 y_i^2 \\
&= \sum_{i\in[d]} \theta^2 x_i^2 + \sum_{i\in[d]} 2(\theta)(1-\theta)x_i y_i + \sum_{i\in[d]} (1-\theta)^2 y_i^2 \\
&\le \theta^2 \sum_{i\in[d]} x_i^2 + \theta(1-\theta)\sum_{i\in[d]} (x_i^2 + y_i^2) + (1-\theta)^2 \sum_{i\in[d]} y_i^2 && \text{(hint)} \\
&\le \theta^2 + 2\theta(1-\theta) + (1-\theta)^2 && \text{(constraint)} \\
&= (\theta + (1-\theta))^2 \\
&= 1
\end{aligned}$$

Since this holds for any $x, y \in \mathcal{B}$ and $\theta \in [0,1]$ we have that $\mathcal{B}$ is convex.

1.3 The linear transformation (5 points)

Suppose $\mathcal{D}$ is a convex set in $\mathbb{R}^d$. For any matrix $A \in \mathbb{R}^{d\times d}$ and vector $b \in \mathbb{R}^d$, show that the following set is also convex:

$$\mathcal{C} = \{x \in \mathbb{R}^d \mid Ax + b \in \mathcal{D}\}$$

Solution Take any $x, y \in \mathcal{C}$ then for $\theta \in [0,1]$ we want to show $\theta x + (1-\theta)y \in \mathcal{C}$. This is equivalent to showing $A(\theta x + (1-\theta)y) + b \in \mathcal{D}$.
We can split b to write the expression as $\theta(Ax+b) + (1-\theta)(Ay+b)$ then since $Ax + b \in \mathcal{D}$, $Ay + b \in \mathcal{D}$ and $\mathcal{D}$ is convex then we have $\theta(Ax+b) + (1-\theta)(Ay+b) \in \mathcal{D}$. Consequently, $\mathcal{C}$ is a convex set.

1.4 Ellipsoid (5 points)

Use the previous two subproblems to show an ellipsoid $\mathcal{E}$ is convex, where for a matrix A in $\mathbb{R}^{d\times d}$ and vector $b \in \mathbb{R}^d$:

$$\mathcal{E} = \{x \in \mathbb{R}^d \mid (x-b)^\top A^\top A(x-b) \leq 1\}. \tag{2}$$

Solution We can rewrite our constraint as

$$\begin{aligned}
(x-b)^T A^T A(x-b) &= [A(x-b)]^T[A(x-b)] && ((Ax)^T = x^T A^T) \\
&= ||A(x-b)||_2^2 && (||x||_2^2 = x^T x\ )(*)
\end{aligned}$$

Then since our norm is always non-negative we have $||A(x-b)||_2 \leq 1$.
Let $\mathcal{B} = \{c \in \mathbb{R}^d | ||x||_2 \leq 1\}$ be the euclidean unit ball then by 1.2, $\mathcal{B}$ is convex as $||x||_2^2 = \sum_{i\in[d]} x_i^2$. We know $A(x-b) \in \mathcal{B}$ so we can rewrite $\mathcal{E}$ as

$$\mathcal{E} = \{x \in \mathbb{R}^d | Ax - Ab \in \mathcal{B}\}.$$

with $A \in \mathbb{R}^{d,d}$ and $Ab \in \mathbb{R}^d$ then by 1.3 we have $\mathcal{E}$ is a convex set.

2 Convex Functions (Michael)

In this problem you want to show that some commonly used functions are convex. You can use the basic definition or the alternative definition mentioned in class in the second lecture.

2.1 The max Operation (5 points)

Suppose $f_1, \cdots, f_m$ are convex functions over $\mathbb{R}^d$, show that

$$f(x) = \max\{f_1(x), \cdots, f_m(x)\}$$

is a convex function over $\mathbb{R}^d$.

What about $g(x) = \min\{f_1(x), \cdots, f_m(x)\}$?

Solution Let $x, y \in \mathbb{R}^d$ then for some $\theta \in [0,1]$ we have $f_1, ..., f_m$ $f(\theta x + (1-\theta)y) = f_j(\theta x + (1-\theta)y)$ for some $j \in [m]$. By convexity

$$f_j(\theta x + (1-\theta)y) \le \theta f_j(x) + (1-\theta) f_j(y) \le \theta f(x) + (1-\theta) f(y).$$

So the function is convex over $\mathbb{R}^d$.

$g(x)$ is not guaranteed to be convex, consider $d = 1$, $m = 2$ and set $f_1(x) = x^2$ and $f_2(x) = (x-2)^2$. If we choose $x = 0, y = 2$ and consider the midpoint then $g\left(\frac{0+2}{2}\right) = g(1) = 1$ but $g(0) = g(2) = 0$ so $g\left(\frac{0+2}{2}\right) > \frac{1}{2}g(0) + \frac{1}{2}g(2)$ which violates convexity.

2.2 1-d Convex Functions (10 points)

(2 Point each bullet point) Prove (or disprove) the convexity of the following functions which map from $\mathbb{R} \to \mathbb{R}$.

- $f(x) = xe^x$.
- $f(x) = (\text{ReLU}(x))^c = \max(0, x)^c$ for any $c \ge 1$.
- $f(x) = \log(1 + e^x)$
- $f(x) = x \log x$ $(x > 0)$
- $f(x) = x \sin(x)$

Solution

- Claim: $f(x) = xe^x$ is not convex.
  We can compute the second derivative. We first calculate the first

  $$f'(x) = e^x + xe^x.$$

  The second would then be

  $$f''(x) = e^x + e^x + xe^x = e^x(2 + x).$$

  Notice that when $x < -2$ $f''(x) < 0$, so the function is not convex along the entire $\mathbb{R}$.

- Claim: $f(x) = (ReLU(x))^c$ is convex. Observe that

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x^c & \text{if } x > 0 \end{cases}.$$

  So the first derivative would be

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ cx^{c-1} & \text{if } x > 0 \end{cases}.$$

  For $x < 0$, the derivative is 0 so $f''(x) = 0$ in this region.
  For $x > 0$, the second derivative is $f''(x) = c(c-1)x^{c-2}$ and since $c \geq 1$, $f''(x) \geq 0$ in this region.
  For $x = 0$, the function is still convex because it is continuous and the subgradients are non-decreasing.
  Thus f is convex.

- Claim: $f(x) = \log(1 + e^x)$ is convex.
  The first derivative $f'(x) = \frac{e^x}{1+e^x}$ and second derivative

$$f''(x) = \frac{e^x(1+e^x) - e^x e^x}{(1+e^x)^2} = \frac{e^x}{(1+e^x)^2}.$$

  Since both numerator and denominator is positive, $f''(x) > 0$ for the entire $\mathbb{R}$ and thus convex.

- Claim: $f(x) = x \log x$ is convex
  The first derivative is $f'(x) = \log x + x \cdot \frac{1}{x} = \log x + 1$ and the second is

$$f''(x) = \frac{1}{x}.$$

  Since $\frac{1}{x} > 0$ along the entire region $x > 0$, f is convex and even strictly convex.

- Claim: $f(x) = x \sin(x)$ is not convex.
  The first derivative is $f'(x) = \sin(x) + \cos(x)$ so the second derivative is

$$f''(x) = \cos(x) + [1 \cdot \cos(x) + x(-\sin(x))] = 2\cos(x) - x\sin(x).$$

  Consider $x = \frac{5\pi}{2}$ then

$$f''\left(\frac{5\pi}{2}\right) = 2\cos\left(\frac{5\pi}{2}\right) - \frac{5\pi}{2}\cos\left(\frac{5\pi}{2}\right) = -\frac{5\pi}{2}.$$

  So f is not convex at $x = \frac{5\pi}{2}$.

2.3 Products and Quotients of Convex Functions (5 points)

In general, the product or quotient of two convex functions is **not** convex. Herein, we explore this further. For this question, suppose that all functions map from $\mathbb{R} \to \mathbb{R}$.

1. Construct two convex functions $f, g : \mathbb{R} \to \mathbb{R}^+$ that are positive on the entire real line and where f/g is **not** convex.
2. Prove that if f, g are convex, both non-decreasing (or non-increasing), and positive functions on $\mathbb{R}$, then fg is convex.
3. Prove that if f is convex non-decreasing, and positive, g is concave, non-increasing, and positive, then f/g is convex.

Solution

1. Consider $f(x) = 1$ and $g(x) = x^2 + 1$ then let $h(x) = \frac{f(x)}{g(x)} = \frac{1}{x^2+1}$.
   To show h is not convex we want to find a point such that $h''(x) < 0$. The first derivative is
   $$h'(x) = -\frac{2x}{(x^2+1)^2}.$$
   and second is
   $$h''(x) = \frac{(x^2+1)^2(-2) - (-2x)2(x^2+1)(2x)}{(x^2+1)^4} = \frac{6x^2-2}{(x^2+1)^3}.$$
   then $h''(0) = -2$ so h is not convex.

2. To show $h := fg$ is convex, we'll observe its second derivative. The first derivative is $h'(x) = f'(x)g(x) + f(x)g'(x)$ and second is
   $$h''(x) = f''(x)g(x) + 2f'(x)g'(x) + f(x)g''(x).$$
   Observe that $f''(x), g''(x) \geq 0$ as f and g is convex, $g(x), f(x) > 0$ as f and g are positive function and $f'(x), g'(x) \geq 0$ as non-decreasing (flip inequality if non-increasing) so each term in the second derivative is non-negative so $h''(x) \geq 0$ so h is convex.

3. We analyze the sign of the second derivative $h''(x)$. First derivative via Quotient Rule:
   $$h'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}$$
   Second derivative calculation:
   $$h''(x) = \frac{[f''(x)g(x) - f(x)g''(x)]g(x)^2 - [f'(x)g(x) - f(x)g'(x)]2g(x)g'(x)}{g(x)^4}$$
   Dividing numerator and denominator by $g(x)$:
   $$h''(x) = \frac{g(x)[f''(x)g(x) - f(x)g''(x)] - 2g'(x)[f'(x)g(x) - f(x)g'(x)]}{g(x)^3}$$

Expanding terms:

$$h''(x) = \frac{f''(x)g(x)^2 - f(x)g(x)g''(x) - 2f'(x)g'(x)g(x) + 2f(x)(g'(x))^2}{g(x)^3}$$

Now we determine the sign of each term based on the given assumptions:

- $f''(x)g(x)^2 \geq 0$: Since f is convex ($f'' \geq 0$) and squares are positive.
- $-f(x)g(x)g''(x) \geq 0$: Since $f, g > 0$ and g is concave ($g'' \leq 0 \implies -g'' \geq 0$).
- $-2f'(x)g'(x)g(x) \geq 0$: Since $g > 0$, f is non-decreasing ($f' \geq 0$), and g is non-increasing ($g' \leq 0$). The term $(-2) \cdot (\text{positive}) \cdot (\text{negative})$ is positive.
- $2f(x)(g'(x))^2 \geq 0$: Since $f > 0$ and squares are non-negative.

Since the numerator is a sum of non-negative terms and $g(x)^3 > 0$, we have $h''(x) \geq 0$. Thus, f/g is convex.

2.4 Properties of KL-Divergence (5 points)

KL-Divergence is a fundamental measure of how one probability distribution differs from another. Formally, if P, Q are discrete distributions over $\mathcal{X}$, we define

$$D_{KL}(P\|Q) = \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}.$$

In this subpart, we will explore some properties of KL.

Let $u, v \in \mathbb{R}^n$ such that $0 < u_i, v_i \leq 1$ and $\sum_{i=1}^n u_i = 1, \sum_{i=1}^n v_i = 1$. Notice that u, v can be thought of as discrete measure over a sample space of n elements. Prove that $D_{KL}(u\|v) \geq 0$. Also, show that $D_{KL}(u\|v) = 0$ if and only if $u = v$.

[*Hint:* $D_{KL}(u\|v) = f(u) - f(v) - \nabla f(v)^T(u - v)$, *for* $f(u) = \sum_{i=1}^n u_i \log u_i$. $f(u)$ *is called the negative entropy of* u.]

Solution

Lemma 1. *The sum of convex functions is convex.*
To show this, suppose $f_1, ..., f_n : \mathbb{R}^d \to \mathbb{R}$ *are convex functions then take any* $x, y \in \mathbb{R}^d$ *and* $\theta \in [0, 1]$ *we have* $f_i(\theta x + (1-\theta)y) \leq \theta f_i(\theta) + (1-\theta)f(y)$ *Consequently*

$$\sum_{i=1}^n f_i(\theta x + (1-\theta)y) \leq \sum_{i=1}^n \theta f_i(x) + (1-\theta)f_i(y).$$

Note: A sum of strictly convex functions are strictly convex, same proof but drop the equality

From problem 2.2 we've established $f(x) = x \log x$ for $x > 0$ is convex. So by Lemma 1, $f(u) = \sum_{i=1}^n u_i \log u_i$ is convex as it's the sum of convex functions. Then by first order convexity condition for any two points $x, y \in \mathbb{R}^n$

$$f(u) \geq f(v) + \nabla f(v)^T(u - v).$$

Rearranging we have $D_{KL}(u\|v) = f(u) - f(v) - \nabla f(v)^T(u-v) \geq 0$.

For the second part of the question to show $D_{KL}(u\|v) = 0$ iff $u = v$. To show the backwards direction if $u = v$ then $f(u) = f(v)$ and $u - v = 0$ so $D_{KL}(u\|v) = 0 - \nabla f(v)^T(0) = 0$. For the forward direction, We'll show the contrapositive that if $u \neq v$ then $D_{KL}(u\|v) \neq 0$. We know $f(u)$ is strictly convex as in problem 2.2 we found $f(x) = x \log x$ is strictly increasing as $f''(x) = \frac{1}{x} > 0$ in our region. The first order taylor approximation is strict underestimater for all distinct points in a strictly convex function. So $f(u) > f(v) + \nabla f(v)^T(u-v)$ and thus $D_{KL}(u\|v) > 0$ whenever $u \neq v$

2.5 Logistic Regression (5 points)

The objective function in the logistic regression problem is of the form

$$f(x) = \sum_{i \in [m]} -\log\left(\frac{1}{1 + \exp\{-y_i\langle a_i, x\rangle\}}\right),$$

where $x, a_1, \ldots, a_m \in \mathbb{R}^d$ and $y_1, \ldots, y_m \in \mathbb{R}$.

Prove that $f(x)$ is convex. [*Hint: Show that the sum of convex functions is convex and that* $x \mapsto -y_i\langle a_i, x\rangle$ *is convex for any* (a_i, y_i).]

Solution First, let us examine the inner term. Let $h_i(x) = -y_i\langle a_i, x\rangle$. This function is affine. We can verify linearity:

$$h_i(tx+(1-t)z) = -y_i\langle a_i, tx+(1-t)z\rangle = t(-y_i\langle a_i, x\rangle)+(1-t)(-y_i\langle a_i, z\rangle) = th_i(x)+(1-t)h_i(z).$$

Since equality holds, Jensen's inequality holds, so $h_i(x)$ is convex (and concave).

From part 2.2, we established that the function $g(z) = \log(1 + e^z)$ is convex.

The composition of a convex function with an affine mapping is convex. Since $g(z)$ is convex and $h_i(x)$ is affine, the composition $f_i(x) = g(h_i(x)) = \log(1 + \exp(-y_i\langle a_i, x\rangle))$ is convex with respect to x.

Finally, since the sum of convex functions is convex, the total objective function:

$$f(x) = \sum_{i=1}^{m} f_i(x) = \sum_{i=1}^{m} \log(1 + \exp(-y_i\langle a_i, x\rangle))$$

is convex.

3 Characterizations of Convexity (10 points) (Johnna)

Throughout this question suppose that your function f is twice differentiable on $\mathbb{R}^d$. In lecture we discussed three characterizations of convexity. In this question we will explore some of these characterizations.

1. **(3 pts)** Show that if f is convex and differentiable, then it must satisfy the condition that for any pair $x, y \in \mathbb{R}^d$,

$$f(y) \geq f(x) + \nabla f(x)^T (y - x).$$

You can use the following directional-derivative characterization of the gradient

$$\nabla f(x)^T v = \lim_{t \to 0} \frac{f(x + tv) - f(x)}{t}.$$

Solution Let $t \in [0, 1]$ then by convexity for any $x, y \in \mathbb{R}^d$ $f((1-t)x + ty) \leq (1-t)f(x) + tf(y)$
We can rewrite this to become

$$f(x + t(y - x)) \leq f(x) + t(f(y) - f(x))$$
$$\frac{f(x + t(y - x)) - f(x)}{t} \leq f(y) - f(x)$$

This holds for any $t \in [0, 1]$ so consider the limit as $t \to 0$ then

$$\lim_{t \to 0} \frac{f(x + t(y - x)) - f(x)}{t} = \nabla f(x)^T (y - x) \leq f(y) - f(x).$$

Rearranging we get the result

$$f(y) \geq f(x) + \nabla f(x)^T (y - x).$$

2. **(4 pts)** Show that if f is differentiable, and has monotone gradient (i.e., we have that $(\nabla f(x) - \nabla f(y))^T (x - y) \geq 0$), then it is convex by the first-order characterization, i.e. satisfies that for any x, y,

$$f(y) \geq f(x) + \nabla f(x)^T (y - x).$$

The second fundamental theorem of calculus is useful to recall: for any differentiable f on $[0, 1]$, $\int_0^1 f'(t)dt = f(1) - f(0)$.

Particularly, an expression you might find useful to play with (try to bound it or re-express it using the fundamental theorem etc.) is,

$$I_1 := \int_0^1 \frac{d}{dt} f((1-t)x + ty)dt.$$

Solution We compute

$$\frac{d}{dt}f((1-t)x+ty) = \nabla f((1-t)x+ty)^T(y-x). \tag{1}$$

Then by fundamental theorem of calculus

$$\int_0^1 \frac{d}{dt}f((1-t)x+ty)dt = f(y)-f(x). \tag{2}$$

We want to show that $f(y)-f(x)-\nabla f(x)^T(y-x) \geq 0$, we can rewrite the LHS using (1) and (2)

$$f(y)-f(x)-\nabla f(x)^T(y-x) = \int_0^1 \nabla f((1-t)x+ty)^T(y-x) - \nabla f(x)^T(y-x)dt$$

Since $\nabla f(x)^T(y-x)$ does not depend on t then we can include it directly in our integral. Combining the terms in the integral we have

$$\int_0^1 \nabla f((1-t)x+ty)^T(y-x)-\nabla f(x)^T(y-x)dt = \int_0^1 (\nabla f((1-t)x+ty)-\nabla f(x))^T(y-x)dt.$$

Since f has monotone gradient $(\nabla f((1-t)x+ty)-\nabla f(x))^T(y-x) \geq 0$ for all $t \in [0,1]$ so $f(y)-f(x)-\nabla f(x)^T(y-x) \geq 0$ and thus f is convex.

3. **(3 pts)** Show that if the epigraph of function $f : \mathbb{R}^n \to \mathbb{R}$ is convex, then f is convex.

$$\text{Epi}(f) = \{(x,t) : x \in \text{dom}(f), t \geqslant f(x)\}$$

Solution Assume Epi(f) is convex then for any pairs $(x_1,t_1),(x_2,t_2) \in \text{Epi}(f)$ we have that for $\theta \in [0,1]$, $\theta(x_1,t_1)+(1-\theta)(x_2,t_2) \in \text{Epi}(f)$. So $\theta x_1 + (1-\theta)x_2 \in dom(f)$ and $\theta t_1 + (1-\theta)t_2 \geq f(\theta x_1 + (1-\theta)x_2)$. We know that $(x, f(x)) \in \text{Epi}(f)$ as trivially $f(x) \geq f(x)$. So taking any pairs $(x_1, f(x_1))$ and $(x_2, f(x_2))$ then $\theta f(x_1)+(1-\theta)f(x_2) \geq f(\theta x_1 + (1-\theta)x_2)$. So f is convex.

4 Partial Minimization (8 points) (Julia)

1. **(3pts)** Let $f : \mathbb{R}^d \times \mathbb{R}^d \to (-\infty, \infty)$ be convex and $\mathbf{C}$ be a convex set in $\mathbb{R}^d$.

Consider the partial minimization of f over the set $\mathbf{C}$:

$$g(x) = \inf_{y \in \mathbf{C}} f(x,y),$$

and assume that g is always finite. Show that g is convex.

Solution For any $x_1, x_2 \in C$ and $\epsilon > 0$ there exist y_1, y_2

$$g(x_1) + \frac{\epsilon}{2} \geq f(x_1,y_1) \text{ and } g(x_2) + \frac{\epsilon}{2} \geq f(x_2,y_2). \tag{Def of inf}$$

Define $x_\theta = \theta x_1 + (1-\theta)x_2$ and $y_\theta = \theta y_1 + (1-\theta)x_2$ and $x_\theta, y_\theta \in C$ by convexity of C. So

$$\begin{aligned} g(x_\theta) &\le f(x_\theta, y_\theta) && \text{(def of inf)} \\ &\le f(\theta x_1 + (1-\theta)x_2, \theta y_1 + (1-\theta)x_2) \\ &\le \theta f(x_1, y_1) + (1-\theta) f(x_2, y_2) && \text{(convexity of } f\text{)} \\ &\le \theta g(x_1) + (1-\theta) g(x_2) + \epsilon && \text{(def of } y_1, y_2\text{)} \end{aligned}$$

So

$$g(\theta x_1 + (1-\theta)x_2) \le \theta g(x_1) + (1-\theta) g(x_2).$$

So g is convex.

2. **(2pts)** Let h_1 and h_2 be two convex functions from $\mathbb{R} \to \mathbb{R}$. Let us define

$$h_1 \Box h_2(x) = \inf_{u \in \mathbb{R}} h_1(u) + h_2(x-u)$$

Is $h_1 \Box h_2$ convex?

Solution Let $g(x) = h_1 \Box h_2$ then for any $x, y \in \mathbb{R}$ and $t \in [0,1]$ we have

$$g(tx + (1-t)y) \le h_1(w) + h_2(tx + (1-t)y - w). \qquad \text{(This holds } \forall w \in \mathbb{R}\text{)}$$

Then for any $u, v \in \mathbb{R}$ let $w = tu + (1-t)v$ and we have

$$\begin{aligned} g(tx + (1-t)y) &\le h_1(w) + h_2(tx + (1-t)y - w) \\ &= h_1(tu + (1-t)v) + h_2(tx + (1-t)y - (tu + (1-t)v)) \\ &= h_1(tu + (1-t)v) + h_2(t(x-u) + (1-t)(y-v)) \\ &\le t h_1(u) + (1-t) h_1(v) + t h_2(x-u) + (1-t) h_2(y-v) \end{aligned}$$

Last inequality from convexity of h_1, h_2 and rearranging we have

$$t\left(h_1(u) + h_2(x-u)\right) + (1-t)\left(h_1(v) + h_2(y-v)\right). \tag{1}$$

Since $g(tx + (1-t)y)$ is a lower bound for (1) for any w so it holds for any choice of u, v thus

$$g(tx + (1-t)y) \le \inf_{u,v \in \mathbb{R}} t\left(h_1(u) + h_2(x-u)\right) + (1-t)\left(h_1(v) + h_2(y-v)\right).$$

Because the terms involving u, v are separable so

$$\begin{aligned} g(tx + (1-t)y) &\le \inf_{u,v \in \mathbb{R}} t\left(h_1(u) + h_2(x-u)\right) + (1-t)\left(h_1(v) + h_2(y-v)\right) \\ &= t \inf_{u \in \mathbb{R}} \left(h_1(u) + h_2(x-u)\right) + (1-t) \inf_{v \in \mathbb{R}} \left(h_1(v) + h_2(y-v)\right) \\ &\le t g(x) + (1-t) g(y) \end{aligned}$$

So g is convex.

3. **(3pts)** Let $f : \mathbb{R} \to \mathbb{R}$ be convex. Define $f^* : \mathbb{R} \to \mathbb{R}$ as

$$f^*(y) = -\inf_{x\in\mathbb{R}} (f(x) - xy)$$

Is $f^*(y)$ concave or convex? Was the convexity of $f(x)$ necessary for your conclusion?

Solution

Lemma 2. $\sup_{x\in\mathbb{R}} f(x) + g(x) \le \sup_{x\in\mathbb{R}} f(x) + \sup_{x\in\mathbb{R}} g(x)$

Consider the set $S = \{f(x)+g(x) | x \in \mathbb{R}\}$ *then by definition of* $\sup$ *we have* $f(x) \le M_f$ *and* $g(x) \le M_g$ *where* $M_f = \sup_{x\in\mathbb{R}} f(x)$ *and* $M_g = \sup_{x\in\mathbb{R}} g(x)$*. Then* $f(x) + g(x) \le M_f + M_g$ *so* $M_f + M_g$ *is an upper bound for* S *and since* $\sup(S)$ *is the least upper bound* $\sup(S) \le M_f + M_g$ *and result follows.*

We can rewrite $f*(y) = \sup_{x\in\mathbb{R}}(xy - f(x))$ For any $u, v \in \mathbb{R}$ and $t \in [0, 1]$

$$\begin{aligned}
f^*(tu + (1-t)v) &= \sup_{x\in\mathbb{R}}(x(tu + (1-t)v) - tf(x) - (1-t)f(x)) \\
&= \sup_{x\in\mathbb{R}}(t(ux - f(x)) + (1-t)(vx - f(x))) \\
&\le t\sup_{x\in\mathbb{R}}(ux - f(x)) + (1-t)\sup_{x\in\mathbb{R}}(vx - f(x)) && \text{(Lemma 2)} \\
&= tf^*(u) + (1-t)f^*(v)
\end{aligned}$$

So f^* is convex and is not dependent on convexity of f^*.

5 Optimization with CVX (20 points) (Canary)

CVX is a framework for disciplined convex programming: it's rarely the fastest tool for the job, but it's widely applicable, and so it's a great tool to be comfortable with. In this exercise we will set up the CVX environment and solve a convex optimization problem.

Generally speaking, for homeworks in this class, your solution to programming-based problems should include plots and whatever explanation necessary to answer the questions asked. In addition, your full code should be submitted to the Homework 1 Gradescope submission slot otherwise you will not get credit for the programming section.

CVX variants are available for each of the major numerical programming languages. There are some minor syntactic and functional differences between the variants but all provide essentially the same functionality. Download the CVX variant of your choosing:

- Matlab: `http://cvxr.com/cvx/`
- Python: `http://www.cvxpy.org/`
- R: `https://cvxr.rbind.io`
- Julia: `https://github.com/JuliaOpt/Convex.jl`

and consult the documentation to understand the basic functionality. Make sure that you can solve the least squares problem $\min_\beta \|y - X\beta\|_2^2$ for an arbitrary vector y and matrix X. Check your answer by comparing with the closed-form solution $(X^TX)^{-1}X^Ty$.

Note: There are certain quirks of CVX that may result in you getting strange errors even if your code is technically correct. We strongly recommend setting your solver to the Splitting Conic Solver (SCS) and sticking to CVX specific functions such as `sum_squares` and `quad_form` if you encounter such errors when attempting the problems below.

Given labels $y \in \{-1, 1\}^n$, and a feature matrix $X \in \mathbb{R}^{n\times p}$ with rows $x_1, \dots x_n$, recall the support vector machine (SVM) problem

$$\begin{aligned}
\min_{\beta,\beta_0,\xi} \quad & \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^n \xi_i \\
\text{subject to} \quad & \xi_i \geq 0,\ i = 1, \dots n \\
& y_i(x_i^T\beta + \beta_0) \geq 1 - \xi_i,\ i = 1, \dots n.
\end{aligned}$$

1. **(5 pts)** Load the training data in `xy_train.csv`. This is a matrix of $n = 200$ row and 3 columns. The first two columns give the first $p = 2$ features, and the third column gives the labels. Using CVX, solve the SVM problem with $C = 1$. Report the optimal criterion value, and the optimal coefficients $\beta \in \mathbb{R}^2$ and intercept $\beta_0 \in \mathbb{R}$.

2. **(5 pts)** Recall that the SVM solution defines a hyperplane

$$\beta_0 + \beta^T x = 0,$$

which serves as the decision boundary for the SVM classifier. Plot the training data and color the points from the two classes differently. Draw the decision boundary on top.

3. **(5 pts)** Now define $\widetilde{X} \in \mathbb{R}^{n \times p}$ to have rows $\widetilde{x}_i = y_i x_i$, $i = 1, \ldots, n$, and solve using CVX the problem

$$\begin{array}{ll} \max\limits_{w} & -\frac{1}{2} w^T \widetilde{X} \widetilde{X}^T w + 1^T w \\ \text{subject to} & 0 \leq w \leq C1, \ w^T y = 0, \end{array}$$

(Above, we use 1 to denote the vector of all 1s.) Report the optimal criterion value; it should match that from part (1). Also report $\widetilde{X}^T w$ at the optimal w; this should mach the optimal β from part (1). Note: this is not a coincidence, and is an example of *duality*, as we will study in detail later in the course.

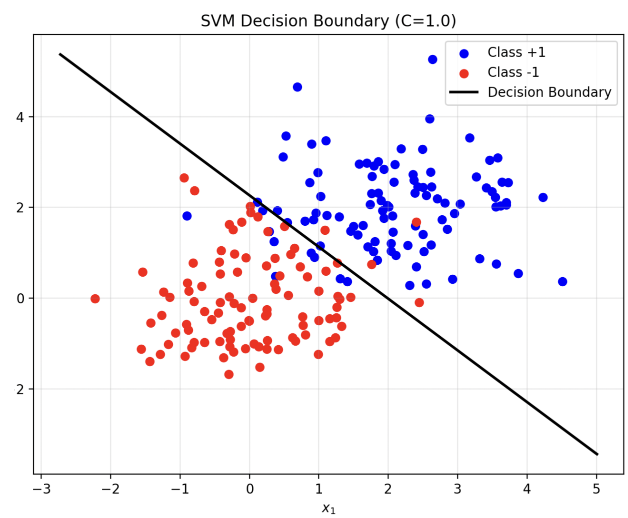
4. **(5 pts)** Investigate many values of the cost parameter $C = 2^a$, as a varies from -5 to 5. For each one, solve the SVM problem, form the decision boundary, and calculate the misclassification error on the test data in `xy_test.csv`. Make a plot of misclassification error (y-axis) versus C (x-axis, which you will probably want to put a log scale). Evaluate at least 50 points in the discretization.

Solution

1. Optimal Criterion Value: 36.7489

   Optimal Coefficients β: $\begin{pmatrix} 1.41967191 \\ 1.24607477 \end{pmatrix}$

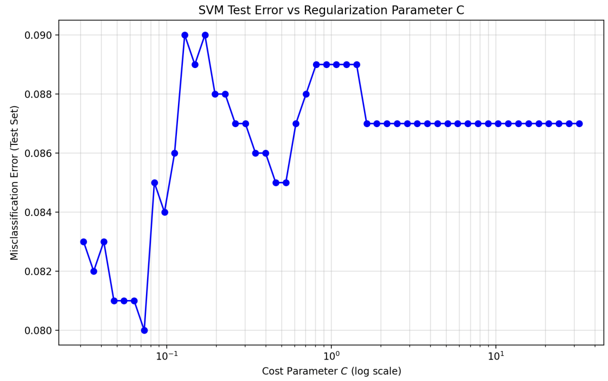
   Optimal Intercept β_0: -2.8237



2.

3. Optimal Criterion Value: 36.5468

   $\tilde{X}^T w = \begin{pmatrix} 1.41922772 \\ 1.24558069 \end{pmatrix}$



4.

Important: Remember that you **MUST** submit all code used in this part to the Programming submission slot on Gradescope otherwise you will not get credit for this section.

6 Collaboration Questions

1. (a) Did you receive any help whatsoever from anyone in solving this assignment? **Solution** No.

   (b) If you answered 'yes', give full details (e.g. "Jane Doe explained to me what is asked in Question 3.4")

   Solution

2. (a) Did you give any help whatsoever to anyone in solving this assignment? **Solution** No.

   (b) If you answered 'yes', give full details (e.g. "I pointed Joe Smith to section 2.3 since he didn't know how to proceed with Question 2")

   Solution

3. (a) Did you find or come across code that implements any part of this assignment? **Solution** No.

   (b) If you answered 'yes', give full details (book & page, URL & location within the page, etc.).

   Solution