MATHS OF DATA SCIENCE

September 27, 2024

Today
Convex functions (Pt 2/2 feat. a board sliced in 'two')

Next Friday (Oct. 4)
<u>NO CLASS</u>

Friday Oct. 11
TWO LECTURES

Friday Oct 25
TWO LECTURES

Last session : Convex function

① Definition and first properties

$f : \mathbb{R}^d \to \mathbb{R}$ is convex on a convex set $\mathcal{X} \subseteq \mathbb{R}^d$

$\iff \forall (x,y) \in \mathcal{X}^2, \forall \alpha \in [0,1],$

$$f(\alpha x + (1-\alpha) y) \leq \alpha f(x) + (1-\alpha) f(y)$$

② Convexity and derivatives

$f : \mathbb{R}^d \to \mathbb{R}$ differentiable on an open set containing $\mathcal{X} \subseteq \mathbb{R}^d$ ($\mathcal{X}$ convex set)

$[f$ convex on $\mathcal{X}]$

$\iff \forall (x,y) \in \mathcal{X}^2, f(y) \geq f(x) + \nabla f(x)^T (y-x)$

↳ Convexity can also be characterized by second-order derivatives

For any twice differentiable function f, we can define the matrix of second-order derivatives, called the Hessian matrix

$$\underbrace{\nabla^2 f(x)}_{\in \mathbb{R}^{d \times d}} := \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right]_{\substack{i=1..d \\ j=1..d}} \quad \text{(in dimension } d)$$

- $\forall x \in \mathbb{R}^d$, $\nabla^2 f(x)$ (when it exists) is a symmetric matrix in $\mathbb{R}^{d \times d}$

$$\left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x) \right)$$

Theorem : Let $\mathcal{X} \subseteq \mathbb{R}^d$ be a convex set and let f be twice continuously differentiable on an open set containing $\mathcal{X}$.

$$[f \text{ is convex on } \mathcal{X}] \iff [\forall x \in \mathcal{X}, \nabla^2 f(x) \succeq 0]$$

$\Longleftrightarrow$ $\left[\forall x \in X, \forall v \in \mathbb{R}^d,\ \underbrace{v^T}_{1\times d} \underbrace{\nabla^2 f(x)}_{d\times d} \underbrace{v}_{d\times 1} \geq 0\right]$

$\Longleftrightarrow$ $\left[\forall x \in X,\ \underbrace{\lambda_{\min}(\nabla^2 f(x))}_{\text{minimum eigenvalue of } \nabla^2 f(x)} \geq 0\right]$

<u>Note</u>: Since $\nabla^2 f(x)$ is a real symmetric matrix, it can be written as $P \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{bmatrix} P^{-1}$ where $P \in \mathbb{R}^{d\times d}$ is invertible and $\lambda_1, \dots, \lambda_d$ are the d <u>real</u> eigenvalues of $\nabla^2 f(x)$

$(\forall i = 1..d, \exists v_i \in \mathbb{R}^d, v_i \neq 0_{\mathbb{R}^d}, \nabla^2 f(x) v_i = \lambda_i v_i)$

<u>Example</u>: $f: \mathbb{R}^d \to \mathbb{R}$
$x \mapsto \ln\left(\sum_{i=1}^d e^{x_i}\right)$ "log sum-exp" / "softmax"

$\forall x \in \mathbb{R}^d$,

$$\nabla^2 f(x) = \begin{bmatrix} \sum_{i\neq 1} e^{x_1+x_i} & -e^{x_1+x_2} & \cdots & -e^{x_1+x_d} \\ -e^{x_1+x_2} & \sum_{i\neq 2} e^{x_2+x_i} & & \vdots \\ \vdots & & \ddots & \\ -e^{x_1+x_d} & & & \sum_{i\neq d} e^{x_d+x_i} \end{bmatrix}$$

Sum first line $= 0$

$\forall i = 1..d$, $[\nabla^2 f(x)]_{ii} = \sum_{j\neq i} |[\nabla^2 f(x)]_{ij}|$

$\Rightarrow \nabla^2 f(x)$ is diagonally dominant hence $\succeq 0$

($A \in \mathbb{R}^{d\times d}$ is diagonally dominant if $\forall i = 1..d$, $[A]_{ii} \geq \sum_{j\neq i} |A_{ij}|$)

③ Extended-value functions

Last time: " Let $\{f_i\}_{i\in I}$ be a family of convex functions from $\mathbb{R}^d$ to $\mathbb{R}$, convex on $\mathbb{R}^d$. Then $\sup_{i\in I} f_i$ is convex on $\mathbb{R}^d$. "

Suppose that $f_i(x) = i \times x \quad \forall i \in \mathbb{N}$ ($d=1$)

but $\sup_{i \in \mathbb{N}} ix = \begin{cases} +\infty & \text{if } x > 0 \\ 0 & \text{if } x \le 0 \end{cases}$

$f(x)$ can be $\pm\infty$

Def: A function $f : \mathcal{X} \to \overline{\mathbb{R}}$ is called an extended-value

For such a function, the by $\text{dom}(f) = \{ x \in \mathcal{X}$

Ex) • Every real-valued function function.

• Given any set $\mathcal{X} \subseteq \mathbb{R}^d$, defined by $\delta_{\mathcal{X}}(x)$

is an extended-value function.

Def: Let $f : \mathcal{X} \subseteq \mathbb{R}^d \to \overline{\mathbb{R}}$. The never takes the value $-\infty$

and if $\text{dom}(f) \neq \emptyset$

→ $\forall \mathcal{X} \subseteq \mathbb{R}^d$, $\delta_{\mathcal{X}}$ (indicator

→ If the function is proper

Then every f_i is convex,

→ Q) Does our definition of convex functions apply in this case?

where $\mathcal{X} \subseteq \mathbb{R}^d$ and $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ function.

effective domain of f is defined

$\mid f(x) < +\infty \}$

$f(x) \in \mathbb{R}$ or $f(x) = -\infty$

$f : \mathbb{R}^d \to \mathbb{R}$ is an extended-value

the indicator function of $\mathcal{X}$

$= \begin{cases} 0 & \text{if } x \in \mathcal{X} \\ +\infty & \text{if } x \notin \mathcal{X} \end{cases}$

function f is called proper if it

(i.e. $f(x) > -\infty \quad \forall x \in \mathcal{X}$)

(i.e. $\exists \bar{x} \in \mathcal{X}, f(\bar{x}) \in \mathbb{R}$)

function) is proper when $\mathcal{X} \neq \emptyset$

, the classical characterization of

convexity applies

$$[f \text{ proper is convex on } \mathbb{R}^d] \iff \left[\forall (x,y)\in(\mathbb{R}^d)^2,\ \forall \alpha\in[0,1],\ f(\alpha x+(1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y)\right]$$

In particular, in the case of the indicator function, this characterization becomes

$$\left[\delta_{\mathcal{X}} \text{ is convex on } \mathbb{R}^d \text{ (as a proper function)}\right] \iff \left[\mathcal{X} \text{ is a convex set}\right]$$

$(\Rightarrow)$ $\forall (x,y)\in \mathcal{X}^2,\ \forall \alpha\in[0,1]$,

$$\delta_{\mathcal{X}}(\alpha x+(1-\alpha)y) \leq \alpha \overbrace{\delta_{\mathcal{X}}(x)}^{=0\ (x\in\mathcal{X})} + (1-\alpha)\overbrace{\delta_{\mathcal{X}}(y)}^{=0\ y\in\mathcal{X}}$$

$$\delta_{\mathcal{X}}(\alpha x+(1-\alpha)y) \leq 0$$

$$\Rightarrow \delta_{\mathcal{X}}(\alpha x+(1-\alpha)y) = 0$$

$$\iff \alpha x+(1-\alpha)y \in \mathcal{X}$$, hence $\mathcal{X}$ convex

$(\Leftarrow)$

x	y	$\alpha x+(1-\alpha)y$	$\delta_{\mathcal{X}}(\alpha x+(1-\alpha)y)$	$\alpha\delta_{\mathcal{X}}(x)+(1-\alpha)\delta_{\mathcal{X}}(y)$
$\in\mathcal{X}$	$\in\mathcal{X}$	$\in\mathcal{X}$	0	$\alpha\times 0+(1-\alpha)\times 0$
$\in\mathcal{X}$	$\notin\mathcal{X}$	$\in\mathcal{X}$	0	$\alpha\times 0+(1-\alpha)\times\infty=\infty$
$\notin\mathcal{X}$	$\notin\mathcal{X}$	$\notin\mathcal{X}$	∞	$\alpha\times\infty+(1-\alpha)\times\infty=\infty$
$\notin\mathcal{X}$	$\notin\mathcal{X}$	$\in\mathcal{X}$	0	∞

The inequality for all cases

(7 cases, the case $x\in\mathcal{X}$, $y\in\mathcal{X}$ and $\alpha x+(1-\alpha)y\notin\mathcal{X}$ does not occur)

↳ Not all extended-value functions are proper

↳ We would like a concept of convexity that extends to non-proper functions.

Def (Epigraph)

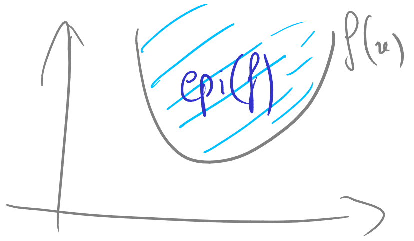
Let $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ be an extended-value function.

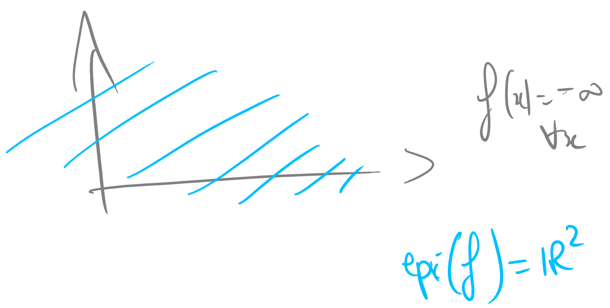
$\mathcal{X}\subseteq\mathbb{R}^d$

The epigraph of f is the set

$$epi(f) := \{ (x,y) \in \mathbb{R}^{d+1} \mid \overset{\in \mathbb{R}}{y} \geq f(\overset{x \in \mathbb{R}^d}{x}) \}$$

$x \in X$





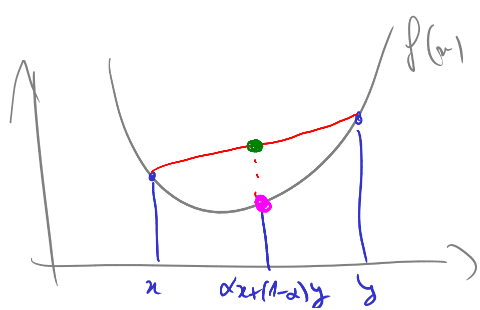
Definition (Convex extended-value function)

Let $f : X \subseteq \mathbb{R}^d \to \overline{\mathbb{R}}$ be an extended-value function defined on a convex set X.

Then f is called convex on X if $epi(f)$ is a convex set in $\mathbb{R}^{d+1}$

Remarks

- In the real-valued case (or in the proper case), this definition of convexity corresponds to the previous definition



f convex means that

$$f(\alpha x+(1-\alpha)y) \leq \alpha f(x)+(1-\alpha)f(y)$$

$$\Leftrightarrow (\alpha x+(1-\alpha)y, \alpha f(x)+(1-\alpha)f(y)) \in epi(f)$$

- If $f: X \to \bar{\mathbb{R}}$ but X is not a convex set in $\mathbb{R}^d$, $\mathrm{epi}(f) = \{(x,y) \in \mathbb{R}^{d+1} \mid x \in X \text{ and } y \geq f(x)\}$ cannot be a convex set

Note: If $f: \mathbb{R}^d \to \bar{\mathbb{R}}$ is differentiable or twice differentiable on an open set containing its domain $\mathrm{dom}(f)$, then convexity of f on can be characterized using the derivatives on $\mathrm{dom}(f)$.

Takeaways

→ Characterization of convexity with the Hessian matrix (when it exists)

→ Extended-value functions: convex when epigraph convex

→ Special cases: proper functions like indicator functions

⇒ All of this to prepare for the convex optimization lectures.

BONUS: Strongly convex functions / sets