CS769 (Jan 2023) - Optimization in Machine Learning
Assignment 1

1 Assignment Policies

The following are the policies regarding this assignment.

1. This assignment needs **to be done individually** by everyone. You are expected to work on the assignments on your own. **Plagiarism** leads to disciplinary action as per DPAC policies.
2. Use LaTeX (online platforms such as Overleaf) for writing theory questions. Handwritten submissions will not be accepted.
3. Submit the assignment as a zip file containing LaTeX source, and the PDF file.
4. Maximum Marks - **60 marks**
5. Assignment is due on Sunday, February 12th 23:59 hrs.

Questions

Question 1. Is $f(x)$ a convex function **True/False** ? Give reasons for your answers. (5 marks)

(a) $f(x) = (\det X)^{1/n}$ on $\mathbf{dom}\ f = \mathbf{S}^n_{++}$

(b) $f(x_1, x_2) = x_1/x_2$ on $\mathbf{R}^2_{++}$

(c) $f(x_1, x_2) = 1/(x_1 x_2)$ on $\mathbf{R}^2_{++}$

(d) $f : \mathbf{R}^n \to \mathbf{R}$, $f(x) = \max_{i=1,2,\ldots,k} \|A^{(i)}x - b^{(i)}\|$, where $A^{(i)} \in \mathbf{R}^{m\times n}, b^{(i)} \in \mathbf{R}^m$, and $\|\cdot\|$ is a norm on $\mathbf{R}^m$.

(e) Gaussian distribution function f

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du$$

Question 2. Consider f to be a convex function, $\lambda_1 > 0$, $\lambda_i \leq 0$ for $i = 2, \ldots, n$ and $\sum_i \lambda_i = 1$. Let $\mathbf{dom}(f)$ be affine, and for $x_1, \ldots, x_n \in \mathbf{dom}(f)$, show that the inequality always holds:

$$f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \geq \sum_{i=1}^{n} \lambda_i f(x_i)$$

(5 marks)

Question 3. We say that a function f is log-convex on the real interval $\mathcal{D} = [a, b]$ if $\forall\ x, y \in \mathcal{D}$ and $\lambda \in [0, 1]$, the function satisfies

$$f\big(\lambda x + (1-\lambda)y\big) \leq f^{\lambda}(x) f^{1-\lambda}(y)$$

We will show that for an increasing log-convex function $f : \mathcal{D} \to \mathbb{R}$ and $0 \leq t \leq 1$,

$$f\left(\frac{a+b}{2}\right) \leq \phi(a,b) \leq \frac{1}{b-a}\int_a^b f(x)dx \leq \psi(a,b,t) \leq \mathcal{L}(f(a), f(b)) \leq \frac{f(a)+f(b)}{2}$$

where

$$\begin{aligned}
\phi(a,b) &= \sqrt{f\left(\frac{3a+b}{4}\right) f\left(\frac{a+3b}{4}\right)} \\
\mathcal{L}(a,b) &= \frac{a-b}{\ln(\frac{a}{b})} \\
\psi(a,b,t) &= (1-t)\mathcal{L}\Big(f\big(ta+(1-t)b\big), f\big(a\big)\Big) + t\mathcal{L}\Big(f\big(b\big), f\big(ta+(1-t)b\big)\Big)
\end{aligned}$$

(a) First, we prove the following inequalities -

(i) For $0 < t < 1$, the following holds

$$t^t(1-t)^{1-t} \geq \frac{1}{2}$$

(2 marks).

(ii) For $0 < a < b$ and $0 \leq t \leq 1$, the following holds

$$\sqrt{ab} \geq \begin{cases} a^{1-t}b^t & t \leq \frac{1}{2} \\ a^t b^{1-t} & t > \frac{1}{2} \end{cases}$$

and

$$\sqrt{ab} \leq \frac{a^{1-t}b^t + a^t b^{1-t}}{2} \leq \frac{a+b}{2}$$

(3 marks)

(b) Show that if f is a positive log-convex function on $[a, b]$, then

$$f\left(\frac{a+b}{2}\right) \leq \frac{1}{b-a}\int_a^b f(x)dx \leq \frac{f(a)+f(b)}{2}$$

(4 marks)

(c) Finally, show that $\sqrt{ab} \leq \mathcal{L}(a,b) \leq \frac{a+b}{2}$ and prove the required inequality. (5 marks)

Question 4. Let $\mathcal{D} = [a, b]$ and $f : \mathcal{D} \to \mathbb{R}$ be a convex or concave $\mathcal{C}^2$ class function. Show that if $|f'(x)| \geq \zeta$ for all $x \in \mathcal{D}$ and $\zeta > 0$, then

$$\left|\int_a^b e^{\iota f(x)}dx\right| \leq \frac{2}{\zeta}$$

where $\iota = \sqrt{-1}$. (5 marks)

Question 5. The basic idea behind many reinforcement learning algorithms is to estimate the action-value function $Q^*(s,a)$ by using the Bellman equation as an iterative update,

$$Q_{i+1}(s,a) = \mathbb{E}_{s'}[r + \gamma \max_{a'} Q_i(s',a')|s,a]$$

where $\{a\}$ are the actions, $\{s\}$ are the states, r is the reward and γ is a discounting factor. In practice, such iterative methods converge to the optimal value function as $i \to \infty$. [If you're not familiar with Reinforcement Learning, read this short introduction to understand the terminologies used: Reinforcement Learning, although it is not required to solve the question.]

It is seen that, this is infeasible and a neural network $Q(s,a,\theta)$ is used as an approximator to estimate this optimal action-value function as $Q(s,a;\theta) \approx Q^*(s,a)$. During training, we minimize the mean-squared error in the Bellman equation, and the loss function of such a network is given as

$$L_i(\theta_i) = \mathbb{E}_{(s,a,r,s') \sim U(D)}[(r + \gamma \max_{a'} Q(s',a';\theta_i^-) - Q(s,a;\theta_i))^2]$$

where $\mathbf{e} = (s,a,r,s')$ are the experiences forming the dataset D. It is known that θ_i^- is fixed.
Find the gradient of the above loss function w.r.t θ_i. (3 marks)

Question 6. Let $x_1, \ldots, x_n$ be non-negative points, and $p_1, \ldots, p_n$ be positive numbers such that $\sum_i p_i = 1$. Define a non-decreasing convex function $f : \text{conv}\{x_1, \ldots, x_n\} \to \mathbb{R}$. Then show that

(a)

$$\prod_{i=1}^{n} x_i^{p_i} \leq \sum_{i=1}^{n} p_i x_i \leq \sum_{i=1}^{n} x_i - (n-1) \prod_{i=1}^{n} x_i^{\frac{1-p_i}{n-1}}$$

(4 marks)

(b)

$$f\left(\prod_{i=1}^{n} x_i^{p_i}\right) \leq \sum_{i=1}^{n} p_i f(x_i) \leq \sum_{i=1}^{n} f(x_i) - (n-1) f\left(\prod_{i=1}^{n} x_i^{\frac{1-p_i}{n-1}}\right)$$

(4 marks)

Question 7.

(a) Show that the following definitions are equivalent:
A function f is L-smooth with Lipschitz constant $L > 0$, if

- $\forall\, \mathbf{x}, \boldsymbol{y} \in \text{dom}(f)$, $\|\nabla f(\mathbf{x}) - \nabla f(\boldsymbol{y})\| \leq L\|\mathbf{x} - \boldsymbol{y}\|$ (i.e, ∇f is L-Lipschitz continuous)
- a quadratic function upper bounds f, i.e, $|f(\boldsymbol{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \boldsymbol{y} - \mathbf{x}\rangle| \leq \frac{L}{2}\|\mathbf{x} - \boldsymbol{y}\|_2^2$

[Hint: Try to express $f(\boldsymbol{y}) - f(\mathbf{x})$ as an integral.] (4 marks)

(b) Let $f : \mathbb{R}^n \to \mathbb{R}$ be such that:

- f is a convex function
- ∇f is Lipschitz-continuous with Lipschitz constant 2μ

Show that, for all $\mathbf{x}, \boldsymbol{y} \in \mathbb{R}^n$,

$$\frac{1}{\mu}\|\nabla f(\mathbf{x}) - \nabla f(\boldsymbol{y})\|_2^2 \leq |f(\boldsymbol{y}) - f(\mathbf{x}) - \nabla f(\mathbf{x})^\top(\boldsymbol{y} - \mathbf{x})| \leq \mu\|\boldsymbol{y} - \mathbf{x}\|^2$$

What can you comment about f in this case? (4 marks)

Question 8. Implement Numerically correct versions of the following functions:

1) Logistic Loss: $L(w) = \sum_{i=1}^{n} \log\left(1 + \exp\left(-y_i w^T x_i\right)\right)$
2) Hinge Loss/SVMs: $L(w) = \sum_{i=1}^{n} \max\left\{0, 1 - y_i w^T x_i\right\}$. Here $y_i \in \{-1, +1\}$.
3) Least Squares Loss: $L(w) = \sum_{i=1}^{n} \left(y_i - w^T x_i\right)^2$. Here $y_i \in R$.

Note: Write your codes in the given notebook: **Assignment1.ipynb** with your implementations of 1), 2), 3), respectively. Do not modify the arguments.

1. Implement the Following Loss Functions with a simple **using simple loop code** in Python. (3 marks)
2. Implement these functions using vectorized code and compare the result with the previous simple loop code. Also, Implement these functions in CVXPY. (6 marks)
3. Plot the graph based on errors of the following functions mentioned above. (3 marks)