

Linear Algebra Overview

Vector: 1-dimensional array of numbers, e.g. $x \in \mathbb{R}^d$.

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix}$$

Vector norm: measures how "large" a vector is.

ℓ_p -norm: $\|x\|_p = \left(\sum_i |x_i|^p \right)^{1/p}$.

Ex. $p=2 \Rightarrow \|x\|_2 = \left(\sum_i x_i^2 \right)^{1/2}$ Euclidean norm

Dot product:

Given $u, v \in \mathbb{R}^d$, $\langle u, v \rangle = u^T v = \sum_i u_i v_i$.

$$\Rightarrow \|x\|_2 = \sqrt{\langle x, x \rangle} \Leftrightarrow \boxed{\|x\|_2^2 = \langle x, x \rangle}$$

Cosine similarity: $\cos(\theta) = \frac{\langle u, v \rangle}{\|u\|_2 \|v\|_2}$ ← dot product
← norm $\|u\|_2$

→ Orthogonality: u, v are orthogonal if $\boxed{\langle u, v \rangle = 0}$

$$\Rightarrow \|u - v\|_2^2 = \sum_{i=1}^d (u_i - v_i)^2$$

$$u = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, v = \begin{pmatrix} 1 \\ -1 \\ \vdots \\ -1 \end{pmatrix}$$

$$\|u - v\|_2^2 = (1-1)^2 + (1-(-1))^2 = 4$$

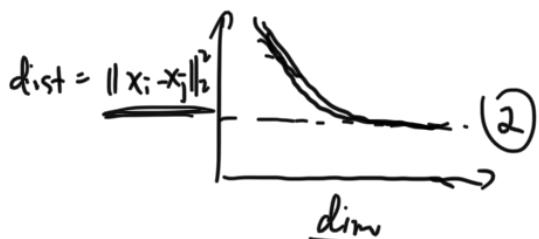
$$\begin{aligned} x &\in \mathbb{R}^d \quad (\text{hypercube}) \\ x_i &\in \left\{ +\frac{1}{\sqrt{d}}, -\frac{1}{\sqrt{d}} \right\} \\ \|x\|_2 &= \left(\sum_i x_i^2 \right)^{1/2} \\ &= \left(\sum_i \frac{1}{d} \right)^{1/2} = 1 \end{aligned}$$

Curse of dimensionality (HW1)

Given $u, v \in \mathbb{R}^d$, $\|u\|_2, \|v\|_2 = 1$,

what's the Euclidean distance between u, v if they are orthogonal?

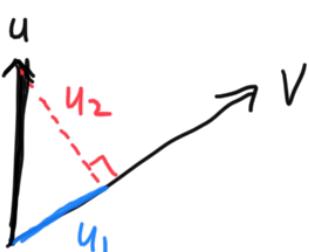
$$\begin{aligned} \|u - v\|_2^2 &= \langle u - v, u - v \rangle = \langle u, u \rangle - 2 \langle u, v \rangle + \langle v, v \rangle \\ &= \|u\|_2^2 + \|v\|_2^2 = 1 + 1 = 2 \end{aligned}$$



Vector projection

$$u = u_1 + u_2,$$

1



where $\langle u_1, u_2 \rangle = 0$,

$$u_1 = \frac{\langle u, v \rangle}{\|v\|_2} \cdot \underbrace{\frac{v}{\|v\|_2}}_{R}$$

$$A = \begin{bmatrix} A_{11} & & \\ & \ddots & \\ & & A_{nn} \end{bmatrix}$$

Matrix: 2-dimensional array of numbers, e.g. $A \in \mathbb{R}^{m \times n}$.

Multiplication: Given $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{n \times d}$,

$$\underline{AB} \in \mathbb{R}^{m \times d} \text{ where } [AB]_{ij} = \sum_{k=1}^n A_{ik} B_{kj}$$

Transpose: $[A^T]_{ij} = \underline{A}_{ji}$.

$$\text{Some properties: } -(AB)^T = B^T A^T.$$

$$- A \in \mathbb{R}^{d \times d} \text{ is } \underline{\text{symmetric}} \Leftrightarrow A = A^T$$

$$\begin{bmatrix} A_{11} & & & & \\ & \ddots & & & \\ & & A_{ii} & & \\ & & & \ddots & -A_{jj} \\ & & & & A_{jj} = A_{ii} \end{bmatrix}$$

Trace: Given $A \in \mathbb{R}^{d \times d}$, $\text{Tr}(A) = \sum_{i=1}^d \underline{A}_{ii} \in \mathbb{R}$

Cyclic property: $\text{Tr}(\underset{R}{ABC}) = \text{Tr}(CAB) = \text{Tr}(BCA)$

$$A = \underline{\frac{x x^T}{\mathbb{R}^{d \times d}}}, \underline{x \in \mathbb{R}^d} \quad \text{tr}(x x^T) = \text{tr}(\underbrace{x^T x}_{\langle x, x \rangle}) = \underline{\|x\|_2^2}$$

Inverse: define the identity matrix $I_d \in \mathbb{R}^{d \times d}$,
where $[I_d]_{i=j} = 1$ $[I_d]_{i \neq j} = 0$

$$\begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & \ddots \\ & & & 1 \end{bmatrix}$$

$$A I_d = A$$

A square matrix $A \in \mathbb{R}^{d \times d}$ is invertible if there exists

$A^{-1} \in \mathbb{R}^{d \times d}$ such that $\underline{AA^{-1} = I_d}$.

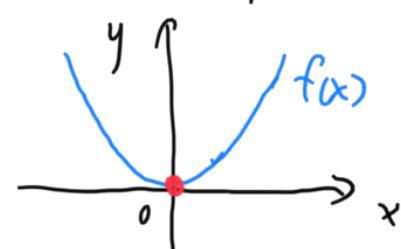
$$ab = 1 \Leftrightarrow b = a^{-1}. \quad AB = I \Rightarrow B = \underline{A^{-1}}$$

Basic Differentiation

Example: consider the function $f(x) = \frac{1}{2}x^2$, $x \in \mathbb{R}$.

Derivative: $f'(x) = x$.

- measures how "fast" the function is changing.



Question: How do we find the minimum of f ?

For convex & differentiable f ,

① find the derivative f'

② solve x^* s.t. $\underline{f'(x^*) = 0}$

What about the vector-valued setting?

$$\text{Example: } f(x) = \frac{1}{2} \underbrace{\langle x, Ax \rangle}_{x \in \mathbb{R}^d, A \in \mathbb{R}^{d \times d}}, \quad \underbrace{ax^2 + bx + c}_{b \in \mathbb{R}^d, c \in \mathbb{R}},$$

① find derivative.

$$f'(x) = \underline{Ax + b} \in \mathbb{R}^d \quad [\text{Matrix Cookbook}]$$

② Set f' to zero.

$$\begin{aligned} Ax &= -b \\ (A^T A)x &= -A^{-1}b \\ x &= -A^{-1}b \end{aligned}$$

Why is this a useful function in ML?

Linear regression: given data $X \in \mathbb{R}^{n \times d}$, labels $y \in \mathbb{R}^n$.

Goal: find $\underline{\theta \in \mathbb{R}^d}$ such that $\underline{X\theta \approx y}$.

$$\text{Squared loss: } \min_{\theta} L(\theta) = \sum_{i=1}^n (\langle x_i, \theta \rangle - y_i)^2$$

$$\begin{aligned} &= \|x\theta - y\|_2^2 \\ &= \langle x\theta - y, x\theta - y \rangle \\ &= \underbrace{\langle \theta, x^T x \theta \rangle}_A - \underbrace{\langle x^T y, \theta \rangle}_B + \underbrace{\langle y, y \rangle}_C \end{aligned}$$

$$\theta = \arg \min_{\theta} L(\theta)$$

$$= \underline{(x^T x)^{-1} x^T y} \quad \text{Least-squares solution}$$