MATHEMATICS FOR DATA SCIENCE

September 13, 2024

Today (Remote lecture)

- Logistics
- Quick overview of the course
- Convex sets (Goal: get ready for tutorials next week)

LOGISTICS

↳ Everything on the course webpage!

https://www.lamsade.dauphine.fr/~croyer/teachMDS.html

↳ For any questions, email me!

clement.royer@lamsade.dauphine.fr

↳ 12 lectures + 12 tutorial sessions in groups

- Lectures Fridays 1.45pm - 3.15pm
- Tutorials Wednesdays 3.30pm - 5pm
  Thursdays 10.15am - 11.45am

⚠ Missing lecture in schedule (Nov. 1st)

↳ Break mid-October, and mid-December

↳ Grading:

70 % Exam

* January
* English / French versions
* Allowed: 1 A4 sheet of notes

30 % Homework (details TBA)

COURSE ROADMAP

Goal: Review mathematical concepts that are useful for modeling and understanding data

1. Convexity ⇒ Certify nice structure within/around data

2. Optimization ⇒ Formulate data science tasks (e.g. regression, classification) as convex optimization problems

3. Applications & Statistics ⇒ Consider data as coming from some probability distribution

Lecture notes

Ch 0 Mathematical background & notations

Ch 1 Convexity

Today: 1.1. Convex sets

Def: A set $X \subseteq \mathbb{R}^n$ is called a convex set if

$\forall (x,y) \in X^2$, $\forall \alpha \in [0,1]$, $\alpha x + (1-\alpha) y \in X$

$\mathbb{R}^n$: set of vectors with n real coordinates

$x \in \mathbb{R}^n$, $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ "column vector"

Ex)

Ex)
- $\mathbb{R}^n$ is a convex set
- A line segment is convex

  $\{\alpha x + (1-\alpha) y \mid \alpha \in [0,1]\}$: line segment between $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$

- For any $x \in \mathbb{R}^n$ and $r \geq 0$:

  Norm: $\forall x \in \mathbb{R}^n$, $\|x\| = \sqrt{\sum_{i=1}^{n} x_i^2}$

  - the closed ball $\{y \in \mathbb{R}^n \mid \|y - x\| \leq r\}$ is convex
  - the open ball $\{y \in \mathbb{R}^n \mid \|y - x\| < r\}$ is convex

↳ Special classes of convex sets

- Affine sets: $\mathcal{X} \subseteq \mathbb{R}^n$ is an affine set if

  $\forall (x,y) \in \mathcal{X}^2, \ \forall \alpha \in \mathbb{R}, \ \alpha x + (1-\alpha) y \in \mathcal{X}$

Ex) $\mathbb{R}^n$, any line $\{\alpha x \mid \alpha \in \mathbb{R}\}$ with $x \in \mathbb{R}^n$

Hyperplane $\{x \in \mathbb{R}^n \mid a^T x = b\}$ where $a \in \mathbb{R}^n, b \in \mathbb{R}$

Scalar/inner product

$a^T x = \sum_{i=1}^{n} a_i x_i$ $\quad (\|x\|^2 = x^T x)$

$a^T = [a_1 \dots a_n]$

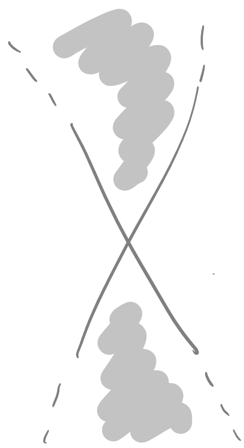
- Convex cones

Def: • $K \subseteq \mathbb{R}^n$ is called a cone if

$\forall x \in K, \ \forall \alpha > 0, \ \alpha x \in K$

• K is called a convex cone if it both a cone and a convex set.

Ex)



A nonconvex cone

Ex) • Open line $\{t > 0\}$ in $\mathbb{R}$

• Half space $\{x \in \mathbb{R}^n \mid a^T x \leq 0\}$

What are these sets good for?

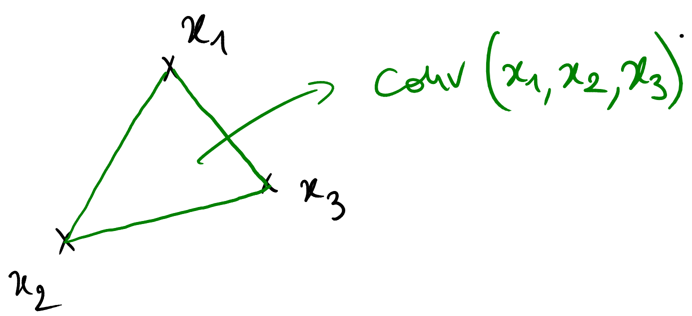
↳ In practice, we are typically given a set of data points $x_1, \ldots, x_k \in \mathbb{R}^n$, not a set!

↳ To analyze this data more easily, we can consider convex sets that contain those points!

Def: Given $(x_1, \ldots, x_k) \in (\mathbb{R}^n)^k$,

- A convex combination of $x_1, \ldots, x_k$ is a vector of the form $\sum_{i=1}^k \alpha_i x_i$ where $\alpha_i \geq 0 \quad \forall i = 1..k$, $\sum_{i=1}^k \alpha_i = 1$
- The set of convex combinations of $x_1, \ldots, x_k$ is called the convex hull of $x_1, \ldots, x_k$. It is a convex set, and we denote it by $\mathrm{conv}(x_1, \ldots, x_k)$.

Ex)



Remarks:
- The notion of convex hull generalizes to sets $X \subseteq \mathbb{R}^n$
- $X \subseteq \mathbb{R}^n$ is convex $\iff \mathrm{conv}(X) = X$

↳ Affine combination: $\sum_{i=1}^k \alpha_i x_i$ with $\sum_{i=1}^k \alpha_i = 1$

↳ Conic combination: $\sum_{i=1}^k \alpha_i x_i$ with $\alpha_i \geq 0 \quad i = 1..k$

Theorem (Separating hyperplane)

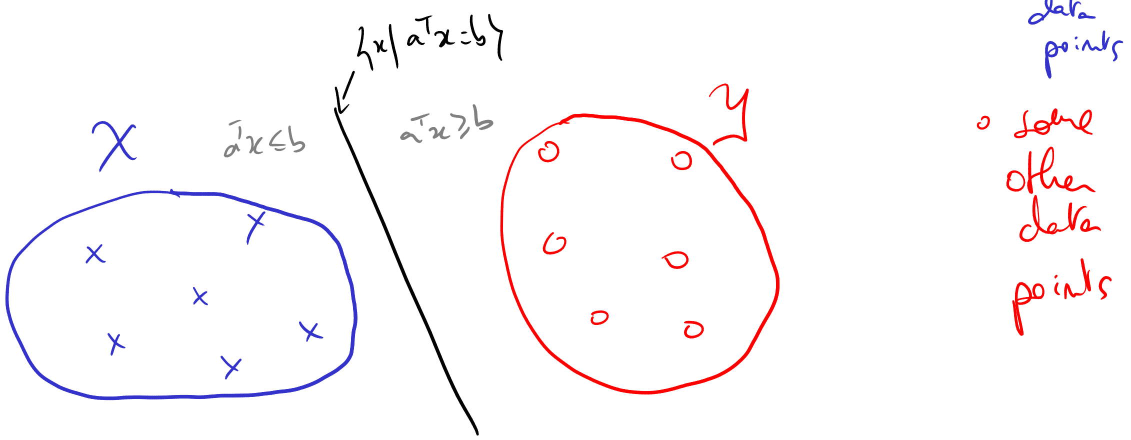
$\mathcal{X} \cap \mathcal{Y} = \emptyset$

For any disjoint convex sets $\mathcal{X}$ and $\mathcal{Y}$, there exist $a \in \mathbb{R}^n$, $\|a\| \neq 0$ and $b \in \mathbb{R}$ such that the hyperplane $\{x \in \mathbb{R}^n \mid a^T x = b\}$ separates $\mathcal{X}$ and $\mathcal{Y}$, that is

$$\forall x \in \mathcal{X}, \quad a^T x \leq b$$

$$\forall y \in \mathcal{Y}, \quad a^T y \geq b$$

Illustration



→ Result at the heart of linear separability / linear binary classification / linear Support vector machines

→ Our motivating example throughout the course

- Convexity: $\exists$ hyperplane / linear classifier
- Optimization: We can find a hyperplane efficiently
- Statistics: We can find a good hyperplane w.r.t. data distribution

Summary

All this (and more details) in Section 1.1. of the lecture notes

- Convex set: definition, to be used for proving convexity of a set
- Special cases: affine sets, convex cones
- Convex hull/combination, affine/conic counterparts
- Motivating example, to be revisited