Analytic Geometry

Introduction to Optimization for Machine Learning
M1 MLSD/AMSD

October 17, 2023

Roadmap

(1) Norms

(2) Inner Products

(3) Lengths and Distances

(4) Angles and Orthogonality

(5) Orthogonal Projections

Norm

- A notion of the length of vectors
- Definition. A norm on a vector space V is a function $\|\cdot\| : V \mapsto \mathbb{R}$, such that for all $\lambda \in \mathbb{R}$ the following hold:
  - Absolutely homogeneous: $\|\lambda \boldsymbol{x}\| = |\lambda| \, \|\boldsymbol{x}\|$
  - Triangle inequality: $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$
  - Positive definite: $\|\boldsymbol{x}\| \geq 0$ and $\|\boldsymbol{x}\| \iff \boldsymbol{x} = \boldsymbol{0}$

Example for $V \in \mathbb{R}^n$

- Manhattan Norm (also called ℓ_1 norm) For $\boldsymbol{x} = [x_1, \cdots, x_n] \in \mathbb{R}^n$,

$$\|\boldsymbol{x}\|_1 :== \sum_{i=1}^{n} |x_i|$$

- Euclidean Norm (also called ℓ_2 norm) For $\boldsymbol{x} \in \mathbb{R}^n$,

$$\|\boldsymbol{x}\|_2 :== \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{\boldsymbol{x}^\top \boldsymbol{x}}$$

Motivation

- Need to talk about the length of a vector and the angle or distance between two vectors, where vectors are defined in abstract vector spaces
- To this end, we define the notion of inner product in an abstract manner.
- Dot product: A kind of inner product in vector space $\mathbb{R}^n$. $\boldsymbol{x}^\mathsf{T}\boldsymbol{y} = \sum_{i=1}^{n} x_i y_i$
- Question. How can we generalize this and do a similar thing in some other vector spaces?

Formal Definition

- An inner product is a mapping $\langle \cdot, \cdot \rangle : V \times V \mapsto \mathbb{R}$ that satisfies the following conditions for all vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$ and all scalars $\lambda \in \mathbb{R}$:
    1. $\langle \boldsymbol{u} + \boldsymbol{v}, \boldsymbol{w} \rangle = \langle \boldsymbol{u}, \boldsymbol{w} \rangle + \langle \boldsymbol{v}, \boldsymbol{w} \rangle$
    2. $\langle \lambda \boldsymbol{v}, \boldsymbol{w} \rangle = \lambda \langle \boldsymbol{v}, \boldsymbol{w} \rangle$
    3. $\langle \boldsymbol{v}, \boldsymbol{w} \rangle = \langle \boldsymbol{w}, \boldsymbol{v} \rangle$
    4. $\langle \boldsymbol{v}, \boldsymbol{v} \rangle \geq 0$ and equal iff $\boldsymbol{v} = \mathbf{0}$
- The pair $(V, \langle \cdot, \cdot \rangle)$ is called an inner product space.

Example

- Example. $V = \mathbb{R}^n$ and the dot product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \boldsymbol{x}^{\mathsf{T}} \boldsymbol{y}$
- Example. $V = \mathbb{R}^2$ and $\langle \boldsymbol{x}, \boldsymbol{y} \rangle := x_1 y_1 - (x_1 y_2 + x_2 y_1) + 2 x_2 y_2$
- Example. $V = \{\text{continuous functions in } \mathbb{R} \text{ over } [a, b]\}$, $\langle u, v \rangle := \int_a^b u(x) v(x) dx$

Symmetric, Positive Definite Matrix

- A square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ that satisfies the following is called symmetric, positive definite (or just positive definite):

$$\forall \boldsymbol{x} \in V \setminus \{\mathbf{0}\} : \boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x} > 0.$$

  If only $\geq$ in the above holds, then $\boldsymbol{A}$ is called symmetric, positive semidefinite.

- $\boldsymbol{A}_1 = \begin{pmatrix} 9 & 6 \\ 6 & 5 \end{pmatrix}$ is positive definite.

- $\boldsymbol{A}_2 = \begin{pmatrix} 9 & 6 \\ 6 & 3 \end{pmatrix}$ is not positive definite.

Inner Product and Positive Definite Matrix (1)

- Consider an n-dimensional vector space V with an inner product $\langle \cdot, \cdot \rangle$ and an ordered basis $B = (\boldsymbol{b}_1, \dots, \boldsymbol{b}_n)$ of V.

- Any $\boldsymbol{x}, \boldsymbol{y} \in V$ can be represented as: $\boldsymbol{x} = \sum_{i=1}^{n} \psi_i \boldsymbol{b}_i$ and $\boldsymbol{y} = \sum_{i=j}^{n} \lambda_j \boldsymbol{b}_j$ for some ψ_i and λ_j, $i, j = 1, \dots, n$.

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \left\langle \sum_{i=1}^{n} \psi_i \boldsymbol{b}_i, \sum_{i=j}^{n} \lambda_j \boldsymbol{b}_j \right\rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} \psi_i \langle \boldsymbol{b}_i, \boldsymbol{b}_j \rangle \lambda_j = \hat{\boldsymbol{x}}^{\mathsf{T}} \boldsymbol{A} \hat{\boldsymbol{y}},$$

  where $\boldsymbol{A}_{ij} = \langle \boldsymbol{b}_i, \boldsymbol{b}_j \rangle$ and $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ are the coordinates w.r.t. B.

- Then, if $\forall \boldsymbol{x} \in V \setminus \{\boldsymbol{0}\} : \boldsymbol{x}^{\mathsf{T}} \boldsymbol{A} \boldsymbol{x} > 0$ (i.e., $\boldsymbol{A}$ is symmetric, positive definite), $\hat{\boldsymbol{x}}^{\mathsf{T}} \boldsymbol{A} \hat{\boldsymbol{y}}$ legitimately defines an inner product (w.r.t. B)

Roadmap

(1) Norms

(2) Inner Products

(3) Lengths and Distances

(4) Angles and Orthogonality

(5) Orthogonal Projections

Length

- Inner product naturally induces a norm by defining:

$$\|x\| := \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$$

- Not every norm is induced by an inner product
- Cachy-Schwarz inequality. For the induced norm by the inner product,

$$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leq \|\boldsymbol{x}\| \; \|\boldsymbol{y}\|$$

Distance

- Now, we can introduce a notion of distance using a norm as:

  Distance. $d(\boldsymbol{x}, \boldsymbol{y}) := \|\boldsymbol{x} - \boldsymbol{y}\| = \sqrt{\langle \boldsymbol{x} - \boldsymbol{y}, \boldsymbol{x} - \boldsymbol{y} \rangle}$

- If the dot product is used as an inner product in $\mathbb{R}^n$, it is Euclidian distance.
- Note. The distance between two vectors does NOT necessarily require the notion of norm. Norm is just sufficient.
- Generally, if the following is satisfied, it is a suitable notion of distance, called metric.
  - *Positive definite*. $d(\boldsymbol{x}, \boldsymbol{y}) \geq 0$ for all $\boldsymbol{x}, \boldsymbol{y}$ and $d(\boldsymbol{x}, \boldsymbol{y}) = 0 \iff \boldsymbol{x} = \boldsymbol{y}$
  - *Symmetric*. $d(\boldsymbol{x}, \boldsymbol{y}) = d(\boldsymbol{y}, \boldsymbol{x})$
  - *Triangle inequality*. $d(\boldsymbol{x}, \boldsymbol{z}) \leq d(\boldsymbol{x}, \boldsymbol{y}) + d(\boldsymbol{y}, \boldsymbol{z})$

Angle, Orthogonal, and Orthonormal

- Using C-S inequality,

$$-1 \leq \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{x}\| \ \|\boldsymbol{y}\|} \leq 1$$

- Then, there exists a unique $\omega \in [0, \pi]$ with

$$\cos \omega = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{x}\| \ \|\boldsymbol{y}\|}$$

- We define ω as the angle between $\boldsymbol{x}$ and $\boldsymbol{y}$.

- Definition. If $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$, in other words their angle is $\pi/2$, we say that they are orthogonal, denoted by $\boldsymbol{x} \perp \boldsymbol{y}$. Additionally, if $\|x\| = \|y\| = 1$, they are orthonormal.

Example

- Orthogonality is defined by a given inner product. Thus, different inner products may lead to different results about orthogonality.
- Example. Consider two vectors $\boldsymbol{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\boldsymbol{y} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$
- Using the dot product as the inner product, they are orthogonal.
- However, using $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^{\mathsf{T}} \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \boldsymbol{y}$, they are not orthogonal.

$$\cos \omega = \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{x}\| \; \|\boldsymbol{y}\|} = -\frac{1}{3} \implies \omega \approx 1.91 \text{ rad } \approx 109.5^\circ$$

Orthogonal Matrix

- Definition. A square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is an orthogonal matrix, iff its columns (or rows) are orthonormal so that

$$\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}} = I = \boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}, \text{ implying } \boldsymbol{A}^{-1} = \boldsymbol{A}^{\mathsf{T}}.$$

  - We can use $\boldsymbol{A}^{-1} = \boldsymbol{A}^{\mathsf{T}}$ for the definition of orthogonal matrices.
  - Fact 1. $\boldsymbol{A}, \boldsymbol{B}$: orthogonal $\implies$ $\boldsymbol{AB}$: orthogonal
  - Fact 2. $\boldsymbol{A}$: orthogonal $\implies$ $\det(\boldsymbol{A}) = \pm 1$

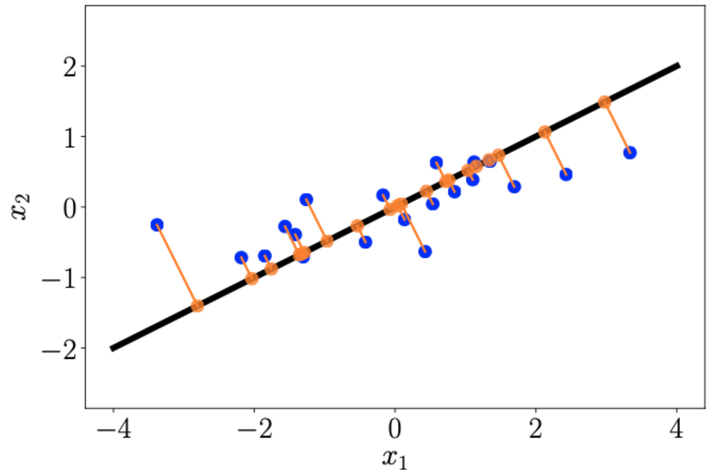
- The linear mapping Φ by orthogonal matrices preserve length and angle (for the dot product)

$$\|\Phi(\boldsymbol{A})\| = \|\boldsymbol{Ax}\|^2 = (\boldsymbol{Ax})^{\mathsf{T}}(\boldsymbol{Ax}) = \boldsymbol{x}^{\mathsf{T}}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{Ax} = \boldsymbol{x}^{\mathsf{T}}\boldsymbol{x} = \|\boldsymbol{x}\|^2$$

$$\cos\omega = \frac{(\boldsymbol{Ax})^{\mathsf{T}}(\boldsymbol{Ay})}{\|\boldsymbol{Ax}\| \ \|\boldsymbol{Ay}\|} = \frac{\boldsymbol{x}^{\mathsf{T}}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{Ay}}{\sqrt{\boldsymbol{x}^{\mathsf{T}}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{Axy}^{\mathsf{T}}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{Ay}}} = \frac{\boldsymbol{x}^{\mathsf{T}}\boldsymbol{y}}{\|\boldsymbol{x}\| \ \|\boldsymbol{y}\|}$$

Projection: Motivation

- Big data: high dimensional
- However, most information is contained in a few dimensions
- Projection: A process of reducing the dimensions (hopefully) without loss of much information[1]
- Example. Projection of 2D dataset onto 1D subspace



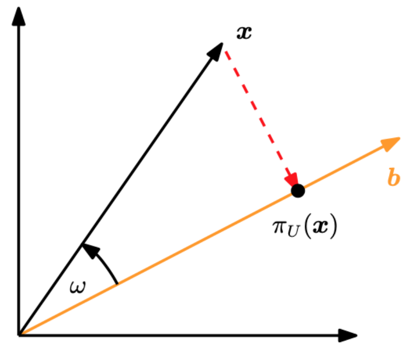
[1] In L10, we will formally study this with the topic of PCA (Principal Component Analysis).

Projection onto Lines (1D Subspaces)

- Consider a 1D subspace $U \subset \mathbb{R}^n$ spanned by the basis $\boldsymbol{b}$.
- For $\boldsymbol{x} \in \mathbb{R}^n$, what is its projection $\pi_U(\boldsymbol{x})$ onto U (assume the dot product)?

$$\langle \boldsymbol{x} - \pi_U(\boldsymbol{x}), \boldsymbol{b} \rangle = 0 \xleftrightarrow{\pi_U(\boldsymbol{x}) = \lambda \boldsymbol{b}} \langle \boldsymbol{x} - \lambda \boldsymbol{b}, \boldsymbol{b} \rangle = 0$$

$$\implies \lambda = \frac{\langle \boldsymbol{b}, \boldsymbol{x} \rangle}{\|\boldsymbol{b}\|^2} = \frac{\boldsymbol{b}^\mathsf{T} \boldsymbol{x}}{\|\boldsymbol{b}\|^2}, \text{ and } \pi_U(\boldsymbol{x}) = \lambda \boldsymbol{b} = \frac{\boldsymbol{b}^\mathsf{T} \boldsymbol{x}}{\|\boldsymbol{b}\|^2} \boldsymbol{b}$$



(a) Projection of $\boldsymbol{x} \in \mathbb{R}^2$ onto a subspace U with basis vector $\boldsymbol{b}$.

- Projection matrix $\boldsymbol{P}_\pi \in \mathbb{R}^{n \times n}$ in $\pi_U(\boldsymbol{x}) = \boldsymbol{P}_\pi \boldsymbol{x}$

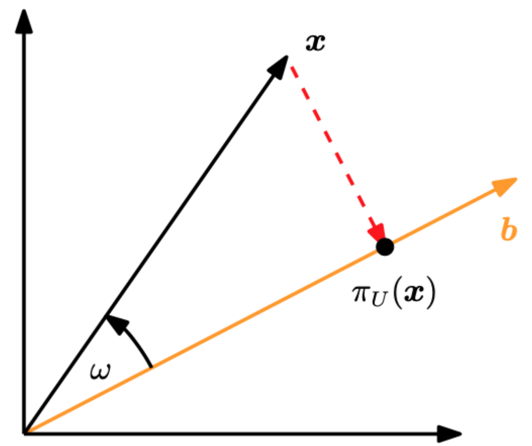
$$\pi_U(\boldsymbol{x}) = \lambda \boldsymbol{b} = \boldsymbol{b} \lambda = \frac{\boldsymbol{b}\boldsymbol{b}^\mathsf{T}}{\|\boldsymbol{b}\|^2} \boldsymbol{x}, \quad \boldsymbol{P}_\pi = \frac{\boldsymbol{b}\boldsymbol{b}^\mathsf{T}}{\|\boldsymbol{b}\|^2}$$

Inner Product and Projection

- We project $\boldsymbol{x}$ onto $\boldsymbol{b}$, and let $\pi_{\boldsymbol{b}}(\boldsymbol{x})$ be the projected vector.
- Question. Understanding the inner project $\langle \boldsymbol{x}, \boldsymbol{b} \rangle$ from the projection perspective?

$$\langle \boldsymbol{x}, \boldsymbol{b} \rangle = \|\pi_{\boldsymbol{b}}(\boldsymbol{x})\| \times \|\boldsymbol{b}\|$$

- In other words, the inner product of $\boldsymbol{x}$ and $\boldsymbol{b}$ is the product of (length of the projection of $\boldsymbol{x}$ onto $\boldsymbol{b}$) $\times$ (length of $\boldsymbol{b}$)



(a) Projection of $\boldsymbol{x} \in \mathbb{R}^2$ onto a subspace U with basis vector $\boldsymbol{b}$.

Example

- $\boldsymbol{b} = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$

$$\boldsymbol{P}_\pi = \frac{\boldsymbol{b}\boldsymbol{b}^\top}{\|\boldsymbol{b}\|^2} = \frac{1}{9}\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 2 \end{pmatrix} = \frac{1}{9}\begin{pmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{pmatrix}$$

For $\boldsymbol{x} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$,

$$\pi_U(\boldsymbol{x}) = \boldsymbol{P}_\pi \boldsymbol{x} = \frac{1}{9}\begin{pmatrix} 1 & 2 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \frac{1}{9}\begin{pmatrix} 5 \\ 10 \\ 10 \end{pmatrix} \in \text{span}[\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}]$$

Projection onto General Subspaces

- $\mathbb{R}^n \to$ 1-Dim
- A basis vector $\boldsymbol{b}$ in 1D subspace

$$\pi_U(\boldsymbol{x}) = \frac{\boldsymbol{b}\boldsymbol{b}^\top \boldsymbol{x}}{\boldsymbol{b}^\top \boldsymbol{b}}, \quad \lambda = \frac{\boldsymbol{b}^\top \boldsymbol{x}}{\boldsymbol{b}^\top \boldsymbol{b}}$$

$$\boldsymbol{P}_\pi = \frac{\boldsymbol{b}\boldsymbol{b}^\top}{\boldsymbol{b}^\top \boldsymbol{b}}$$

- $\mathbb{R}^n \to m$-Dim, $(m < n)$
- A basis matrix $B = \left(\boldsymbol{b}_1, \cdots, \boldsymbol{b}_m\right) \in \mathbb{R}^{n \times m}$

$$\pi_U(\boldsymbol{x}) = \boldsymbol{B}(\boldsymbol{B}^\top \boldsymbol{B})^{-1} \boldsymbol{B}^\top \boldsymbol{x}, \ \boldsymbol{\lambda} = (\boldsymbol{B}^\top \boldsymbol{B})^{-1} \boldsymbol{B}^\top \boldsymbol{x}$$

$$\boldsymbol{P}_\pi = \boldsymbol{B}(\boldsymbol{B}^\top \boldsymbol{B})^{-1} \boldsymbol{B}^\top$$

- $\lambda \in \mathbb{R}^1$ and $\boldsymbol{\lambda} \in \mathbb{R}^m$ are the coordinates in the projected spaces, respectively.
- $(\boldsymbol{B}^\top \boldsymbol{B})^{-1} \boldsymbol{B}^\top$ is called pseudo-inverse.
- How to derive is analogous to the case of 1-D lines (see pp. 71).

Example: Projection onto 2D Subspace

- $U = \text{span}[\begin{pmatrix}1\\1\\1\end{pmatrix}, \begin{pmatrix}0\\1\\2\end{pmatrix}] \subset \mathbb{R}^3$ and $\boldsymbol{x} = \begin{pmatrix}6\\0\\0\end{pmatrix}$. Check that $\{\begin{pmatrix}1 & 1 & 1\end{pmatrix}^\top, \begin{pmatrix}0 & 1 & 2\end{pmatrix}^\top\}$ is a basis.
- Let $\boldsymbol{B} = \begin{pmatrix}1 & 0\\1 & 1\\1 & 2\end{pmatrix}$. Then, $\boldsymbol{B}^\top\boldsymbol{B} = \begin{pmatrix}1 & 1 & 2\\0 & 1 & 2\end{pmatrix}\begin{pmatrix}1 & 0\\1 & 1\\1 & 2\end{pmatrix} = \begin{pmatrix}3 & 3\\3 & 5\end{pmatrix}$
- Can see that $\boldsymbol{P}_\pi = \boldsymbol{B}(\boldsymbol{B}^\top\boldsymbol{B})^{-1}\boldsymbol{B}^\top = \frac{1}{6}\begin{pmatrix}5 & 2 & -1\\2 & 2 & 2\\-1 & 2 & 5\end{pmatrix}$, and

  $\pi_U(\boldsymbol{x}) = \frac{1}{6}\begin{pmatrix}5 & 2 & -1\\2 & 2 & 2\\-1 & 2 & 5\end{pmatrix}\begin{pmatrix}6\\0\\0\end{pmatrix} = \begin{pmatrix}5\\2\\-1\end{pmatrix}$