Convexity and Jensen's Inequality

Andrew Nobel

August, 2020

Euclidean Norm and Inner Product

Given: Vector $x = (x_1, \ldots, x_d)^t \in \mathbb{R}^d$

- Inner product $\langle x, y \rangle \; = \; x^t y \; = \; \sum_{i=1}^d x_i \, y_i$
- Norm $||x|| \; = \; \langle x, x \rangle^{1/2} \; = \; (x_1^2 + \cdots + x_d^2)^{1/2}$

Basic Properties: For every $x, y \in \mathbb{R}^d$

- Non-negative: $||x|| \geq 0$ with equality if and only if $x = 0$
- Scaling: For $a \in \mathbb{R}$, $||a\, x|| = |a| \, ||x||$
- Triangle inequality: $||x + y|| \; \leq \; ||x|| + ||y||$
- Reverse triangle inequality: $| \, ||x|| - ||y|| \, | \; \leq \; ||x - y||$
- Cauchy-Schwarz inequality: $|x^t y| \; \leq \; ||x|| \, ||y||$

Convex Sets

Definition: A set $C \subseteq \mathbb{R}^d$ is *convex* if for every $x, y \in C$ and every $\alpha \in [0,1]$ the point $\alpha x + (1-\alpha)y \in C$.

Interpretation

- Vector $\alpha x + (1-\alpha)y$ called *convex combination* of x, y with weight α
- Set $\{\alpha x + (1-\alpha)y : \alpha \in [0,1]\}$ is just the line between x and y
- So C is convex if the line between any two points in C is contained in C

Examples of Convex Sets

- Simple examples: $C = \mathbb{R}^d$, $\emptyset$, $\{0\}$
- Open norm ball $B(x_0, r) := \{x : ||x - x_0|| < r\}$
- Halfspace $H(w, b) = \{x : w^t x \geq b\}$ with direction w and offset b
- Hyperplane $\partial H(w, b) = \{x : w^t x = b\}$ (n-1)-dimensional
- Polyhedron $\{x : Ax \leq c\}$ where $\leq$ understood componentwise
- Probability simplex $\{u : u_i \geq 0$ and $\sum_{i=1}^{d} u_i = 1\}$

Basic Properties of Convex Sets

Fact

(a) If C_λ with $\lambda \in \Lambda$ are convex sets then so is their intersection $\cap_{\lambda \in \Lambda} C_\lambda$

(b) If $A, B \subseteq \mathbb{R}^d$ are convex then so is $A + B = \{u + v : u \in A$ and $v \in B\}$

(c) If $A \subseteq \mathbb{R}^d$ is convex and $c \in \mathbb{R}$ then $cA = \{cu : u \in A\}$ is convex

Convex Functions

Definition: Let $C \subseteq \mathbb{R}^d$ be convex. A function $f : C \to \mathbb{R}$ is *convex* if for every $x, y \in C$ and every $\alpha \in (0, 1)$,

$$f(\alpha x + (1-\alpha)y) \ \le \ \alpha f(x) + (1-\alpha) f(y) \tag{0.1}$$

Convexity of C ensures that $f(\cdot)$ is defined at $\alpha x + (1-\alpha)y$

Interpretation: For each $x, y \in C$ the line connecting $(x, f(x))$ and $(y, f(y))$ lies *above* the graph $\{(u, f(u)) : u \in C\} \subseteq \mathbb{R}^{d+1}$ of f

Definition

- $f : C \to \mathbb{R}$ is *strictly convex* if (0.1) holds with $\le$ replaced by $<$
- $f : C \to \mathbb{R}$ is *concave* if (0.1) holds with $\le$ replaced by $\ge$

Verifying Convexity and Concavity

1. Check the definition: In many cases it is possible to directly check the definition

2. Second derivative condition

- A function $f : C \to \mathbb{R}$ is convex if the matrix $\nabla^2 f(x)$ of second partial derivatives is well-defined and non-negative definite for each $x \in C$
- A function $f : C \to \mathbb{R}$ is concave if the matrix $\nabla^2 f(x)$ of second partial derivatives is well-defined and non-positive definite for each $x \in C$

Special case $d = 1$: f is convex if $f'' \geq 0$ and concave if $f'' \leq 0$

Examples of Convex/Concave Functions

Case $d = 1$

- $f(x) = |x|$ is convex, but *not* strictly convex
- $f(x) = x^2$, e^x, e^{-x}, x^{-1}, and $x \log x$ are strictly convex
- $f(x) = \log x$, $\sqrt{x}$ are strictly concave

Case $d \geq 2$

- $f(x) = ||x||$ is convex
- $f(x) = \langle x, u \rangle + b$, affine function, is convex and concave
- $f(x) = \sup_{u \in A} \langle x, u \rangle$, where $A \subseteq \mathbb{R}^d$ is bounded, is convex
- $f(x) = x^t A x$, quadratic form, is convex if $A \geq 0$, concave if $A \leq 0$

Basic Properties of Convex Functions

Fact: Let $C \subseteq \mathbb{R}^d$ be convex and let $f : C \to \mathbb{R}$

(a) If f is convex if $-f$ is concave, and vice-versa

(b) If f is convex or concave, it is continuous on the interior of C

(c) If $\{f_\lambda : \lambda \in \Lambda\}$ are convex functions on C then $f = \sup_{\lambda \in \Lambda} f_\lambda$ is convex

Supporting Hyperplanes

Fact: Let $C \subseteq \mathbb{R}^d$ be convex. If $f : C \to \mathbb{R}$ is convex, then for every $u \in C$ there is a vector $v \in \mathbb{R}^d$ such that

$$f(x) \;\geq\; f(u) + \langle v, x - u \rangle \;\text{ for each }\; x \in C$$

- The vector v is called a *subgradient* of f at x (not necessarily unique)
- Lower bound $h(x) := f(u) + \langle v, x - u \rangle$ is affine with $h(u) = f(u)$
- Graph of function h is a hyperplane supporting the graph of f at u
- We can express f as the maximum of its lower bounds

Jensen's Inequality

Recall: The expected value of a random vector $X = (X_1, \ldots, X_d)^t$ is defined by

$$\mathbb{E}X = (\mathbb{E}X_1, \ldots, \mathbb{E}X_d)^t \in \mathbb{R}^d$$

Jensen's Inequality: Let $C \subseteq \mathbb{R}^d$ be convex and suppose that $X \in C$. Provided that all expectations are well-defined, the following hold.

(1) The expectation $\mathbb{E}X \in C$

(2) If $f : C \to \mathbb{R}$ is convex then $f(\mathbb{E}X) \leq \mathbb{E}f(X)$. If f is strictly convex and X is not constant then the inequality is strict.

(3) If $f : C \to \mathbb{R}$ is concave then $f(\mathbb{E}X) \geq \mathbb{E}f(X)$. If f is strictly concave and X is not constant then the inequality is strict.

Note: Definition of convexity is a special case of (2) for random vector $X \in C$ with $\mathbb{P}(X = x) = \alpha$ and $\mathbb{P}(X = y) = 1 - \alpha$

Applications of Jensen's Inequality

Case $d = 1$

- $\mathbb{E}X^2 \geq (\mathbb{E}X)^2 \quad \mathbb{E}e^X \geq e^{\mathbb{E}X} \quad \mathbb{E}(X \log X) \geq (\mathbb{E}X) \log(\mathbb{E}X)$
- $\mathbb{E} \log X \leq \log \mathbb{E}X \quad \mathbb{E}\sqrt{X} \leq \sqrt{\mathbb{E}X}$

Case $d \geq 2$

- $\mathbb{E}||X|| \geq ||\mathbb{E}X||$
- $\mathbb{E}(\langle X, u \rangle + b) = \langle \mathbb{E}X, u \rangle + b$
- $\mathbb{E}(X^t A X) \leq (\mathbb{E}X)^t A (\mathbb{E}X)$ if $A \leq 0$

AM-GM inequality: If $a_1, \ldots, a_n$ are positive then $(\prod_{i=1}^n a_i)^{1/n} \leq n^{-1} \sum_{i=1}^n a_i$

Holder's Inequality

Fact: Let $a, b \geq 0$ and $1 < p, q < \infty$ be such that $1/p + 1/q = 1$. Then

$$\frac{1}{p}\, a^p \;+\; \frac{1}{q}\, b^q \;\geq\; ab$$

Holder's Inequality: Let $1 < p, q < \infty$ be such that $1/p + 1/q = 1$. If X, Y are random variables such that $\mathbb{E}|X|^p, \mathbb{E}|Y|^q$ are finite then

$$|\mathbb{E}XY| \;\leq\; \mathbb{E}|XY| \;\leq\; (\mathbb{E}|X|^p)^{1/p}\, (\mathbb{E}|Y|^q)^{1/q}$$

Cauchy-Schwartz: If $\mathbb{E}X^2$, $\mathbb{E}Y^2$ finite then $\mathbb{E}|XY| \leq \sqrt{\mathbb{E}X^2\, \mathbb{E}Y^2}$

Note: General integral version of Holder's inequality:

$$\int |f(x)g(x)|\, d\mu(x) \;\leq\; \left(\int |f(x)|^p\, d\mu(x)\right)^{1/p} \left(\int |g(x)|^q\, d\mu(x)\right)^{1/q}$$