MATHEMATICS OF DATA SCIENCE

September 20, 2024

Today (First in-person lecture!)

- Convex functions (Sec. 1.2 of the lecture notes)
  ⇒ Might finish this next time

- ⚠ Might be schedule changes on October 4

CONVEX FUNCTIONS

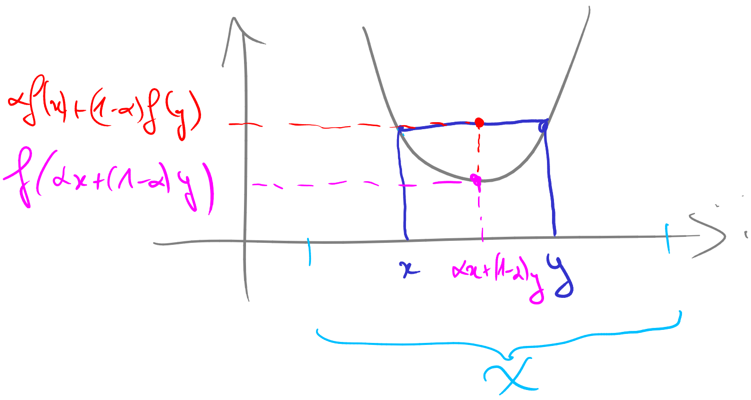
1) First definitions and properties

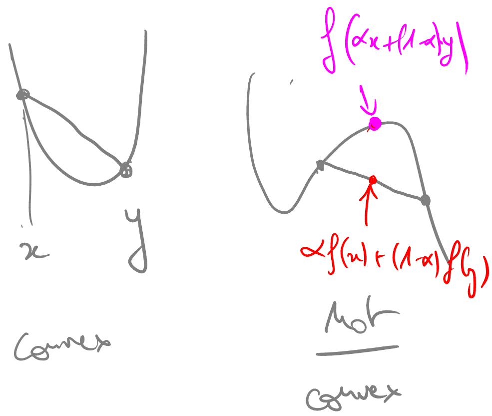
Def: Let $f : \mathbb{R}^d \to \mathbb{R}$ and let $\mathcal{X} \subseteq \mathbb{R}^d$ be a convex set.

- The function f is called convex on $\mathcal{X}$ if

$$\forall (x,y) \in \mathcal{X}^2, \forall \alpha \in [0,1], \quad f(\underbrace{\alpha x + (1-\alpha) y}_{\in \mathcal{X}}) \leq \alpha f(x) + (1-\alpha) f(y)$$

- The function f is called concave on $\mathcal{X}$ if $(-f)$ is convex on $\mathcal{X}$.





Examples:

- $\forall A \in \mathbb{R}^{m\times d}, \forall b \in \mathbb{R}^m, \; x \mapsto Ax+b$ is convex on $\mathbb{R}^d$ (NB: "convex on $\mathbb{R}^d$" $\equiv$ "convex")

$\Rightarrow$ that function is also concave!

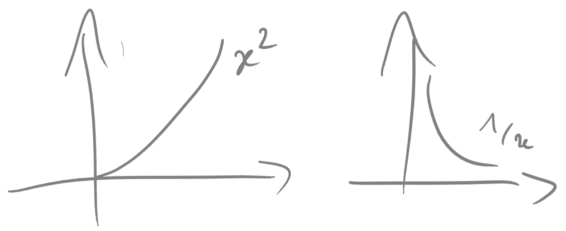
$\forall (x,y) \in (\mathbb{R}^d)^2, \forall \alpha \in [0,1],$

$$A(\alpha x + (1-\alpha)y) + b = \alpha A x + (1-\alpha) A y + b = \alpha (Ax+b) + (1-\alpha)(Ay+b)$$

$=$ is both $\leq$ and $\geq$

Linear functions are the only functions that are both concave and convex

- $\begin{array}{c} \mathbb{R} \to \mathbb{R} \\ x \mapsto x^a \end{array}$ is convex for $a > 1$ or $a < 0$
on $\mathbb{R}_{++} = \{x \in \mathbb{R} \mid x > 0\}$
$\mathbb{R}_{+} = \{x \in \mathbb{R} \mid x \geq 0\}$



- Any norm in $\mathbb{R}^d$ is convex on $\mathbb{R}^d$

(Euclidean / ℓ_2 norm) $\quad \|x\| := \sqrt{\sum_{i=1}^{d} x_i^2}$

$\|x\|_1 := \sum_{i=1}^{d} |x_i| \qquad \|x\|_\infty = \max_{1 \leq i \leq d} |x_i|$

...

(Follows from $N(x+y) \leq N(x) + N(y)$, $N(\lambda x) = \lambda N(x) \quad \forall \lambda \geq 0$) axioms for a norm $N(\cdot)$

- $x \mapsto \ln(x)$ (logarithm) is concave on $\mathbb{R}_{++}$
($x \mapsto -\ln(x)$ is convex)

- $x \mapsto e^x$ convex on $\mathbb{R}$

- $\begin{array}{c} \mathbb{R}^d \to \mathbb{R} \\ x \mapsto \ln\left(\sum_{i=1}^{d} e^{x_i}\right) \end{array}$ "log-sum exp" is convex on $\mathbb{R}^d$

- $x \mapsto \frac{1}{2}\|x\|^2$ is convex on $\mathbb{R}^d$.

...

$\hookrightarrow$ One way to check that a function is convex (other than checking the definition, which can be difficult when $d \geq 3$) is to decompose it into "building blocks" and to check that the decomposition only involves convex functions and operations that preserve convexity.

Conic combination of f_i
$x \longmapsto \sum_{i=1}^{k} \alpha_i f_i(x)$ with $\alpha_i \geq 0 \; \forall i = 1..k$

Operations that preserve convexity (Non-exhaustive list)

- If $f_1, \ldots, f_k$ are convex functions on $\mathbb{R}^d$ from $\mathbb{R}^d$ to $\mathbb{R}$, then any conic combination of the f_i's is convex on $\mathbb{R}^d$

- If $f : \mathbb{R}^m \to \mathbb{R}$ convex on $\mathbb{R}^m$ and $g : \mathbb{R}^d \to \mathbb{R}$ is defined by $g(x) = f(Ax+b)$ with $A \in \mathbb{R}^{m \times d}$, $b \in \mathbb{R}^m$ then g is convex on $\mathbb{R}^d$ (composition of convex and linear is convex)

- If $f_1, \ldots, f_k$ are convex functions (on $\mathbb{R}^d$) from $\mathbb{R}^d$ to $\mathbb{R}$ and if $g : \mathbb{R}^k \to \mathbb{R}$ is nondecreasing in each of its arguments, then $h : x \longmapsto g(f_1(x), f_2(x), \ldots, f_k(x))$ is convex.

NB: g need not be convex

Ex) Can use this to show convexity of $\ln\left(\sum e^{x_i}\right)$

$$g : \mathbb{R}^k \to \mathbb{R}$$
$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix} \mapsto g(y_1, \ldots, y_k)$$

g nondecreasing in y_1: $\forall (y_2, \ldots, y_k)$, $\forall (y_1, z_1)$,
$$y_1 \leq z_1 \Rightarrow g(y_1, y_2, \ldots, y_k) \leq g(z_1, y_2, \ldots, y_k)$$

- If f_1 and f_2 are convex functions on $\mathbb{R}^d$, $x \longmapsto \max(f_1(x), f_2(x))$ ($\mathbb{R}^d \to \mathbb{R}$) is convex

More generally, for any family (not necessarily finite) of functions $\{f_i\}_{i \in I}$ such that $f_i : \mathbb{R}^d \to \mathbb{R}$ is convex on $\mathbb{R}^d$,

$$x \longmapsto \sup_{i \in I} f_i(x) \quad \text{is convex}$$

NB: $\forall \{a_i\}_{i\in I} \in \mathbb{R}^I$,

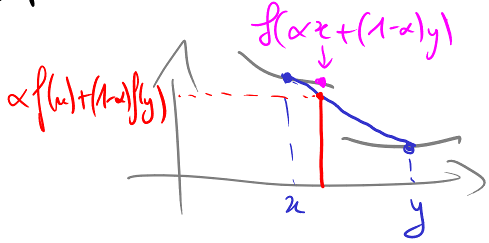
$\sup_{i \in I} \{a_i\}$ = smallest value v such that $v \geq a_i \ \forall i \in I$

Ex) $a_i = -\frac{1}{i+1}$ $I = \mathbb{N}$

$\sup_{i \in I} a_i = 0$

2) Convexity and derivatives

Theorem: Any function from $\mathbb{R}^d \to \mathbb{R}$ that is convex on $\mathbb{R}^d$ is continuous on $\mathbb{R}^d$.



Theorem: Let $f : \mathbb{R}^d \to \mathbb{R}$ be defined on some convex set $\mathcal{X} \subseteq \mathbb{R}^d$ and differentiable on an open set containing $\mathcal{X}$.

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

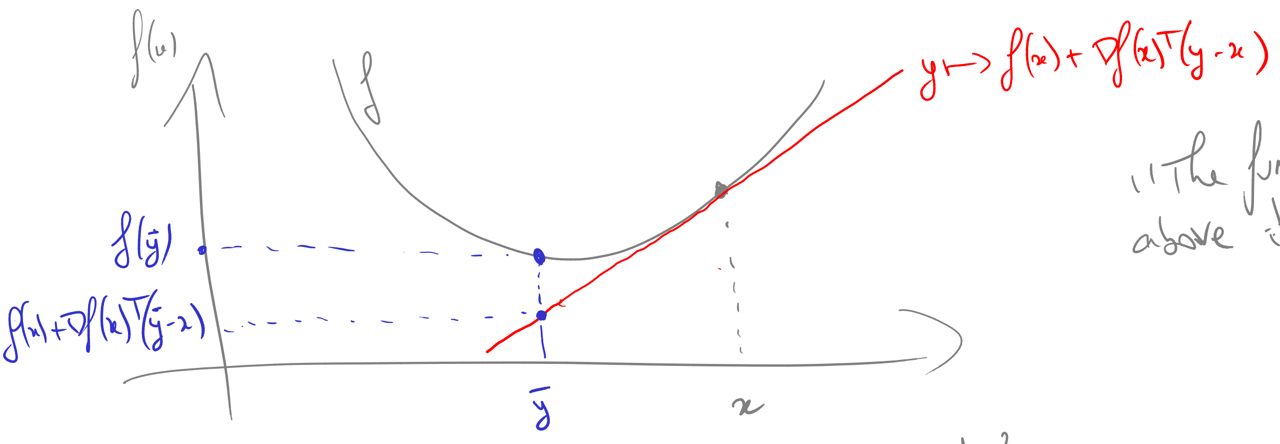
$$\iff \left[\forall (x,y) \in \mathcal{X}^2,\ f(y) \geq f(x) + \underbrace{\nabla f(x)}_{\in \mathbb{R}^d}{}^T \underbrace{(y-x)}_{\in \mathbb{R}^d}\right]$$

(First-order characterization of convexity)

$\nabla f(x)$ is the gradient of f at x

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_d}(x) \end{bmatrix} \in \mathbb{R}^d$$

Aparté : f is continuous on $\mathbb{R}^d$ if $\forall \varepsilon > 0$, $\exists \delta > 0$ such that $\|x - y\| \leq \delta \Rightarrow |f(x) - f(y)| \leq \varepsilon$



"The function stays above its tangents"

f convex $\quad \forall (x, \bar{y}) \in (\mathbb{R}^d)^2,$

$$f(\bar{y}) \geq f(x) + \nabla f(x)^T (\bar{y} - x) \qquad (*)$$

Remark : For differentiable functions, $(*)$ is another way to prove that a function is convex

Example : $q : \mathbb{R}^d \to \mathbb{R}$

$x \mapsto \frac{1}{2} x^T A x + b^T x + c$

quadratic function

$A \succeq 0$ (positive semidefinite matrix)

$b \in \mathbb{R}^d$

$c \in \mathbb{R}$

$A \in \mathbb{R}^{d \times d}$ is positive semi-definite (PSD) if

1) A is symmetric ($A^T = A$)

2) $\forall v \in \mathbb{R}^d$,

$$\underbrace{\underbrace{v^T}_{1 \times d} \underbrace{A}_{d \times d} \underbrace{v}_{d \times 1}}_{1 \times 1} \geq 0$$

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & & \\ a_{m1} & \cdots & a_{mm} \end{bmatrix}$$

$$A^T = \begin{bmatrix} a_{11} & & a_{m1} \\ & \ddots & \\ a_{1m} & & a_{mm} \end{bmatrix}$$

Ex) $A = \begin{bmatrix} 3 & 2 \\ 4 & 5 \end{bmatrix}$

$A^T = \begin{bmatrix} 3 & 4 \\ 2 & 5 \end{bmatrix}$

We write $A \succeq 0$ when A is PSD

q is convex on $\mathbb{R}^d$ $\quad q(x) = \frac{1}{2}x^TAx + b^Tx + c$

$$Dq(x) = Ax + b$$

$\forall (x,y) \in (\mathbb{R}^d)^2,$

$$q(y) = \frac{1}{2} y^TAy + b^Ty + c$$

$$= \frac{1}{2} y^TAy + b^Ty + \overbrace{\frac{1}{2}x^TAx + b^Tx + c}^{q(x)} - \frac{1}{2}x^TAx - b^Tx$$

$$= q(x) + \frac{1}{2} y^TAy - \frac{1}{2} x^TAx + b^Ty - b^Tx .$$

$$= q(x) + \frac{1}{2} y^TAy - \frac{1}{2} x^TAx + b^T(y-x)$$

$$= q(x) + (Ax)^T(y-x) + b^T(y-x) + \frac{1}{2} y^TAy - \frac{1}{2} x^TAx - (Ax)^T(y-x)$$

we want $q(y) \geq q(x) + Dq(x)^T(y-x)$

$$= q(x) + (Ax+b)^T(y-x) + \frac{1}{2} y^TAy - \frac{1}{2} x^TAx - \underbrace{x^TA(y-x)}$$

$(Ax)^T = x^TA^T$
$= x^TA$

$$q(y) = q(x) + Dq(x)^T(y-x) + \underbrace{\frac{1}{2} y^TAy - \frac{1}{2} x^TAx - x^TA(y-x)}_{\geq 0}$$

Since $\frac{1}{2} y^TAy - \frac{1}{2} x^TAx - x^TA(y-x) = \frac{1}{2} y^TAy - x^TAy + \frac{1}{2} x^TAx$

$$= \frac{1}{2} \underbrace{(y-x)^TA(y-x)} \geq 0$$

$v^TAv \geq 0$
because A PSD

$$\frac{1}{2}(y-x)^TA(y-x) = \frac{1}{2} y^TAy - \frac{1}{2} x^TAy - \frac{1}{2} y^TAx + \frac{1}{2} x^TAx$$

$$= \frac{1}{2} y^TAy - x^TAy + \frac{1}{2} x^TAx$$

by symmetry of A

↳ Note that not all quadratics are convex

Ex: $d = 1$, $q(x) = -x^2$