Labs
Optimization for Machine Learning
Spring 2024

Saarland University
CISPA Helmholtz Center for Information Security
Sebastian Stich
TAs: Yuan Gao & Xiaowen Jiang
https://cms.cispa.saarland/optml24/

Problem Set 1, April 16, 2024 (Convexity, Python Setup)

Convexity

Exercise 1. Recognizing convex functions

- *Give an example of a function $f : \mathbb{R} \to \mathbb{R}$ which is convex but not strictly convex.*
- *Give an example of a function $f : \mathbb{R} \to \mathbb{R}$ which is strictly convex yet not bounded below.*

Exercise 2. Properties of convexity

- *Give an example of two univariate convex functions f and g such that $h(\mathbf{x}) = f(\mathbf{x})g(\mathbf{x})$ is not convex.*
- *Give an example of two univariate convex functions f and g such that $h(\mathbf{x}) = f(g(\mathbf{x}))$ is not convex.*
- *Give an example of two univariate convex functions f and g such that $h(\mathbf{x}) = \min(f(\mathbf{x}), g(\mathbf{x}))$ is not convex.*

Exercise 3. Recognizing simple quadratics

Let $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c$ be a function where $\mathbf{A} \in \mathbb{R}^{d\times d}$ is a matrix and $\mathbf{x}, \mathbf{b} \in \mathbb{R}^d$ are vectors.

- *What is the necessary and sufficient condition on $\mathbf{A}$ for f to be convex?*
- *What is the sufficient condition on $\mathbf{A}$ for f to be strictly convex?*

Exercise 4. Norms

- Show that any norm defined on $\mathbb{R}^d$ is convex.
- Solve Exercise **??** from the lecture note:

  A seminorm is a function $f : \mathbb{R}^d \to \mathbb{R}$ satisfying the following two properties for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ and all $\lambda \in \mathbb{R}$.

  (i) $f(\lambda\mathbf{x}) = |\lambda| f(\mathbf{x})$,
  (ii) $f(\mathbf{x} + \mathbf{y}) \le f(\mathbf{x}) + f(\mathbf{y})$ (triangle inequality).

  Prove that every seminorm is convex.

Exercise 5. Jensen's inequality

*Prove Jensen's inequality (Lemma **??**): Let $f : \mathbb{R}^d \to \mathbb{R}$ be a convex function. Show that for any $\mathbf{x}_1, ..., \mathbf{x}_n \in \mathbb{R}^d$ and any $\lambda_1, ..., \lambda_n \ge 0$ such that $\sum_{i=1}^n \lambda_i = 1$, it holds that:*

$$f(\lambda_1\mathbf{x}_1 + ... + \lambda_n\mathbf{x}_n) \le \lambda_1 f(\mathbf{x}_1) + ... + \lambda_n f(\mathbf{x}_n)$$

Let $\mathbf{X}$ be a random variable and let $\mathbb{E}$ denote expectation, what can you say about $f(\mathbb{E}[\mathbf{X}])$?

Exercise 6. Least square *(Exercise **??** from lecture note)*

Suppose that we have centered observations $(\mathbf{x}_i, y_i)$ such that $\sum_{i=1}^n \mathbf{x}_i = \mathbf{0}, \sum_{i=1}^n y_i = 0$. Let $w_0^\star, \mathbf{w}^\star$ be the global minimum of the least squares objective

$$f(w_0, \mathbf{w}) = \sum_{i=1}^n (w_0 + \mathbf{w}^\top \mathbf{x}_i - y_i)^2.$$

Prove that $w_0^\star = 0$. Also, suppose $\mathbf{x}_i'$ and y_i' are such that for all i, $\mathbf{x}_i' = \mathbf{x}_i + \mathbf{q}$, $y_i' = y_i + r$. Show that $(w_0, \mathbf{w})$ minimizes f if and only if $(w_0 - \mathbf{w}^\top \mathbf{q} + r, \mathbf{w})$ minimizes

$$f'(w_o, \mathbf{w}) = \sum_{i=1}^{n} (w_0 + \mathbf{w}^\top \mathbf{x}_i' - y_i')^2.$$

Exercise 7. Logistic regression (optional) *(Exercise **??** from lecture note)*
Consider the logistic regression problem with two classes. Given a training set P consisting of datapoint and label pairs $(\mathbf{x}, y)$ where $\mathbf{x} \in \mathbb{R}^d$ and $y \in \{-1, +1\}$, we define our loss ℓ for weight vector $\mathbf{w} \in \mathbb{R}^d$ to be

$$\ell(\mathbf{w}) = \sum_{(\mathbf{x},y)\in P} -\ln\left(z(y\mathbf{w}^\top \mathbf{x})\right) ,$$

where $z(s) = 1/(1 + \exp(-s))$. We say that the weight vector $\mathbf{w}$ is a separator *for P if for all $(\mathbf{x}, y) \in P$,*

$$y(\mathbf{w}^\top \mathbf{x}) \geq 0 .$$

A separator is said to be trivial *if for all $(\mathbf{x}, y) \in P$,*

$$y(\mathbf{w}^\top \mathbf{x}) = 0 .$$

For example $\mathbf{w} = 0$ is a trivial separator. Depending on the data P, there may be other trivial separators.

Prove the following statement: the function ℓ has a global minimum if and only if all separators are trivial.

Getting Started with Python

Seeing the algorithms in action is more fun! Unfortunately, I do not have the capacity to provide support for programming exercises this year. Nevertheless, you are are welcome to explore the Python notebooks provided by the EPFL course.

Google Colab.

We recommend running the provided notebooks in the cloud using Google Colab. This way, you do not have to install anything, and you can even get a free GPU. (Of course, you can also setup up python locally on your machine).

The first practical exercise is a primer on NumPy, a scientific computing library for Python. You can open the corresponding notebook in Colab with this link:

colab.research.google.com/github/epfml/OptML_course/blob/master/labs/ex00/npprimer.ipynb

For computational efficiency, avoid `for`-loops in favor of NumPy's built-in commands. These commands are vectorized and thoroughly optimized and bring the performance of numerical Python code (like for Matlab) closer to lower-level languages like C.