Homework 1
Convex Sets and Convex Functions [*]

10-425/10-625 Introduction to Convex Optimization
http://425.mlcourse.org

OUT: Sep. 7, 2023
DUE: Sep. 17, 2023
TAs: Roochi and Akash

Instructions

- **Collaboration Policy**: Please read the collaboration policy in the syllabus.
- **Late Submission Policy:** See the late submission policy in the syllabus.
- **Submitting your work:** You will use Gradescope to submit answers to all questions and code.
  - **Written:** You will submit your completed homework as a PDF to Gradescope. For each problem, please clearly indicate the question number (e.g. 3.2). Submissions can be handwritten, but must be clearly legible; otherwise, you will not be awarded marks. Alternatively, submissions can be written in LaTeX. You may use the LaTeX source of this assignment (included in the handout .zip) as your starting point. For multiple choice / select all questions, simply write the letter(s) (e.g. A, B, C) corresponding to your chosen answer.
  - **Programming:** You will submit your code for programming questions to Gradescope. There is no autograder. We will examine your code by hand and may award marks for its submission.
- **Materials:** The data and reference output that you will need in order to complete this assignment is posted along with the writeup and template on the course website.

Question	Points
Optimization Basics	5
Convex Sets	26
Convex Functions	27
Characterizations of Convexity	10
Optimization with CVX	27
Collaboration Questions	2
Total:	97

[*]Compiled on Thursday 7$^{\text{th}}$ September, 2023 at 15:12

1 Optimization Basics (5 points)

1.1. (1 point) **True or False:** For a convex optimization problem with objective function f, where x is a local minimum and x^* a global minimum, $f(x)$ may be greater than $f(x^*)$.

A. True

B. False

1.2. (1 point) **True or False:** If $f(x) : \mathbb{R} \to \mathbb{R}$ is a convex function, then the following is a valid constraint in a convex optimization problem: $\sin(x) \leq 0$

A. True

B. False

1.3. (2 points) **Select all that apply:** Given $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^n, c \in \mathbb{R}^m$, and $f(x) = A(x+b)+c$. Under which of the following conditions is $f(x)$ affine but not linear?

A. $Ab = 0; c \neq 0$

B. $Ab \neq 0; c = 0$

C. $Ab + c = 0$

D. $Ab + c \neq 0$

E. None of the above

1.4. (1 point) **Short answer:** Consider the following convex optimization problem for $x \in \mathbb{R}$.

$$\begin{aligned} \min_x \quad & x^2 + 3x + 2 \\ \text{subject to} \quad & -\log(x) \leq 1 \end{aligned}$$

What are the implicit constraints for this problem?

2 Convex Sets (26 points)

2.1. (a) (3 points) **Show** that a polyhedron $\{x \in \mathbb{R}^n : Ax \leq b\}$, for some $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$ is convex.

(b) (3 points) **Show** that the polyhedron is closed.

2.2. (4 points) Let $A \in \mathbb{R}^{m \times n}$. **Show** that if $S \subseteq \mathbb{R}^m$ is convex then so is $A^{-1}(S) = \{x \in \mathbb{R}^n : Ax \in S\}$, which is called the preimage of S under the map $A : \mathbb{R}^n \to \mathbb{R}^m$.

2.3. (a) (3 points) **Show** that if $S_i \subseteq \mathbb{R}^n, i \in I$ is a collection of convex sets, then their intersection $\cap_{i \in I} S_i$ is also convex.

(b) (3 points) **Show** that the same statement holds if we replace "convex" with "closed".

2.4. (a) (4 points) **Show** that the unit ball $\mathcal{B} := \{x : \|x\|_2 \leq 1\}$ is convex

(b) (3 points) **Show** that If $\mathcal{D}$ is a convex set in $\mathbb{R}^d$ and $A \in \mathbb{R}^{d \times d}$ is a matrix and $b \in \mathbb{R}^d$ is a vector then

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

is also convex.

(c) (3 points) **Show** that an Ellipsoid is convex: An Ellipsoid $\mathcal{E}$ in $\mathbb{R}^d$ is a set defined as: for some matrix A in $\mathbb{R}^{d \times d}$ and vector $b \in \mathbb{R}^d$:

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

3 Convex Functions (27 points)

3.1. (3 points) **Prove** that $f(x) = e^x$ is convex using the zero'th-order, first-order, or second-order criterion of a convex function

3.2. (4 points) **Prove** that $f(x) = \log(1 + e^x)$ is convex using the zero'th-order, first-order, or second-order criterion of a convex function

3.3. (5 points) **Prove** that $f(x) = ReLU(x)$ is convex using the zero'th-order, first-order, or second-order criterion of a convex function

3.4. (5 points) Suppose f is a convex function over $\mathbb{R}^d$, **prove** that for any positive integer d', any matrix $A \in \mathbb{R}^{d \times d'}$ and any vector $b \in \mathbb{R}^d$, $g(x) := f(Ax + b)$ is a convex function over $\mathbb{R}^{d'}$ as well.

3.5. (5 points) Suppose $f_1, \cdots, f_m$ are convex functions over $\mathbb{R}^d$, **show** that for any $\lambda_1, \cdots, \lambda_m \geq 0$,

$$f(x) := \sum_{i \in [m]} \lambda_i f_i(x)$$

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

3.6. (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 $x, a_1, \ldots, a_m \in \mathbb{R}^d$. **Prove** that $f(x)$ is convex.

4 Characterizations of Convexity (10 points)

4.1. (10 points) **Show** that if $f : \mathbb{R} \to \mathbb{R}$ is convex then the fact that $f''(x) \geq 0$ for every x, implies that for any x, y,

$$f(y) \geq f(x) + f'(x)(y - x).$$

Throughout this question assume that the function f is twice differentiable on $\mathbb{R}$.

Hint: One possible route to show this is to first show that (a) a function which is *monotone* satisfies the above inequality (i.e. is convex by the first-order characterization), and then to (b) show that a function whose second derivative is ≥ 0 is monotone.

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

Particularly, two expressions you might find useful to play with (try to bound them or re-express them using the fundamental theorem etc.) are:

$$I_1 := \int_0^1 \frac{d}{dt} f((1-t)x + ty)dt,$$
$$I_2 := \int_0^1 \frac{d}{dt} \left(f'((1-t)x + ty)\right) dt \in \mathbb{R}.$$

5 Optimization with CVX (27 points)

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 corresponding programming slot on Gradescope—there is no auto-grader.

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 python CVX variant:

- Python: `http://www.cvxpy.org/`

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 $(X^TX)^{-1}X^Ty$.

Given labels $y \in \{-1, 1\}^n$, and a feature matrix $X \in \mathbb{R}^{n\times p}$ with rows $x_1, \dots x_n \in \mathbb{R}^p$, recall the hard-margin support vector machine (SVM) problem over parameters $\beta \in \mathbb{R}^p$ and $\beta_0 \in \mathbb{R}$:

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

5.1. (5 points) **Numerical answer:** Load the training data in `TrainLinSep.csv` and `TrainNoSep.csv`. Both of these datasets are matrices of $n = 200$ rows and 3 columns. The first two columns give the first $p = 2$ features, and the third column gives the labels. Using CVX, solve the hard-margin SVM problem on both datasets. Report the optimal coefficients $\beta \in \mathbb{R}^2$ and intercept $\beta_0 \in \mathbb{R}$ (round your answers to four digits after the decimal point). If the problem has no solution, answer "infeasible".

5.2. (4 points) **Plot:** Recall that the SVM solution defines a hyperplane

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

which serves as the decision boundary for the SVM classifier. For each of the two datasets, plot the training data and color the points from the two classes differently. Draw the decision boundary learned by the hard-margin SVM on top. If the optimal β, β_0 do not exist, answer "infeasible".

Hint: use the plotDecisionBoundary() function provided in order to produced plots.

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 soft-margin support vector machine (SVM) problem over parameters $\beta \in \mathbb{R}^p$ and $\beta_0 \in \mathbb{R}$ and slack variables $\xi_i \in \mathbb{R}$:

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

5.3. (5 points) **Numerical answer:** Now use CVX to solve the *soft-margin* SVM problem on `TrainLinSep.csv` and `TrainNoSep.csv` with $C = 0.1$. Report the optimal coefficients $\beta \in \mathbb{R}^2$ and intercept $\beta_0 \in \mathbb{R}$ (round your answers to four digits after the decimal point). If the problem has no solution, answer "infeasible".

5.4. (4 points) **Plot:** Now use the coefficients from the *soft-margin* SVM formulation to plot the learned decision boundary with $C = 0.1$. For each of the two datasets, plot the training data and color the points from the two classes differently. Draw the decision boundary on top. If the optimal β, β_0 do not exist, answer "infeasible".

Hint: use the plotDecisionBoundary() function provided in order to produced plots.

5.5. (3 points) **Short answer:** How does the separability of the data affect the SVM solution for the hard-margin formulation and the soft-margin formulation?

5.6. (6 points) **Plot:** Investigate many values of the cost parameter $C = 2^a$, as a varies from -5 to 5. For each one, solve the soft-margin SVM problem, form the decision boundary, and calculate the misclassification error on the *test* data. Make two plots (one for each test dataset: `TestLinSep.csv` and `TestNoSep.csv`) of misclassification error (y axis) versus C (x axis, which you should plot on log_2 scale).

Note: Set solver to ECOS if you're having issues with the default solver not yielding a solution.

5.7. (0 points) Did you upload your code to the appropriate programming slot on Gradescope?
Hint: The correct answer is 'yes'.

6 Collaboration Questions (2 points)

After you have completed all other components of this assignment, report your answers to these questions regarding the collaboration policy. Details of the policy can be found in the syllabus.

6.1. (1 point) Did you collaborate with anyone on this assignment? If so, list their name or Andrew ID and which problems you worked together on.

6.2. (1 point) Did you find or come across code that implements any part of this assignment? If so, include full details.