Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Department of Computer Science
Institute of Theoretical Computer Science
Bernd Gärtner and Niao He

Optimization for Data Science **Graded Assignment 1** **FS23**

- The solution is due on **March 28, 2023** by **11:59 pm**. Please submit your solution as a PDF on Moodle. The name of the file should follow the format `GA4-{Legi number}`, e.g., `GA1-19-123-456`. After uploading your solution, please make sure that the status is “Submitted for grading". You should receive an automatic email that confirms your submission. Please notify us if you don’t receive this.
- If you want to submit your solution within six hours before the deadline and a technical problem prevents you from submitting it on Moodle, you can send your solution as PDF to `saeed.ilchi@inf.ethz.ch`. The same submission deadline still applies. If you encounter any trouble with the submission process, complain timely.
- Please solve the exercises carefully and typeset your solution using LaTeX or a similar typesetting program. A tutorial can be found at `http://www.cadmo.ethz.ch/education/thesis/latex`. Handwritten solutions will not be graded! The same applies to solutions written with any kind of tablet device and stylus, etc.
- For geometric drawings that can easily be integrated into LaTeX documents, we recommend the drawing editor IPE, which you can find at `http://ipe7.sourceforge.net/`.
- Keep in mind the following premises:
  - When writing in English, write short and simple sentences.
  - When writing a proof, write precise statements.
- This is a theory course, which means: if an exercise does not explicitly say “you do not need to prove your answer” or “justify intuitively”, then a formal proof is **always** required. You can of course refer in your solutions to the lecture notes and to the exercises, if a result you need has already been proved there.
- We would like to stress that the ETH Disciplinary Code applies to this Graded Assignment as it constitutes part of your final grade. The only exception we make to the Code is that we encourage you to verbally discuss the tasks with your colleagues. It is strictly prohibited to share any (hand-)written or electronic (partial) solutions with any of your colleagues. We are obliged to inform the Rector of any violations of the Code.
- As with all exercises, the material of the graded assignments is relevant for the exam.

Separating Points on the Unit Interval (20 points)

Consider a learning problem where the data source $\mathcal{X} = [0, 1]$ is the unit interval and each sample point $X \in \mathcal{X}$ is drawn uniformly from $\mathcal{X}$ and is labeled as zero if $X < p^*$ and labeled as 1 otherwise, where p^* is an unknown parameter. Suppose we want to model finding p^* with 0-1-loss and the class of hypotheses is $\mathcal{H} = [0, 1]$. Provide a function $f(\cdot, \cdot)$ such that for any $0 \leqslant \varepsilon, \delta \leqslant 1$ and given $n \geqslant f(\varepsilon, \delta)$ many samples, any hypothesis $H \in \mathcal{H}$ with zero empirical risk[1] has low expected risk with probability at least $1 - \delta$. That is:

$$\ell(H) \leqslant \varepsilon.$$

In other words, if $n \geqslant f(\varepsilon, \delta)$, then the probability of existence of a hypothesis with zero empirical risk but with expected risk more than ε is at most δ.

Continuous Convex Functions (35 points)

Let $f : \mathbb{R}^d \to \mathbb{R}$ be a continuous function. Show that the following are equivalent:

(a) f is a convex function.

(b) For all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, the following inequality holds:

$$\int_0^1 f(\mathbf{y} + \lambda(\mathbf{x} - \mathbf{y}))d\lambda \leqslant \frac{f(\mathbf{x}) + f(\mathbf{y})}{2}.$$

Gradient Descent with Inexact Gradient Oracle (45 points)

Consider an unconstrained optimization problem $\min_{\mathbf{x} \in \mathbb{R}^d} f(\mathbf{x})$. Assume f is μ-strongly convex and L-Lipschitz smooth. Now we only have access to an inexact gradient $g(\mathbf{x})$ at each point $\mathbf{x}$ such that $||g(\mathbf{x}) - \nabla f(\mathbf{x})|| \leqslant \delta$ with $\delta > 0$. Consider gradient descent with this inexact gradient:

$$\mathbf{x}_{t+1} = \mathbf{x}_t - \gamma g(\mathbf{x}_t),$$

where $\gamma > 0$ is the step-size. Define $\mathbf{x}^* \triangleq \arg\min_{\mathbf{x} \in \mathbb{R}^d} f(x)$ and $f^* = f(\mathbf{x}^*)$.

(a) Show that for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, we have

$$f(\mathbf{x}) \geqslant f(\mathbf{y}) + \langle g(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle + \frac{\mu}{4}||\mathbf{x} - \mathbf{y}||^2 - \frac{\delta^2}{\mu},$$

and moreover,

$$\frac{1}{\mu}||g(\mathbf{y})||^2 \geqslant f(\mathbf{y}) - f^* - \frac{\delta^2}{\mu}.$$

[1] Observe that there is always at least one such hypothesis.

(b) Show that for any $\mathsf{x}, \mathsf{y} \in \mathbb{R}^d$ we have

$$f(\mathsf{x}) \leqslant f(\mathsf{y}) + \langle g(y), \mathsf{x} - \mathsf{y} \rangle + L\|\mathsf{x} - \mathsf{y}\|^2 + \frac{\delta^2}{2L}.$$

(c) Show that by running gradient descent with inexact gradient and setting $\gamma = \frac{1}{2L}$, we have

$$f(\mathsf{x}_{t+1}) - f^* \leqslant \left(1 - \frac{\mu}{4L}\right)(f(\mathsf{x}_t) - f^*) + \frac{3\delta^2}{4L}.$$

This directly implies

$$f(\mathsf{x}_T) - f^* \leqslant \left(1 - \frac{\mu}{4L}\right)^T (f(\mathsf{x}_0) - f^*) + \frac{3\delta^2}{\mu}.$$

(d) Find a function that is μ-strongly-convex and show that the algorithm above can not guarantee to find a point x such that $f(\mathsf{x}) - f^* < \frac{\delta^2}{2\mu}$.