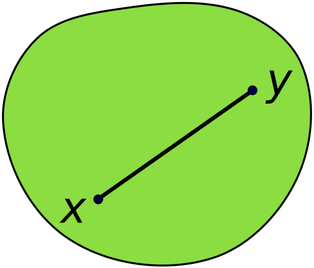
Optimization in Machine Learning

Mathematical Concepts: Convexity



Learning goals

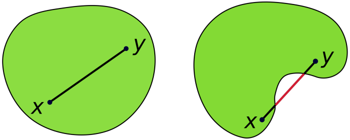
- Convex sets
- Convex functions

CONVEX SETS

A set of $\mathcal{S} \subseteq \mathbb{R}^d$ is **convex**, if for all $\mathbf{x}, \mathbf{y} \in \mathcal{S}$ and all $t \in [0, 1]$ the following holds:

$$\mathbf{x} + t(\mathbf{y} - \mathbf{x}) \in \mathcal{S}$$

Intuitively: Connecting line between any $\mathbf{x}, \boldsymbol{y} \in \mathcal{S}$ lies completely in $\mathcal{S}$.



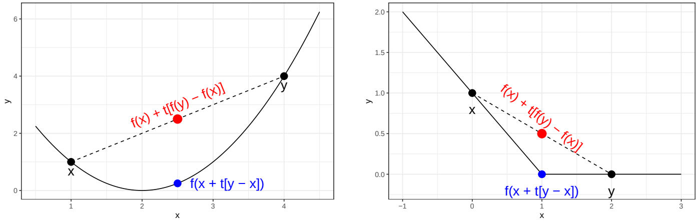
Left: convex set; right: not convex. Source: Wikipedia.

CONVEX FUNCTIONS

Consider $f : \mathcal{S} \to \mathbb{R}$, $\mathcal{S}$ convex.
f is **convex** if for all $\mathbf{x}, \mathbf{y} \in \mathcal{S}$ and all $t \in [0, 1]$

$$f(\mathbf{x} + t(\mathbf{y} - \mathbf{x})) \leq f(\mathbf{x}) + t(f(\mathbf{y}) - f(\mathbf{x})).$$

Intuitively: Connecting line lies above function.



Left: Strictly convex function. Right: Convex, but not strictly.

Strictly convex if "$<$" instead of "$\leq$". **Concave** (strictly) if the equation holds with "$\geq$" ("$>$"), respectively.

Note: f (strictly) concave $\Leftrightarrow -f$ (strictly) convex.

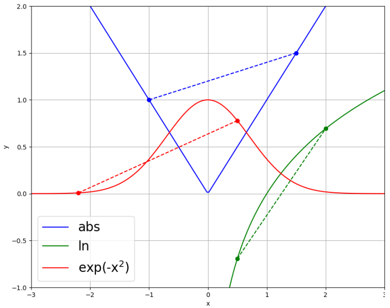
EXAMPLES

Convex function: $f(x) = |x|$.

Proof:
$$\begin{aligned} f\left(\mathbf{x} + t(\mathbf{y} - \mathbf{x})\right) &= |\mathbf{x} + t(\mathbf{y} - \mathbf{x})| = |(1-t)\mathbf{x} + t \cdot \mathbf{y}| \leq |(1-t)\mathbf{x}| + |t \cdot \mathbf{y}| \\ &= (1-t)|\mathbf{x}| + t|\mathbf{y}| = |\mathbf{x}| + t \cdot (|\mathbf{y}| - |\mathbf{x}|) \\ &= f(\mathbf{x}) + t \cdot (f(\mathbf{y}) - f(\mathbf{x})) \end{aligned}$$

Concave function: $f(x) = \log(x)$.

Neither nor: $f(x) = \exp(-x^2)$ (but log-concave)



PROVE CONVEXITY VIA HESSIAN

Let $f \in \mathcal{C}^2$ and $H(\mathbf{x})$ its Hessian.

f is **convex iff** $H(\mathbf{x})$ is positive semidefinite (p.s.d.) for all $\mathbf{x} \in \mathcal{S}$, i.e. if for all points $\mathbf{x}$ and all vectors $\mathbf{d} \neq 0$:

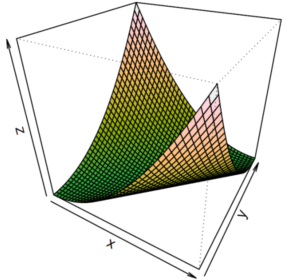
$$\mathbf{d}^\top \nabla^2 f(\mathbf{x}) \mathbf{d} \geq 0.$$

If $H(\mathbf{x})$ positive definite (strict "$>$"), f is strictly convex.

Alternatively: Matrix p.s.d. $\Leftrightarrow$ all eigenvalues ≥ 0.

PROVE CONVEXITY VIA HESSIAN

Example: $f(\mathbf{x}) = x_1^2 + x_2^2 - 2x_1x_2$, $\nabla f(\mathbf{x}) = \begin{pmatrix} 2x_1 - 2x_2 \\ 2x_2 - 2x_1 \end{pmatrix}$, $H(\mathbf{x}) = \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix}$.

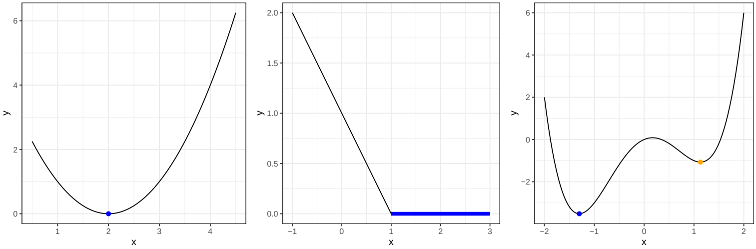


f is convex since $H(\mathbf{x})$ is p.s.d. for all $\mathbf{x}$:

$$\begin{aligned} \mathbf{d}^\top \begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix} \mathbf{d} &= \mathbf{d}^\top \begin{pmatrix} 2d_1 - 2d_2 \\ -2d_1 + 2d_2 \end{pmatrix} = 2d_1^2 - 2d_1d_2 - 2d_1d_2 + 2d_2^2 \\ &= 2d_1^2 - 4d_1d_2 + 2d_2^2 = 2(d_1 - d_2)^2 \geq 0. \end{aligned}$$

CONVEX FUNCTIONS IN OPTIMIZATION

- For a convex function, every local optimum is a global one
- A strictly convex function at most one optimal point



Left: Strictly convex; exactly one local minimum, which is also global. Middle: Convex, but not strictly; all local optima are global ones, but not unique. Right: Not convex.

CONVEX FUNCTIONS IN OPTIMIZATION

"...in fact, the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity."

- R. Tyrrell Rockafellar, in SIAM Review, 1993