Algebraic geometry in machine learning

Jackson Van Dyke

October 20, 2020

Data-Pop alliance

"Data-Pop Alliance is a collaborative laboratory created by

- the Harvard Humanitarian Initiative,
- MIT Connection Science, and
- Overseas Development Institute.

We bring together researchers, experts, practitioners, and activists to change the world with data through three pillars of work:

1. diagnosing local realities and human problems with data and AI;
2. mobilizing capacities, communities, and ideas towards more data literate societies, and
3. transforming the systems and processes that underpin our societies and countries."

Find out more on their website.

Motivation

$$\text{raw data} \quad \rightsquigarrow \quad \{v_i\} \in \mathbb{R}^N \quad \rightsquigarrow \quad \text{subspace} \subseteq \mathbb{R}^N$$

Example

If we start with k images, we can split it into N squares and take the grayscale values to get k vectors in $\mathbb{R}^N$. Then we can

- take the span,
- take the affine span, or
- take the smallest ellipsoid containing the vectors.

Before doing anything else with these subspaces, we want to develop some notion of distance between them.

Table of contents

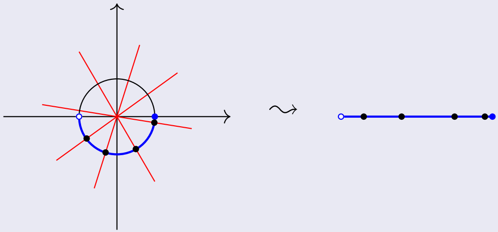
Distance

Question

What is the distance between two linear subspaces?

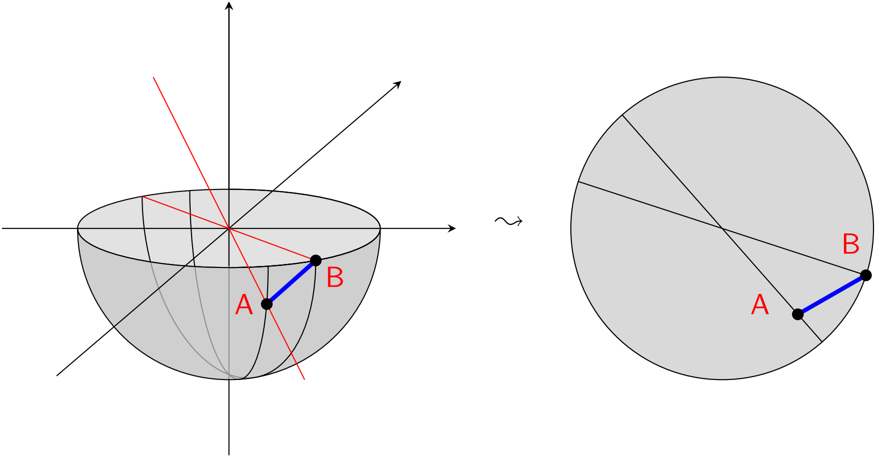
Example

For lines in $\mathbb{R}^2$, we just need to take the angle.



So now we want to formalize this in high dimensions.

Higher-dimensional picture



distance (A,B) = blue.

Higher-dimensional setup

Let $a_1, \ldots, a_k \in \mathbb{R}^N$ and $b_1, \ldots, b_k \in \mathbb{R}^N$ be (separately) linearly independent sets of vectors. Write their spans as:

$$A \coloneqq \operatorname{Span}\{a_1, \ldots, a_k\} \subset \mathbb{R}^N \qquad B \coloneqq \operatorname{Span}\{b_1, \ldots, b_k\} \subset \mathbb{R}^N .$$

Since the vectors were linearly independent, A and B are both k-dimensional linear subspaces of $\mathbb{R}^N$.

Therefore A and B are points of the **Grassmannian**.

$$A, B \in \operatorname{Gr}(k, N) \coloneqq \left\{ k - \text{dim'l linear subspaces of } \mathbb{R}^N \right\} .$$

Principal vectors and angles

- Write $\widehat{a}_1 \in A$ and $\widehat{b}_1 \in B$ for the vectors which

$$\begin{aligned} \text{maximize} \qquad & a^T b \\ \text{such that} \qquad & \|a\| = \|b\| = 1 \end{aligned}$$

  for $a \in A$, $b \in B$.
- Write $\widehat{a}_2 \in A$ and $\widehat{b}_2 \in B$ for the vectors which

$$\begin{aligned} \text{maximize} \qquad & a^T b \\ \text{such that} \qquad & \|a\| = \|b\| = 1 \\ & a^T \widehat{a}_1 = 0, \quad b^T \widehat{b}_1 = 0 \end{aligned}$$

  for $a \in A$ and $b \in B$.
- In general we ask for $\widehat{a}_j$ (resp. $\widehat{b}_j$) to be orthogonal to $\widehat{a}_i$ (resp. $\widehat{b}_i$) for all $i < j$.

Grassmann distance

- We can think of the principal vectors as forming a basis which is convenient for measuring angles.
- Define the **principal angles** θ_j by

$$\cos\theta_j = \widehat{a}_j^T \widehat{b}_j \ .$$

  Note that $\theta_1 \leq \ldots \leq \theta_k$.
- The **Grassmann distance** between the linear subspaces A and B is given by:

$$d_k(A, B) = \left(\sum_{i=1}^{k} \theta_i^2 \right)^{1/2} \ .$$

Principal angles in $\mathrm{Gr}(2,3)$

Consider two planes in $\mathbb{R}^3$ given by

$$A = \mathrm{Span}(e_1, e_2)$$
$$B = \mathrm{Span}(e_2, e_1 + e_3) \ .$$

The principal vectors are:

$$\widehat{a_1} = e_1 \qquad\qquad \widehat{b_1} = e_2$$
$$\widehat{a_2} = e_2 \qquad\qquad \widehat{b_2} = e_3$$

So the principal angles are:

$$\theta_1 = 0 \qquad\qquad \theta_2 = \pi/2$$

and

$$d(A, B) = \pi/2 \ . \tag{1}$$

Metric

We have been using the word "distance" a bit loosely.

Technically, d defines a **metric** on $\mathrm{Gr}(k, N)$ because it satisfies:

1. $d(A, B) = 0$ if and only if $A = B$,
2. $d(A, B) = d(B, A)$, and
3. $d(A, C) \leq d(A, B) + d(B, C)$

for all A, B, and $C \in \mathrm{Gr}(k, N)$.

An example

- By separating images into three regions:

$$2 \text{ images of someone's face } \rightsquigarrow v_1, v_2 \in \mathbb{R}^3$$

- If v_1 and v_2 are linearly independent, we get a plane:

$$F := \operatorname{Span}(v_1, v_2) = \{m_1 v_1 + m_2 v_2 \mid m_1, m_2 \in \mathbb{R}\} \subset \mathbb{R}^3 .$$

- For two new photos of someone, again we get a plane and we can take the distance to F as a way to compare to the original photos.
- But what if I only have one picture of someone, and I want to compare it to the two I started with?

Question

How do we compare subspaces of different dimensions?

Schubert varieties

- For $k \leq \ell$, we would like a notion of distance between
$$A \in \mathrm{Gr}(k, N) \qquad\qquad B \in \mathrm{Gr}(\ell, N) \ .$$
- Consider the set of ℓ-planes containing A:
$$\Omega_+(A) := \{P \in \mathrm{Gr}(\ell, N) \mid A \subseteq P\}$$
and the set of all k-planes containing B:
$$\Omega_-(B) := \{P \in \mathrm{Gr}(k, N) \mid P \subseteq B\} \ .$$
These are examples of **Schubert varieties**. E.g.
$$\Omega_+(\text{the line}) = \{\text{planes containing the line}\}$$
$$\Omega_-(\text{the plane}) = \{\text{lines contained in the plane}\} \ .$$
- **Strategy:** measure distance from A to $\Omega_-(B)$, and B to $\Omega_+(A)$ and compare.

Distance between linear subspaces of different dimensions

The distance from A to $\Omega_-(B)$ is given by:

$$\delta_- = \min\{d_k(P,A) \mid P \in \Omega_-(B)\} \ .$$

and the distance from B to $\Omega_+(A)$ is given by

$$\delta_+ = \min\{d_\ell(P,B) \mid P \in \Omega_+(A)\} \ .$$

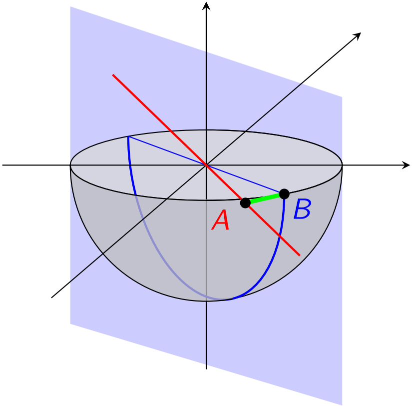
Theorem 1 (Ye-Lim 2016 [YL16])

$\delta_+ = \delta_-$, *and the common value is:*

$$\delta(A,B) = \left(\sum_{i=1}^{\min(k,\ell)} \theta_i^2\right)^{1/2} \ .$$

Example

Now A is still a line, but B is a plane, both still in $\mathbb{R}^3$.



The distance is the only principal angle that can be defined: the first one. So

$$\delta(\mathsf{A}, \mathsf{B}) = \text{green} \; .$$

Metric?

- Recall d was a metric on $\mathrm{Gr}(k, N)$.
- The space of all linear subspaces in all dimensions is the **doubly infinite Grassmannian**: $\mathrm{Gr}(\infty, \infty) = \sqcup_{k=1}^{\infty} \mathrm{Gr}(k, \infty)$.

Question

Does δ define a metric on $\mathrm{Gr}(\infty, \infty)$?

No: it only satisfies symmetry.

$$\delta(A, B) = 0 \quad \iff \quad A \subseteq B \text{ or } B \subseteq A$$

Counterexample

Let $L_1, L_2 \in \mathrm{Gr}(1, N)$, $P \in \mathrm{Gr}(2, N)$ such that $L_1, L_2 \subset P$.
Triangle inequality $\implies \delta(L_1, L_2) = \delta(L_1, P) = 0$. **Contradiction.**

Premetric.

Instead, δ is what is called a **premetric** (or **distance**) on $\mathrm{Gr}(\infty, \infty)$, since it satisfies:

1. $d(A, B) \geq 0$,
2. $d(A, A) = 0$, and
3. $d(A, B) = d(B, A)$

for all $A, B \in \mathrm{Gr}(\infty, \infty)$.

This can be thought of more as a way to measure *separation*, in the sense of the distance between a point and a set.

Metric after all?

- Recall we can express $\delta(A,B) = \left(\sum_{i=1}^{\min(k,\ell)=k} \theta_i^2\right)^{1/2}$.
- Instead of stopping at k, we can just define $\theta_i = \pi/2$ for $i \gneq k$.
- Then

$$d_\infty(A,B) = \left(\sum_{i=1}^{\max(k,\ell)=\ell} \theta_i^2\right)^{1/2} \tag{2}$$

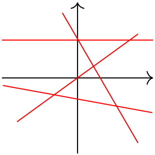
  is a **metric** on $\mathrm{Gr}(\infty,\infty)$.
- When restricted to $\mathrm{Gr}(k,\infty)$, this agrees with d_k.
- Geometrically this is saying that we stabilize the smaller subspace by crossing with copies of $\mathbb{R}$ and then taking the ℓ metric.

Affine subspaces

- Let $A \in \mathrm{Gr}(k, N)$ be a k-dimensional linear subspace and $b \in \mathbb{R}^N$ to be thought of as the "shift" away from the origin.
- Write $\{a_1, \ldots, a_k\}$ for some basis of A.
- The associated **affine subspace** is:

$$A + b := \left\{ m_1 a_1 + \ldots + m_k a_k + b \in \mathbb{R}^N \,\middle|\, \lambda_i \in \mathbb{R} \right\} \subset \mathbb{R}^N .$$

In particular, they don't have to contain the origin.

E.g. $\mathrm{Graff}(0, N) = \mathbb{R}^N$, and $\mathrm{Graff}(1, N) =$ 

Together, the affine subspaces form the **Grassmannian of affine subspaces**:

$$\mathrm{Graff}(k, N) = \left\{ k\text{-dim'l affine subspaces of } \mathbb{R}^N \right\} .$$

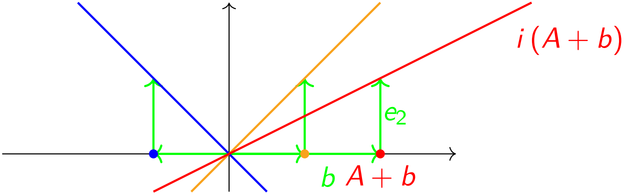
Embedding Graff in (a bigger) Gr

- **Strategy:** view affine subspaces as linear subspaces of a higher-dimensional space, and take d_{Gr}:

$$\mathrm{Graff}\,(k, N) \xhookrightarrow{i} \mathrm{Gr}\,(k+1, N+1)$$

$$A + b \longmapsto \mathrm{Span}\,(A \cup \{b + e_{n+1}\})$$

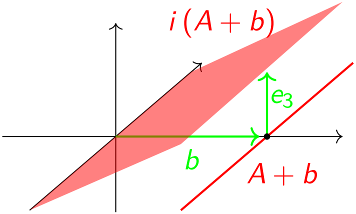
- When $k = 0$ and $N = 1$, i sends points of $\mathbb{R}$ to lines of $\mathbb{R}^2$.
- Given a point $\bullet$, taking this span is the same as drawing a line from the point a unit distance above $\bullet$ through the origin.



Embedding Graff in (a bigger) Gr

$$\mathrm{Graff}\,(1,2) \xhookrightarrow{i} \mathrm{Gr}\,(2,3)$$

$$A + b \longmapsto \mathrm{Span}\,(A \cup \{b + e_3\})$$



A metric on Graff

We use this embedding to define the distance between two affine subspaces:

$$d_{\mathrm{Graff}(k,N)}\left(A+b, B+c\right) := d_{\mathrm{Gr}(k+1,N+1)}\left(i\left(A+b\right), i\left(B+c\right)\right) \ .$$

- d_{Graff} is a metric because d_{Gr} is.
- If $b = c = 0$, this is just the usual Grassmannian distance.
- Just as the distance between linear subspaces was calculated using the principal angles, there are **affine principal angles** such that this distance is written as before.
- These angles are also computationally manageable.

An example

- By separating two images into three regions we get $v_1, v_2 \in \mathbb{R}^3$.
- If they are linearly independent, we get a line L which contains those points:

$$L := \{m_1 v_1 + m_2 v_2 \mid m_1 + m_2 = 1, m_1, m_2 \in \mathbb{R}\} \subset \mathbb{R}^3 \ .$$

  This is the **affine span/hull of v_1 and v_2**.
- The affine hull is the smallest affine subspace containing the data. In particular, it is contained in the linear subspace F from before.
- For two new photos of someone, again we get a line and we can take the distance to L to compare to the originals.

Question

How do we compare subspaces of different dimensions?

Distance for inequidimensional affine subspaces

For $k \leq \ell$, we would like a notion of distance between

$$A + b \in \mathrm{Graff}(k, N) \qquad\qquad B + c \in \mathrm{Graff}(\ell, N) \ .$$

As in the linear case, define

$$\begin{aligned}\Omega_+(A+b) &:= \{P + q \in \mathrm{Graff}(\ell, N) \mid A + b \subseteq P + q\} \\ \Omega_-(B+c) &:= \{P + q \in \mathrm{Graff}(k, N) \mid P + q \subseteq B + c\} \ .\end{aligned}$$

Theorem 2 (Lim-Wong-Ye 2018 [LWY18])

$d_{\mathrm{Graff}(k,N)}(A + b, \Omega_-(B + c)) = d_{\mathrm{Graff}(\ell,N)}(B + c, \Omega_+(A + b))$, *and it is explicitly given via the affine principle angles.*

d_{Graff} is a metric because d_{Gr} is.

Generalizations

- This whole story holds for ellipsoids in $\mathbb{R}^N$ as well.
- The distance between two ellipsoids is the distance between the matrices defining them.
- Therefore it reduces to the analogous calculations in the cone of real symmetric/complex Hermitian matrices.
- These techniques should extend to any situation where your space looks like a space of matrices.

Future directions

- A **category** is (roughly) a collection of objects and arrows between the objects which satisfy some conditions.
- In [DHKK13], the authors define a notion of distance between any two objects of a category.

Example

The collection of half-dimensional subspaces of a given even-dimensional manifold[a] fit naturally into a category called the **Fukaya category**. Roughly, we have an object for every subspace, and an arrow whenever they intersect.

[a] Technically they're Lagrangians in a symplectic manifold.

Question

Is this a useful distance for our purposes? Is it computable?

Sheaves

- Consider some subset $S \subseteq \mathbb{R}^N$ of samples equipped with a function $f_S : S \to \{0, 1\}$.
- This should be thought of as telling us if a datum does (1) or does not (0) satisfy something.
- This defines a sheaf $\widetilde{S}$ by assigning an open subset U to:

$$\widetilde{S}(U) := \left\{ f : \mathbb{R}^N \to \{0, 1\} \,\middle|\, f|_S = f_S \right\} . \tag{3}$$

- The function which guesses if each datum in $\mathbb{R}^N$ does or does not satisfy something (consistently with the samples S) is a section of $\widetilde{S}$.

Graph quotients

In [Lin17], they study graph quotients as “sheaves” over graphs.

Example 1

Consider the following (real) tweets:

Dating a **skeleton**.

If you’re **skeleton** you can buy velveeta with bones.

12 foot Home Depot **skeleton**.

- Each word is meant to be a vertex; edges are assumed to connect the vertices together in some way.
  - E.g. the edges join together neighboring words.
- It’s a “sheaf” in the sense that the stalk over “skeleton” consists of the different instances of skeleton in the data.

Summary:

Assume we have a way to pass from raw data to a subspace:

$$\text{raw data} \quad \rightsquigarrow \quad \{v_i\} \in \mathbb{R}^N \quad \rightsquigarrow \quad \text{subspace} \subseteq \mathbb{R}^N$$

When the subspace is linear, affine, or an ellipsoid, there is a metric (or premetric) which on the space of such subspaces (of any dimension!) which is realistic to calculate.

So we can distinguish data by measuring the distance between the associated subspaces.

References

George Dimitrov, Fabian Haiden, Ludmil Katzarkov, and Maxim Kontsevich, *Dynamical systems and categories*, 2013.

Vepstas Linas, *Graph quotients: A topological approach to graphs*, Bulletin of the Novosibirsk Computing Center. Series: Computing Science. (2017), no. 41.

Lek-Heng Lim, Ken Sze-Wai Wong, and Ke Ye, *The grassmannian of affine subspaces*, 2018.

Ke Ye and Lek-Heng Lim, *Schubert varieties and distances between subspaces of different dimensions*, SIAM J. Matrix Anal. Appl. **37** (2016), no. 3, 1176–1197. MR 3543155