Discussion 6

1. Convexity of Functions

Definition. A function $f\colon \mathbb{R}^n \to \mathbb{R}$ is convex if $\operatorname{dom}(f)$ is a convex set and if for all $\vec{x}, \vec{y} \in \operatorname{dom}(f)$ and $\theta \in [0,1]$, we have,

$$f(\theta\vec{x} + (1-\theta)\vec{y}) \leq \theta f(\vec{x}) + (1-\theta)f(\vec{y}). \tag{1}$$

The function f is strictly convex if the inequality is strict.

Definition. A function $f : \mathbb{R}^n \to \mathbb{R}$ is concave if $\operatorname{dom}(f)$ is a convex set and if for all $\vec{x}, \vec{y} \in \operatorname{dom}(f)$ and θ with $0 \leq \theta \leq 1$, we have,

$$f(\theta\vec{x} + (1-\theta)\vec{y}) \geq \theta f(\vec{x}) + (1-\theta)f(\vec{y}). \tag{2}$$

The function f is strictly concave if the inequality is strict.

Property. A function f is concave if and only if $-f$ is convex. An affine function is both convex and concave.

Property: Jensen's inequality. The inequality in Equation (1) is known as **Jensen's Inequality**. This can be extended to convex combinations of more than one point. If f is convex, and $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k \in \operatorname{dom}(f)$, and $\theta_1, \theta_2, \ldots, \theta_k \geq 0$ with $\sum_{i=1}^k \theta_i = 1$ then,

$$f(\theta_1\vec{x}_1 + \theta_2\vec{x}_2 + \cdots + \theta_k\vec{x}_k) \leq \theta_1 f(\vec{x}_1) + \theta_2 f(\vec{x}_2) + \cdots + \theta_k f(\vec{x}_k). \tag{3}$$

Property: first order condition. Suppose f is differentiable. Then f is convex if and only if $\operatorname{dom}(f)$ is convex and

$$f(\vec{y}) \geq f(\vec{x}) + \nabla f(\vec{x})^\top(\vec{y} - \vec{x}), \tag{4}$$

for all $\vec{x}, \vec{y} \in \operatorname{dom}(f)$.

Property: Second order condition. Suppose f is twice differentiable. Then f is convex if and only if, $\operatorname{dom}(f)$ is convex and the Hessian of f, $\nabla^2 f(\vec{x})$, is positive semi-definite for all $\vec{x} \in \operatorname{dom}(f)$.

(a) **Restriction to a line**.

Show that a function f is convex if and only if for all $\vec{x} \in \operatorname{dom}(f)$ and all $\vec{v}$, the function $g : \operatorname{dom}(g) \to \mathbb{R}$ given by $g(t) = f(\vec{x} + t\vec{v})$ is convex for $\operatorname{dom}(g) = \{t \in \mathbb{R} \mid \vec{x} + t\vec{v} \in \operatorname{dom}(f)\}$.

Solution: In the first direction: assume f is convex and consider $\vec{x} \in \operatorname{dom}(f)$, $\vec{v}$ and the function $g : \operatorname{dom}(g) \to \mathbb{R}$ given by $g(t) = f(\vec{x} + t\vec{v})$ where $\operatorname{dom}(g) = \{t \in \mathbb{R} \mid \vec{x} + t\vec{v} \in \operatorname{dom}(f)\}$.

Because f is convex, $\operatorname{dom}(f)$ is convex, therefore $\operatorname{dom}(g)$ is also convex. For $t_1, t_2 \in \operatorname{dom}(g)$ and $\lambda \in [0,1]$:

$$\begin{aligned} g(\lambda t_1 + (1-\lambda)t_2) &= f(\vec{x} + (\lambda t_1 + (1-\lambda)t_2)\vec{v}) && (5) \\ &= f(\lambda(\vec{x} + t_1\vec{v}) + (1-\lambda)(\vec{x} + t_2\vec{v})) && (6) \\ &\leq \lambda f(\vec{x} + t_1\vec{v}) + (1-\lambda)f(\vec{x} + t_2\vec{v}) && (7) \\ &= \lambda g(t_1) + (1-\lambda)g(t_2) && (8) \end{aligned}$$

Therefore g is convex.

In the other direction: Consider $\vec{x}_1, \vec{x}_2 \in \text{dom}(f)$ and $\lambda \in [0,1]$. Define $g : t \to f(\vec{x}_2 + t(\vec{x}_1 - \vec{x}_2))$. g is convex and $0 \in \text{dom}(g)$ and $1 \in \text{dom}(g)$, so $[0,1] \in \text{dom}(g)$. Therefore $\lambda\vec{x}_1 + (1-\lambda)\vec{x}_2 \in \text{dom}(f)$ and $\text{dom}(f)$ is convex.

Because g is convex:

$$g(\lambda 1 + (1-\lambda)0) = g(\lambda) \leq \lambda g(1) + (1-\lambda)g(0) \tag{9}$$

$$f(\vec{x}_2 + \lambda(\vec{x}_1 - \vec{x}_2)) \leq \lambda f(\vec{x}_2 + 1(\vec{x}_1 - \vec{x}_2)) + (1-\lambda)f(\vec{x}_2 + 0(\vec{x}_1 - \vec{x}_2)) \tag{10}$$

$$f(\lambda\vec{x}_1 + (1-\lambda)\vec{x}_2) \leq \lambda f(\vec{x}_1) + (1-\lambda)f(\vec{x}_2) \tag{11}$$

Therefore f is convex.

(b) **Non-negative weighted sum**.

Show that the non-negative weighted sum of convex functions is convex: i.e. if $f_1, \ldots, f_n$ are n convex functions from $\mathbb{R}^n$ to $\mathbb{R}$ and $w_1, \ldots, w_n \in \mathbb{R}_+$ are n positive scalars, then the function:

$$f = \sum_{i=1}^{n} w_i f_i \tag{12}$$

is convex. To make the question easier, you can assume that the functions $f_1, \ldots, f_n$ are twice-differentiable.

Solution: Check convexity by using the second order condition. First, the weighted sum of twice-differentiable function is also twice-differentiable:

$$\nabla^2 f = \nabla^2 \left(\sum_{i=1}^{n} w_i f_i\right) \tag{13}$$

$$= \sum_{i=1}^{n} w_i \nabla^2 f_i \qquad \text{(linearity of } \nabla^2\text{)} \tag{14}$$

Next we check that $\nabla^2 f$ is PSD.

$$\forall \vec{y}, \forall \vec{x} \quad \vec{y}^\top(\nabla^2 f(\vec{x}))\vec{y} = \vec{y}^\top(\sum_{i=1}^{n} w_i \nabla^2 f_i(\vec{x}))\vec{y} \tag{15}$$

$$= \sum_{i=1}^{n} w_i \vec{y}^\top(\nabla^2 f_i(\vec{x}))\vec{y} \tag{16}$$

$$\geq 0 \qquad (\vec{y}^\top(\nabla^2 f_i(\vec{x}))\vec{y} \geq 0, \text{ because } f_i \text{ is convex}) \tag{17}$$

So $\forall \vec{x},\ \nabla^2 f(\vec{x})$ is PSD, so f is convex.

(c) **Point-wise maximum**.

Show that if f_1 and f_2 are convex functions then their pointwise maximum f, defined by

$$f(\vec{x}) = \max(f_1(\vec{x}), f_2(\vec{x})), \tag{18}$$

with $\text{dom}(f) = \text{dom}(f_1) \cap \text{dom}(f_2)$, is also convex.

Solution: Because f_1 and f_2 are convex, then $\text{dom}(f_1)$ and $\text{dom}(f_2)$ are convex sets. Because convexity of sets is preserved under intersection, $\text{dom}(f) = \text{dom}(f_1) \cap \text{dom}(f_2)$ is also convex.

$$\text{epi}(f) = \{(\vec{x}, t) \mid \vec{x} \in \text{dom}(f), f(\vec{x}) \leq t\} \tag{19}$$

$$= \{(\vec{x}, t) \mid \vec{x} \in \mathrm{dom}(f), \max(f_1(\vec{x}), f_2(\vec{x})) \le t\} \tag{20}$$
$$= \{(\vec{x}, t) \mid \vec{x} \in \mathrm{dom}(f_1) \cap \mathrm{dom}(f_2), f_1(\vec{x}) \le t \text{ and } f_2(\vec{x}) \le t\} \tag{21}$$
$$= \{(\vec{x}, t) \mid \vec{x} \in \mathrm{dom}(f_1), f_1(\vec{x}) \le t\} \cap \{(\vec{x}, t) \mid \vec{x} \in \mathrm{dom}(f_2), f_2(\vec{x}) \le t\} \tag{22}$$
$$= \mathrm{epi}(f_1) \cap \mathrm{epi}(f_2) \tag{23}$$

Because f_1 and f_2 are convex, then $\mathrm{epi}(f_1)$ and $\mathrm{epi}(f_2)$ are convex. Because convexity of sets is preserved under intersection, $\mathrm{epi}(f)$ is convex. Because of the equivalence between the convexity of functions and the convexity of their epigraphs, f is convex.

2. **Convexity of Constraint Sets**

Let $f_1, \ldots, f_m, h_1, \ldots, h_p \colon \mathbb{R}^n \to \mathbb{R}$ be functions. Let $S \subseteq \mathbb{R}^n$ be defined as

$$S \doteq \left\{ \vec{x} \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} f_i(\vec{x}) \le 0 & \forall i = 1, \ldots, m \\ h_j(\vec{x}) = 0 & \forall j = 1, \ldots, p \end{array} \right\}. \tag{24}$$

Show that if $f_1, \ldots, f_m$ are convex functions, and $h_1, \ldots, h_p$ are affine functions, then S is a convex set.

Solution: Let $\vec{x}, \vec{y} \in S$ and let $\theta \in [0, 1]$. Then for any $i = 1, \ldots, m$, we have

$$\begin{aligned} f_i(\theta\vec{x} + (1-\theta)\vec{y}) &\le \theta \underbrace{f_i(\vec{x})}_{\le 0} + (1-\theta) \underbrace{f_i(\vec{y})}_{\le 0} \\ &\le 0. \end{aligned}$$

And for any $j = 1, \ldots, p$, we have

$$\begin{aligned} h_j(\theta\vec{x} + (1-\theta)\vec{y}) &= \theta \underbrace{h_j(\vec{x})}_{=0} + (1-\theta) \underbrace{h(\vec{y})}_{=0} \\ &= 0. \end{aligned}$$

Thus $\theta\vec{x} + (1-\theta)\vec{y} \in S$. Thus S is convex.

3. **Ridge Regression**

Prove that the optimal solution to the ridge regression problem:

$$\min_{\vec{w} \in \mathbb{R}^p} \|X\vec{w} - \vec{y}\|_2^2 + \lambda \|\vec{w}\|_2^2, \tag{25}$$

where $X \in \mathbb{R}^{n \times p}$, $\lambda > 0$ and $\vec{y} \in \mathbb{R}^n$, is given by:

$$\vec{w}^* = (X^\top X + \lambda I)^{-1} X^\top \vec{y}. \tag{26}$$

Solution: We begin by taking the gradient of of the objective function

$$f(\vec{w}) = \|X\vec{w} - \vec{y}\|_2^2 + \lambda \|\vec{w}\|_2^2 = \vec{w}^\top X^\top X \vec{w} - 2\vec{y}^\top X \vec{w} + \|\vec{y}\|^2 + \lambda \|\vec{w}\|^2 \tag{27}$$

with respect to $\vec{w}$ and setting it to zero, we get:

$$\nabla f(\vec{w}^*) = 2X^\top X \vec{w}^* + 2\lambda \vec{w}^* - 2X^\top \vec{y} = 0 \implies \vec{w}^* = (X^\top X + \lambda I)^{-1} X^\top \vec{y}. \tag{28}$$