Exercises - convex analysis

Caio Corro

1 Convex functions

Are the following functions convex? concave? neither? (prove your answers)

1. Linear and affine functions: $f : \mathbb{R}^n \to \mathbb{R}$ defined as $f(\boldsymbol{u}) = \langle \boldsymbol{a}, \boldsymbol{u} \rangle + b$, where $\boldsymbol{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$.
2. Exponential function: $f : \mathbb{R} \to \mathbb{R}$ defined as $f(u) = \exp u$.
3. Logarithmic function: $f : \mathbb{R}_{++} \to \mathbb{R}$ defined as $f(u) = \log u$.
4. Sigmoid function: $f : \mathbb{R} \to]0, 1[$ defined as $f(u) = \frac{\exp(u)}{1+\exp(u)}$.
5. Softplus function: $f : \mathbb{R} \to \mathbb{R}$ defined as $f(u) = \log(1 + \exp u)$.
6. L2 regularizer: $f : \mathbb{R}^n \to \mathbb{R}$ defined as $f(\boldsymbol{u}) = \frac{1}{2}\|\boldsymbol{u}\|_2^2$.
7. L1 regularizer: $f : \mathbb{R}^n \to \mathbb{R}$ defined as $f(\boldsymbol{u}) = \sum_{i=1}^n |u_i|$.
8. Shannon entropy: $f : \triangle(k) \to \mathbb{R}_+$ defined as $f(\boldsymbol{u}) = -\sum_i u_i \log u_i$.
9. Fermi-Dirac entropy: $f : [0, 1] \to \mathbb{R}_+$ defined as $f(u) = -u \log u - (1 - u) \log(1 - u)$.

2 Inequalities and applications

1. Non-negativity of the KL divergence.
   (a) Prove that $\forall u \in \mathbb{R}_{++} : \log u \leq u - 1$.
   (b) Let $\boldsymbol{\mu}, \boldsymbol{\theta} \in \triangle(k)$. We assume that $\forall i : \mu_i > 0 \implies \theta_i > 0$. Prove that the KL divergence between $\boldsymbol{\mu}$ and $\boldsymbol{\theta}$, defined as

   $$D_{\text{KL}}[\boldsymbol{\mu}, \boldsymbol{\theta}] = \sum_i \mu_i \log \frac{\mu_i}{\theta_i}$$

   is non-negative, i.e. $D_{\text{KL}}[\boldsymbol{\mu}, \boldsymbol{\theta}] \geq 0$.
2. Convexity of the log-partition function via Jensen's inequality.
   (a) Let U be a convex set, $k \geq 2$, $\boldsymbol{u}^{(1)}, ..., \boldsymbol{u}^{(k)} \in U$ and $\boldsymbol{\mu} \in \triangle(k)$. Prove that

   $$\sum_{i=1}^k \mu_i \boldsymbol{u}^{(i)} \in U$$

   (hint: proof by induction).
   (b) (Jensen's inequality) Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be a convex function. Prove that for any $k \geq 2$, $\boldsymbol{u}^{(1)}, ..., \boldsymbol{u}^{(k)}$, $\boldsymbol{\mu} \in \triangle(k)$:

   $$\sum_{i=1}^k \mu_i f(\boldsymbol{u}^{(i)}) \geq f(\sum_{i=1}^k \mu_i \boldsymbol{u}^{(i)})$$

(c) Prove that the log-partition function, $f : \mathbb{R}^n \to \mathbb{R}$ defined as $f(\boldsymbol{u}) = \log \sum_{i=1}^n \exp(u_i)$ is convex using Jensen's inequality (hint: use the definition based on the Hessian of the function).

3. Convexity of the log-partition function via Hölder's inequality.

   (a) (Young's inequality) Let $u, v, p, q \in \mathbb{R}_{++}$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Prove that:

   $$uv \leq \frac{u^p}{p} + \frac{v^q}{q}.$$

   Hint:

   $$\begin{aligned} uv &= \exp(\log(uv)) \\ &= \exp(\log u + \log v) \\ &= \exp(\frac{1}{p} \log u^p + \frac{1}{q} \log v^q). \end{aligned}$$

   (b) (Hölder's inequality) Let $u_1...u_n, v_1...v_n \in \mathbb{R}_{++}$ and $p, q \in \mathbb{R}_{++}$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Prove that:

   $$\sum_i u_i v_i \leq (\sum_i u_i^p)^{\frac{1}{p}} (\sum_i v_i^q)^{\frac{1}{q}}$$

   (hint: use Young's inequality)

   (c) Prove that the log-partition function, $f : \mathbb{R}^n \to \mathbb{R}$ defined as $f(\boldsymbol{u}) = \log \sum_{i=1}^n \exp(u_i)$ is convex using Hölder's inequality.

3 Subgradients

You should prove the following examples twice: once using the definition of subgradient, and then a simpler proof using the "rules" we saw in the course.

1. (L1 reg) Let $f(\boldsymbol{a}) = \sum_i |a_i|$. Prove that $\boldsymbol{g}$ defined as follows is a subgradient of f at $\boldsymbol{a}$:

   $$g_i = \begin{cases} -1 & \text{if } a_i < 0 \\ 1 & \text{if } a_y > 0 \\ 0 & \text{otherwise} \end{cases}$$

   Hint: check each coordinate independently and them sum over inequalities.

2. (Hinge loss, binary) Let $f(w) = \max(0, 1-w)$. Prove that g defined as follows is a subgradient of f at w:

   $$g = \begin{cases} 0 & \text{if } w > 1 \\ -1 & \text{otherwise} \end{cases}$$

The hinge loss for multiclass classification $\ell : E(k) \times \mathbb{R}^k \to \mathbb{R}_{++}$ is defines as:

$$\ell(\boldsymbol{y}, \boldsymbol{w}) = \max(0, -\langle \boldsymbol{w}, \boldsymbol{y} \rangle + 1 + \max_{\boldsymbol{y}' \in E(k) \setminus \{\boldsymbol{y}\}} \langle \boldsymbol{w}, \boldsymbol{y}' \rangle)$$

1. Prove that the following formulation is equivalent:

   $$\bar{\ell}(\boldsymbol{y}, \boldsymbol{w}) = -\langle \boldsymbol{w}, \boldsymbol{y} \rangle + \max_{\boldsymbol{y}' \in E(k)} \langle \boldsymbol{w} + (\mathbf{1} - \boldsymbol{y}), \boldsymbol{y}' \rangle$$

   where $\mathbf{1}$ is a vector of dimension k full of ones. We call the inner maximization problem in this loss "loss-augmented inference".

2. Use this second formulation to compute one subgradient w.r.t. input $\boldsymbol{w}$.

4 Fenchel conjugates

Compute the Fenchel conjugates of the following functions:

1. Exponential function: $f : \mathbb{R} \to \mathbb{R}$ defined as $f(u) = \exp u$. Then compute explicitly the biconjuguate.
2. Negative logarithmic function: $f : \mathbb{R}_{++} \to \mathbb{R}$ defined as $f(u) = -\log u$. Then compute explicitly the biconjuguate.
3. Linear and affine functions: $f : \mathbb{R}^n \to \mathbb{R}$ defined as $f(\boldsymbol{u}) = \langle \boldsymbol{a}, \boldsymbol{u} \rangle + b$, where $\boldsymbol{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Then compute explicitly the biconjuguate.
4. L2 regularizer: $f : \mathbb{R}^n \to \mathbb{R}$ defined as $f(\boldsymbol{u}) = \frac{1}{2}\|\boldsymbol{u}\|_2^2$.
5. Softplus function: $f : \mathbb{R} \to \mathbb{R}$ defined as $f(u) = \log(1 + \exp u)$.
6. Log-partition function: $f : \mathbb{R}^n \to \mathbb{R}$ defined as $f(\boldsymbol{u}) = \log \sum_{i=1}^n \exp(u_i)$.

5 Fenchel conjugates properties

1. Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be a function defined as $f(\boldsymbol{u}) = \sum_{i=1}^n f_i(u_i)$. Prove that:

$$f^*(\boldsymbol{t}) = \sum_{i=1}^n f_i^*(t_i)$$

2. Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be a function defined as $f(\boldsymbol{u}) = \alpha h(\boldsymbol{u})$, $\alpha > 0$. Prove that:

$$f^*(\boldsymbol{t}) = \alpha h^*(\alpha^{-1}\boldsymbol{t})$$