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 YOUR SOLUTION HERE

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:

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

Where x_i is the i-th coordinate of x.

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

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

Solution YOUR SOLUTION HERE

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 YOUR SOLUTION HERE

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 YOUR SOLUTION HERE

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 YOUR SOLUTION HERE

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 $c \geq 1$.
- $f(x) = \log(1 + e^x)$
- $f(x) = x \log x$ $(x > 0)$
- $f(x) = x \sin(x)$

Solution YOUR SOLUTION HERE

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 YOUR SOLUTION HERE

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 YOUR SOLUTION HERE

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 YOUR SOLUTION HERE

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 YOUR SOLUTION HERE

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 YOUR SOLUTION HERE

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

YOUR SOLUTION HERE

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 YOUR SOLUTION HERE

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 YOUR SOLUTION HERE

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

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

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

Solution YOUR SOLUTION HERE

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 $(X^T X)^{-1} X^T y$.

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

$$\begin{aligned} \max_{w} \quad & -\frac{1}{2} w^T \widetilde{X} \widetilde{X}^T w + 1^T w \\ \text{subject to} \quad & 0 \le w \le C1, \; w^T y = 0, \end{aligned}$$

(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 YOUR SOLUTION HERE

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** Yes / 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** Yes / 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** Yes / No.

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

   Solution