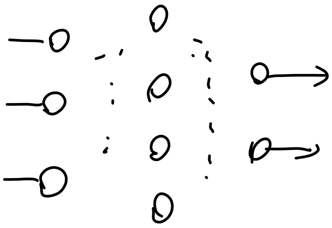
Lec2 Analytic Geometry 1: Norm, Inner Product and Positive Definite

Data Set
(Input, Output)



$\Rightarrow$ predict "unseen" data

$\Rightarrow$ prediction closed to the truth

$\Rightarrow$ ground-truth y, prediction $\hat{y}$

$\|y-\hat{y}\|$ as small as possible

norm

1. Norms

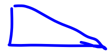
Def. A norm on a vector space V is a function

$$\|\cdot\|: V \longmapsto \mathbb{R}$$

such that for any $\lambda \in \mathbb{R}$, and $\vec{x}, \vec{y} \in V$, the following hold.

(1) $\|\lambda \vec{x}\| = |\lambda| \, \|\vec{x}\|$: absolute homogeneous

(2) $\|\vec{x}+\vec{y}\| \le \|\vec{x}\| + \|\vec{y}\|$ triangle inequality

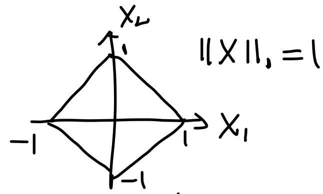


(3) $\|\vec{x}\| \ge 0$ and $\|\vec{x}\| = 0 \iff \vec{x} = \vec{0}$ positive definitness

iff.

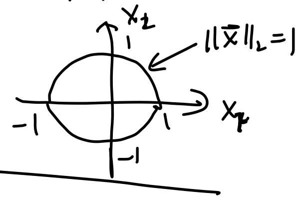
Example: (1) ℓ_1 norm (Manhattan Norm)

For $\vec{x} \in \mathbb{R}^n$, $\|x\|_1 := \sum_{i=1}^{n} |x_i|$



(2) ℓ_2 norm (Euclidean Norm)

For $\vec{x} \in \mathbb{R}^n$, $\|\vec{x}\|_2 := \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{\vec{x}^T \vec{x}}$



Exercise: prove: consider vector $\vec{x} = [x_1, x_2, \ldots, x_n]^T \in \mathbb{R}^n$

showing $\|x\|_\infty := \max(|x_1|, |x_2|, \ldots, |x_n|)$ is a norm

proof: (cond. 1) consider $\lambda \in \mathbb{R}$, $\|\lambda x\|_\infty = \max(|\lambda x_1|, |\lambda x_2|, \ldots |\lambda x_n|)$

$= \max(|\lambda| |x_1|, \ldots |\lambda| |x_n|)$, $|\cdot|$ is a norm

$= |\lambda| \max(|x_1|, \ldots |x_n|)$ ①

(cond. 2) consider $\vec{x} = [x_1, \ldots, x_n]^T$ $\vec{y} = [y_1, \ldots y_n]^T$

left: $\|\vec{x}+\vec{y}\|_\infty = \max(|x_1+y_1|, \ldots, |x_n+y_n|)$

right: $\|\vec{x}\|_\infty + \|\vec{y}\|_\infty = \max(|x_1|, \ldots |x_n|) + \max(|y_1|, \ldots, |y_n|)$

since for any n, $|x_n+y_n| \le |x_n| + |y_n|$, m, s.t.

$\|\vec{x}+\vec{y}\| = \max(|x_1+y_1|, \ldots, |x_n+y_n|) := |x_m+y_m| \le |x_m| + |y_m|$ ①

$\le \max(|x_1| \ldots |x_n|) + \max(|y_1| \ldots |y_n|)$

cond 3. obvious ①

2. (General) Inner Product

1) Dot product: for any $\vec{x}, \vec{y} \in \mathbb{R}^n$, $\vec{x}^T \vec{y} = \sum_{i=1}^{n} x_i y_i$

2) Bilinear Mapping f

Def. Give a vector space V. For all $\vec{x}, \vec{y} \in V$, $\psi, \lambda \in \mathbb{R}$,

$f(\cdot,\cdot): V \times V \mapsto \mathbb{R}$ is a bilinear map if.

(1) $f(\lambda\vec{x}+\psi\vec{y}, \vec{z}) = \lambda f(\vec{x},\vec{z}) + \psi f(\vec{y},\vec{z})$ (linear in 1st argument)

(2) $f(\vec{x}, \lambda\vec{y}+\psi\vec{z}) = \lambda f(\vec{x},\vec{y}) + \psi f(\vec{x},\vec{z})$ (······ 2nd argument)

3) Symmetric

$\vec{x} \times \vec{y} \neq \vec{y} \times \vec{x}$

Let V be a vector space and $f: V \times V \mapsto \mathbb{R}$ be a bilinear map. Then f is symmetric if $f(\vec{x},\vec{y}) = f(\vec{y},\vec{x})$

4) Positive definite.

Let V be a vector space, and $f: V \times V \mapsto \mathbb{R}$ be a bilinear map. Then f is positive definite if $\forall \vec{x} \in V \setminus \{0\}$ we have $f(\vec{x},\vec{x}) > 0$ and $f(\vec{0},\vec{0}) = 0$ (always non-negative)

5. General Inner Product

Def: Consider a vecter space V and a mapping $f: V \times V \to \mathbb{R}$.

f is called an inner product

① f is bilinear; ② f is symetric; ③ f is positive definite.

denote such a functino $f(\vec{x},\vec{y})$ as $\langle \vec{x},\vec{y} \rangle$

6. Property of (general) inner product.

Consider a vector space V. inner product $\langle \cdot \rangle: V \times V \mapsto \mathbb{R}$ and an ordered basis $B = [\vec{b}_1, \dots \vec{b}_n]$ matrix

For any $\vec{x}, \vec{y} \in V$, with

$\vec{x} = \sum_{i=1}^{n} \psi_i \vec{b}_i$, $\vec{y} = \sum_{j=1}^{n} \lambda_j \vec{b}_j$, with suitable $\psi_i, \lambda_i \in \mathbb{R}$

Then $\langle \vec{x}, \vec{y} \rangle = \langle \sum_{i=1}^{n} \psi_i \vec{b_i}, \sum_{j=1}^{n} \lambda_j \vec{b_j} \rangle$

$= \sum_{i=1}^{n} \psi_i \langle \vec{b_i}, \sum_{j=1}^{n} \lambda_j \vec{b_j} \rangle$

$= \sum_{i=1}^{n} \psi_i \sum_{j=1}^{n} \lambda_j \langle \vec{b_i}, \vec{b_j} \rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} \psi_i \langle \vec{b_i}, \vec{b_j} \rangle \lambda_j = \hat{x}^T A \hat{y}$

with $\hat{x} := [\psi_1, \dots, \psi_n]^T$

$\hat{y} := [\lambda_1, \dots, \lambda_n]^T$, $\quad A_{n\times n} = \begin{bmatrix} \langle \vec{b_1}, \vec{b_1} \rangle & \cdots & \langle \vec{b_1}, \vec{b_n} \rangle \\ \langle \vec{b_2}, \vec{b_1} \rangle & \cdots & \langle \vec{b_2}, \vec{b_n} \rangle \\ \vdots & & \\ \langle \vec{b_n}, \vec{b_1} \rangle & \cdots & \langle \vec{b_n}, \vec{b_n} \rangle \end{bmatrix}$

A is symmetry

Exercise Consider $V = \mathbb{R}^2$ with inner product $\langle \cdot \rangle : V \times V \to \mathbb{R}$ and a basis $B = [\vec{q_1}, \vec{q_2}]$ of V where $\vec{q_1}$ $[1, 1]^T$, $\vec{q_2} = [1, -2]^T$. $\langle \vec{q_1}, \vec{q} \rangle$

$\vec{x} := 2\vec{q_1} + 3\vec{q_2}$, $\vec{y} := -\vec{q_1} + 2\vec{q_2}$, compute $\langle \vec{x}, \vec{y} \rangle$, A matrix?

Normal computation: $\langle \vec{x}, \vec{y} \rangle = \langle 2\vec{q_1} + 3\vec{q_2}, -\vec{q_1} + 2\vec{q_2} \rangle$

$= -2\langle \vec{q_1}, \vec{q_1} \rangle - 3\langle \vec{q_2}, \vec{q_1} \rangle + 4\langle \vec{q_1}, \vec{q_2} \rangle + 6\langle \vec{q_2}, \vec{q_2} \rangle$ ①

$\langle \vec{q_1}, \hat{q_1} \rangle = 2$, $\langle \vec{q_2}, \vec{q_1} \rangle = \langle \vec{q_1}, \vec{q_2} \rangle = 1 - 2 = -1$, $\langle \vec{q_2}, \vec{q_2} \rangle = 5$

$\Rightarrow$ ① $\langle \vec{x}, \vec{y} \rangle = -4 + 3 - 4 + 30 = 25$

$\hat{x} = [2, 3]^T \quad \hat{y} = [-1, 2]^T$

also $A = \begin{bmatrix} \langle \vec{q_1}, \vec{q_1} \rangle & \langle \vec{q_1}, \vec{q_2} \rangle \\ \langle \vec{q_2}, \vec{q_1} \rangle & \langle \vec{q_2}, \vec{q_2} \rangle \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 5 \end{bmatrix}$

$\langle \vec{x}, \vec{y} \rangle$

$= \hat{x}^T A y = [2 \;\; 3] \begin{bmatrix} 2 & -1 \\ -1 & 5 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix}$

$= [1 \;\; 13] \begin{bmatrix} -1 \\ 2 \end{bmatrix} = 25$

7. Symmetric and Positive Definite Matrix.

Recall the positive define of a norm (inner product), we should have $\forall \vec{x} \in V \setminus \{\vec{0}\}$, $\vec{x}^T A x > 0$, we call a symmetric matrix A satisfying this condition as positive definite matrix.

Moreover, if $\forall \vec{x} \in V \setminus \{\vec{0}\}$, $x^T A x \geq 0$, $\Rightarrow$ A is a symmetric positive semidefinite matrix

Proposition: For any $A \in \mathbb{R}^{n\times n}$, $A^T A$ is symmetric positive semidefinite.

proof: by def, for any $\vec{x} \in V$ $A\vec{x} \in V$ $\vec{x}^T A^T A \vec{x} \geq 0$, $M := A^T A$

$(A\vec{x})^T A \vec{x}$ $\quad \vec{x}^T M \vec{x} \geq 0$

Proposition 2. For any $M \in \mathbb{R}^{n\times n}$ that is symmetric, positive definite, iff there exists $A \in \mathbb{R}^{n\times n}$ with full rank, s.t. $M = A^T A$

Theorem 3. For a real-valued finite dimensional vector space V and an ordered basis B of V, it holds that $\langle\cdot,\cdot\rangle : V\times V \mapsto \mathbb{R}$ is an inner product if and only if there exists a symmetric, positive definite matrix $A \in \mathbb{R}^{n\times n}$, with

$$\langle \vec{x}, \vec{y}\rangle = \hat{x}^T A \hat{y}$$

where $\hat{x}, \hat{y}$ are the coordinate representation of $\vec{x}, \vec{y}$ w.r.t. B.

Fact: any inner product induce a norm

$\|\vec{x}\| = \sqrt{\langle\vec{x},\vec{x}\rangle} = \sqrt{\vec{x}^T M \vec{x}}$ for some sym. pos definite matrix

Thm Cauchy-Schwarz Inequality

For an inner product vector space $(V, \langle\cdot,\cdot\rangle)$, the induce norm $\|\cdot\|$ satisfies the Cauchy-Schwarz Inequality $|\langle\vec{x},\vec{y}\rangle| \le \|\vec{x}\|\|\vec{y}\|$

Proof: consider $\langle\vec{x},\vec{y}\rangle$ corresponds to a symmetric, positive definite matrix M and matrix N with $M = N^T N$

$\langle\vec{x},\vec{y}\rangle = \vec{x}^T M \vec{y} = \hat{\vec{x}}^T \hat{\vec{y}}$, with $\hat{\vec{x}} := N\vec{x}$, $\hat{\vec{y}} := N\vec{y}$

$$\|\vec{x}\| = \sqrt{\vec{x}^T M \vec{x}} = \sqrt{\vec{x}^T N^T N \vec{x}} = \sqrt{\hat{\vec{x}}^T \hat{\vec{x}}} = \|\hat{\vec{x}}\|_2$$

Hence, to prove $|\langle\vec{x},\vec{y}\rangle| \le \|\vec{x}\|\|\vec{y}\|$, is equivalent.

$$|\hat{x}^T\hat{y}| \le \|\hat{x}\|_2\|\hat{y}\|_2 \quad \cdots\cdots\cdots (1)$$

Consider $\hat{x} := [a_1, \ldots, a_n]^T$, $\hat{y} := [b_1, \ldots, b_n]^T$, we can prove (1) by proving $\left(\sum_{i=1}^n a_i b_i\right)^2 \le \left(\sum_{i=1}^n a_i^2\right)\left(\sum_{i=1}^n b_i^2\right)$

consider the following polynomial

$$f(x) = \left(\sum_{i=1}^n a_i^2\right)x^2 - 2\left(\sum_{i=1}^n a_i b_i\right)x + \sum_{i=1}^n b_i$$

$$= \sum_{i=1}^n (a_i^2 x^2 - 2a_i b_i + b_i^2) = \sum_{i=1}^n (a_i x - b_i)^2 \ge 0$$

meaning that discriminant of $f(x)$ is not positive.

$$\Rightarrow 4\left(\sum_{i=1}^n a_i b_i\right)^2 - 4\left(\sum_{i=1}^n a_i^2\right)\left(\sum_{i=1}^n b_i^2\right) \le 0.$$

Prove: Consider a positive definite matrix $M \in \mathbb{R}^{n\times n}$ and a vector space $V = \mathbb{R}^n$, prove $\vec{x} \in V$, $\|\vec{x}\|_M := \sqrt{\vec{x}^T M x}$ is a norm proof: (cond. 1) consider $\lambda \in \mathbb{R}$

$$\|\lambda \vec{x}\|_M = \sqrt{\lambda^2 \vec{x}^T M \vec{x}} = \sqrt{\lambda^2} \cdot \sqrt{\vec{x}^T M \vec{x}} = |\lambda| \, \|\vec{x}\|_M$$

(cond. 2) Consider $\vec{x}, \vec{y} \in V$

$$\|\vec{x}+\vec{y}\|_M \le \|\vec{x}\|_M + \|\vec{y}\|_M$$

$$\|\vec{x}+\vec{y}\|_M^2 \le (\|\vec{x}\|_M + \|\vec{y}\|_M)^2$$

$$= \|\vec{x}\|_M^2 + 2\|\vec{x}\|_M \|\vec{y}\|_M + \|\vec{y}\|_M^2$$

$$\|\vec{x}+\vec{y}\|_M^2 = (\vec{x}+\vec{y})^T M (\vec{x}+\vec{y})$$

$$= \vec{x}^T M \vec{x} + \vec{x}^T M \vec{y} + \vec{y}^T M x + \vec{y}^T M \vec{y}$$

$$= \vec{x}^T M \vec{x} + 2\vec{x}^T M \vec{y} + \vec{y}^T M \vec{y}$$

$$= \|\vec{x}\|_M^2 + \|\vec{y}\|_M^2 + 2\vec{x}^T M \vec{y} \quad ①$$

Let $M = N^T N$

w.l.o.g. $\vec{x}' := N\vec{x}$, $\vec{y}' = N\vec{y}$

$$\vec{x}^T M \vec{y} = \vec{x}'^T \vec{y}' \le \|\vec{x}'\|_2 \|\vec{y}'\|_2 \quad ②$$

$$② \Rightarrow ① \le \|\vec{x}\|_M^2 + \|\vec{y}\|_M^2 + 2\|\vec{x}'\| \|\vec{y}'\|$$

$$= \|\vec{x}\|_M^2 + \|\vec{y}\|_M^2 + 2\|\vec{x}\|_M \|\vec{y}\|_M = (\|x\|_M + \|y\|_M)^2.$$

(cond. 3) obvious

3. Distance

Consider an inner product space $(V, \langle \cdot \rangle)$. Then the distance between $\vec{x}$, and $\vec{y}$, with $\vec{x}, \vec{y} \in V$ is

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

Remark: inner product induced a norm, and distance between two vectors, distance is related to norm.

Remark: different norm corresponds to different distance

Exercise : Concept length

compute the length of a vector $\vec{x} = [1, 1]^T$ using

① Dot product $= \sum x_i y_i$

② $\langle \vec{x}, \vec{y} \rangle = x^T \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix} y = x_1 y_1 - \frac{1}{2}(x_1 y_2 + x_2 y_1) + x_2 y_2.$

Answer: (1) $\|\vec{x}\| = \sqrt{\vec{x}^T \vec{x}} = \sqrt{1^2 + 1^2} = \sqrt{2}$ Euclidean Space

(2) $\langle \vec{x}, \vec{x} \rangle = x^2 - x_1 x_2 + x_2^2 = 1 - 1 + 1 = 1 \Rightarrow \|\vec{x}\| = 1$

Remark: length of a vector is related to the norm selected, length may be different if norm is different

Def. Metric

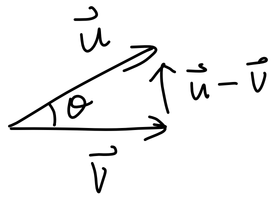
The mapping of $V \times V \mapsto \mathbb{R}$ for which $(\vec{x}, \vec{y})$ maps to $d(\vec{x}, \vec{y})$ is called a metric, which satisfies

1) positive definite: $d(\vec{x}, \vec{y}) \geq 0$ for any $\vec{x}, \vec{y} \in V$, $d(\vec{x}, \vec{y}) = 0$ iff $\vec{x} = \vec{y}$

2) symmetric: $d(\vec{x}, \vec{y}) = d(\vec{y}, \vec{x})$ for any $\vec{x}, \vec{y} \in V$

3) triangular inequality: $d(\vec{x}, \vec{z}) \leq d(\vec{x}, \vec{y}) + d(\vec{y}, \vec{z})$

Def. Angle



law of cosines

$$\|\vec{u} - \vec{v}\|^2 = \|\vec{u}\|^2 + \|\vec{v}\|^2 - 2\|\vec{u}\|\|\vec{v}\| \cos\theta \quad ①$$

note that $\langle \vec{u} - \vec{v}, \vec{u} - \vec{v} \rangle = \|\vec{u}\|^2 + \|\vec{v}\|^2 - \underline{2\langle \vec{u}, \vec{v} \rangle}$ ②

① ② $\Rightarrow \langle \vec{u}, \vec{v} \rangle = \|\vec{u}\|\|\vec{v}\| \cos\theta$

Assume that $\vec{x} \neq \vec{0}$, $\vec{y} \neq \vec{0}$, then by the Cauchy-Schwarz inequality

$$-1 \leq \frac{\langle \vec{x}, \vec{y} \rangle}{\|\vec{x}\|\|\vec{y}\|} \leq 1 \Longleftarrow \underline{|\langle \vec{x}, \vec{y} \rangle| \leq \|\vec{x}\|\|\vec{y}\|}$$

Thus, there exists a unique $\theta \in [0, \pi]$ s.t.

$$\boxed{\cos\theta = \frac{\langle \vec{x}, \vec{y} \rangle}{\|\vec{x}\|\|\vec{y}\|}} \Rightarrow \theta: \text{angle between } \vec{x} \text{ and } \vec{y}$$

In particular. $\vec{x} \perp \vec{y} \Rightarrow \langle \vec{x}, \vec{y} \rangle = 0 \Rightarrow \vec{x}$ and $\vec{y}$ and orthogonal.

Further more, if $\|\vec{x}\| = \|\vec{y}\| = 1$, then $\vec{x}, \vec{y}$ are both orthonormal

Def. Orthogonal Matrix

A square matrix $A \in \mathbb{R}^{n\times n}$ is orthogonal matrix iff its columns are orthonormal so that $A^TA = AA^T = I$ implies $A^{-1} = A^T$

$\Rightarrow$ inverse of such matrix can be obtained by simply transposing the matrix

$$A^TA = \begin{bmatrix} \vec{a}_1^T \\ \vec{a}_2^T \\ \vdots \\ \vec{a}_n^T \end{bmatrix} [\vec{a}_1 \cdots \vec{a}_n] = \begin{bmatrix} \vec{a}_1^T\vec{a}_1 & \vec{a}_1^T\vec{a}_2 & \cdots & \vec{a}_1^T\vec{a}_n \\ \vec{a}_2^T\vec{a}_1 & \vec{a}_2^T\vec{a}_2 & \cdots & \vec{a}_2^T\vec{a}_n \\ \vdots & & & \vdots \\ \vec{a}_n^T\vec{a}_1 & \cdots & \cdots & \vec{a}_n^T\vec{a}_n \end{bmatrix} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}$$

Remark: Transformation by orthogonal matrixes do NOT change the length of the vector

$$\|A\vec{x}\|^2 = \underbrace{\vec{x}^TA^T}_{(A\vec{x})^T}A\vec{x} = \vec{x}^TI\vec{x} = \vec{x}^T\vec{x} = \|\vec{x}\|^2$$

self learning: orthonormal basis (see 3.5), orthogonal complement (sec 3.6)

Def. Inner product of Function.

Given two functions $u, v: \mathbb{R} \mapsto \mathbb{R}$, the inner product of u and v is defined as $\langle u, v\rangle := \int_a^b u(x)v(x)dx$, with $a, b < \infty$

A generalization of inner product to infinite number of entries.

Remark: norm and orthogonality of a function (among functions) can also be defined in a similar way (sum $\to$ integral)

☆ Related Cocepts: convolution of two function.

$$u * v = \int_0^t f(\tau)\, g(t-\tau)d\tau.$$

↳ Convolution NN $\longrightarrow$ image recognition.

---

general inner product $\xrightarrow{\text{extended to function}}$ convolution

↳ CNN

positive definite matrix $\leftarrow$ general inner product

N^TN

general inner product $\downarrow$ induce

norm $\to$ length

$\|y - \hat{y}\|$

norm $\to$ distance (metric)

norm $\to$ angle $\to$ orthogonal $\to$ orthonormal (basis)

Sec. 3.1 ~ Sec 3.7