Homework 6

This homework is due at 11 PM on Wednesday, October 11, 2023.

Submission Format: Your homework submission should consist of a single PDF file that contains all of your answers (any handwritten answers should be scanned), as well as a printout of your completed Jupyter notebook(s).

1. **Convex or Concave**

   Determine whether the following functions are convex, strictly convex, concave, strictly concave, both or neither.

   (a) $f(x) = \mathrm{e}^x - 1$ on $\mathbb{R}$.

   (b) $f(x_1, x_2) = x_1 x_2$ on $\mathbb{R}^2_{++}$ (i.e. when $x_1 > 0$ and $x_2 > 0$).

   (c) The log-likelihood of a set of points $\{x_1, \ldots, x_n\}$ that are normally distributed with mean μ and finite variance $\sigma > 0$ is given by:

   $$f(\mu, \sigma) = n \log\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 \tag{1}$$

   i. Show that if we view the log likelihood for fixed σ as a function of the mean, i.e

   $$g(\mu) = n \log\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 \tag{2}$$

   then g is strictly concave (equivalently, we say f is strictly concave in μ).

   ii. **(OPTIONAL)** Show that if we view the log likelihood for fixed μ as a function of the inverse of the variance, i.e

   $$h(z) = n \log\left(\frac{\sqrt{z}}{\sqrt{2\pi}}\right) - \frac{z}{2}\sum_{i=1}^{n}(x_i - \mu)^2 \tag{3}$$

   then h is strictly concave (equivalently, we say f is strictly concave in $z = \frac{1}{\sigma^2}$). Note that we have used the dummy variable z to denote $\frac{1}{\sigma^2}$.

   iii. **(OPTIONAL)** Show that f is not jointly concave in μ, $\frac{1}{\sigma^2}$. *HINT: We say a function $w(x, y)$ with $x \in \mathcal{R}^m$ and $y \in \mathcal{R}^n$ is jointly convex if*

   $$w\left(\lambda(x_1, y_1) + (1-\lambda)(x_2, y_2)\right) \leq \lambda w((x_1, y_1)) + (1-\lambda) w((x_2, y_2)). \tag{4}$$

   This is the same as letting $z = (x, y)$ and saying f is convex in z. We can define joint concavity in a similar fashion by reversing the inequalities.

   (d) Prove that $f(x) = \log(1 + \mathrm{e}^x)$ is convex. Note that this implies that $g(x) = -f(x) = \log\left(\frac{1}{1+\mathrm{e}^x}\right)$ is concave. Compare this to $h(x) = \frac{1}{1+\mathrm{e}^x}$, is $h(x)$ convex or concave?

2. Further characterizations of convexity

Show that $\sigma_1 : \mathbb{R}^{m \times n} \to \mathbb{R}_+$, the function that maps a matrix to its largest singular value, is a convex function, with domain $\mathbb{R}^{m \times n}$.

HINT: You may express $\sigma_1(A)$ using the ℓ^2 operator norm of A:

$$\sigma_1(A) = \max_{\vec{x} \in \mathbb{R}^n : \|\vec{x}\|_2 = 1} \|A\vec{x}\|_2 \,,$$

This question proves that this norm is convex, so you may not use the fact that norms are convex.

3. Convex and strictly convex functions

(a) Recall that a function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be strictly convex if it satisfies Jensen's inequality with strict inequality, i.e., $\forall \vec{x} \neq \vec{y} \in \mathbb{R}^n$ and $\forall t \in (0, 1)$, we have

$$f(t\vec{x} + (1-t)\vec{y}) < tf(\vec{x}) + (1-t)f(\vec{y})$$

Show that for a strictly convex function $f : \mathbb{R}^n \to \mathbb{R}$, the problem

$$\min_{\vec{x} \in \mathbb{R}^n} f(\vec{x}) \tag{5}$$

has at most one solution.

HINT: Try to argue by contradiction assuming that there are two solutions $\vec{x}_1, \vec{x}_2$ which achieve the minimum value. Argue that using these two points you can find another point in $\mathbb{R}^n$ with strictly smaller function value.

(b) Prove that for all convex optimization problems $\min_{\vec{x} \in \mathcal{X}} f(\vec{x})$, where f is a convex function and $\mathcal{X}$ is a convex subset of its domain, all local minima are global minima. You may not assume that f is differentiable.

HINT: Start with assuming $\vec{x}^\star$ is a local minimum that is not global. Then there must exist some $\vec{\tilde{x}}$ satisfying $f(\vec{\tilde{x}}) < f(\vec{x}^\star)$. Use the definition of the convexity of a function to prove by contradiction.

4. First Order Criteria for Convexity, Strict Convexity, and Strong Convexity

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function with domain $\text{dom}(f)$. Note that requiring that f is differentiable automatically implies that we are assuming that $\text{dom}(f)$ is an open set.

(a) Show that f is convex iff it holds that $\text{dom}(f)$ is a convex set and for all $\vec{x}, \vec{y} \in \text{dom}(f)$ we have

$$\left(\nabla f(\vec{y}) - \nabla f(\vec{x})\right)^T (\vec{y} - \vec{x}) \geq 0. \tag{6}$$

Remark: When a function $g : \mathbb{R}^n \to \mathbb{R}^n$ satisfies the condition $\left(g(\vec{y}) - g(\vec{x})\right)^T (\vec{y} - \vec{x}) \geq 0$ for all $\vec{x}, \vec{y} \in \text{dom}(g)$, we say that that g is *monotone*. Note that this is consistent with the use of the term "monotone" to refer to a function $g : \mathbb{R} \to \mathbb{R}$ that is monotonically increasing (although one often uses the term in this case to also apply to a function $g : \mathbb{R} \to \mathbb{R}$ that is monotonically decreasing). Thus the condition in (6) is saying that ∇f is monotone.

(b) Recall that a function $f : \mathbb{R}^n \to \mathbb{R}$ with domain $\text{dom}(f)$ is said to be strictly convex if $\text{dom}(f)$ is a convex set and for all $\vec{x} \neq \vec{y} \in \text{dom}(f)$ and $\lambda \in (0, 1)$ we have

$$f(\lambda\vec{x} + (1-\lambda)\vec{y}) < \lambda f(\vec{x}) + (1-\lambda) f(\vec{y}).$$

Show that f is strictly convex iff it holds that $\text{dom}(f)$ is a convex set and for all $\vec{x} \neq \vec{y} \in \text{dom}(f)$ we have

$$\left(\nabla f(\vec{y}) - \nabla f(\vec{x})\right)^T (\vec{y} - \vec{x}) > 0. \tag{7}$$

Remark: When a function $g : \mathbb{R}^n \to \mathbb{R}^n$ satisfies the condition $\left(g(\vec{y}) - g(\vec{x})\right)^T (\vec{y} - \vec{x}) > 0$ for all $\vec{x} \neq \vec{y} \in \text{dom}(g)$ we say that that g is *strictly monotone*.

(c) Let $m > 0$. A function $f : \mathbb{R}^n \to \mathbb{R}$ is called m-*strongly convex* if the function

$$h(\vec{x}) := f(\vec{x}) - \frac{m}{2}\|\vec{x}\|_2^2,$$

with $\text{dom}(h) := \text{dom}(f)$, is convex. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function with domain $\text{dom}(f)$ (note that this means $\text{dom}(f)$ must be an open set). Given $m > 0$, show that f is m-strongly convex iff it holds that $\text{dom}(f)$ is a convex set and for all $\vec{x}, \vec{y} \in \text{dom}(f)$ we have

$$\left(\nabla f(\vec{y}) - \nabla f(\vec{x})\right)^T (\vec{y} - \vec{x}) \geq m\|\vec{x} - \vec{y}\|_2^2. \tag{8}$$

Remark: When a function $g : \mathbb{R}^n \to \mathbb{R}^n$ satisfies the condition $\left(g(\vec{y}) - g(\vec{x})\right)^T (\vec{y} - \vec{x}) > m\|\vec{x} - \vec{y}\|^2$ for all $\vec{x}, \vec{y} \in \text{dom}(g)$ we say that that g is *strongly monotone* or *coercive* (confusingly, the term "coercive" is also used in a different sense, which we will encounter later). Thus the condition in (8) is saying that ∇f is strongly monotone.

5. Homework Process

With whom did you work on this homework? List the names and SIDs of your group members.

NOTE: If you didn't work with anyone, you can put "none" as your answer.