Optimization for Machine Learning
CS-439

Lecture 1: Introduction & Convexity

Recap

Martin Jaggi & Nicolas Flammarion

EPFL – `github.com/epfml/OptML_course`

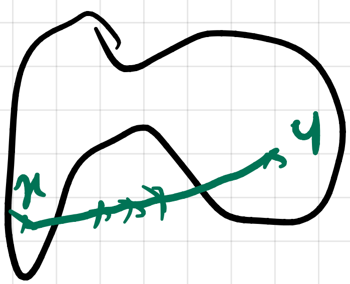
February 22, 2021

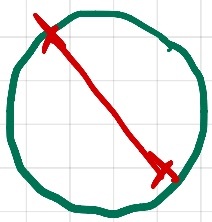
- Cauchy Schwarz

  - $|x^T y| \le \|x\|_2 \|y\|_2$ for $x, y \in \mathbb{R}^d$
  - very basic
  - lot of proofs ; i.e, $\left\| \frac{x}{\|x\|_2} - \frac{y}{\|y\|_2} \right\|^2 \geq 0$

- Convex set

  - C is convex iff $\lambda x + (1-\lambda) y \in C$
    for all $x, y \in C$ and $\lambda \in [0,1]$
  - Ex :





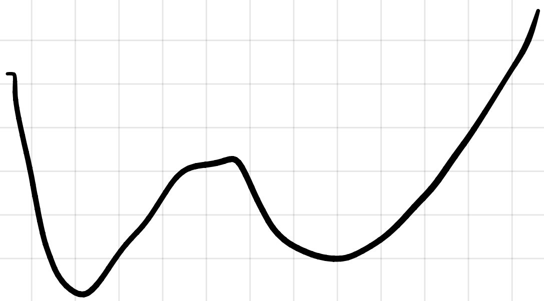
- Prop: Intersection of convex sets is convex
  Projection onto convex set is well defined

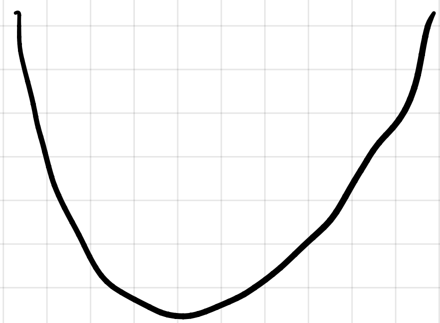
$$P_C(y) = \arg\min_{x \in C} \|x - y\|_2$$

Convex functions

- f is convex iff $f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y)$

- Ex





- Prop :- all local minimum is a global one
  - Jensen's inequality $\sum \lambda_i = 1$

$$f\left(\sum \lambda_i x_i \right) \leq \sum \lambda_i f(x_i)$$

  - Convex functions are continuous
  - convex analysis
  - very important in optimization
    → cvx pb can be solved !

How to show that a function is cvx?

Toolbox:

- usual cvx fcts : linear, affine, exp, norm ...
- operations that preserve cvx :
  - sum $f(x) + g(x)$
  - multiplication by a positive scalar
  - max $\max\{f(x), g(x)\}$
  - composition (not always, under some condition)

k^{th} order characterization of cvx

- 1° : $f(y) \geq f(x) + \nabla f(x)^T (y-x)$

  graph of f is above all its tangent hyperplane

- 2^{nd} $\nabla^2 f(x) \geq 0$

Minimizing a convex function

- Local minima are global
- critical point are global minima (you just have to set the gradient to 0)
- global minima are not unique except if the function is strictly cvx

$$f(\lambda x + (1-\lambda) y) < \lambda f(x) + (1-\lambda) f(y)$$

- Constrained minimization $\min f(x)$, $x \in C$

$$\nabla f(x_*)^T (x - x_*) \geq 0$$

- Existence of a minimizer: (not always, e.g., exp)