CSED490Y: Optimization for Machine Learning

Week 02-1: Basics

Namhoon Lee

POSTECH

Spring 2022

Linear algebra

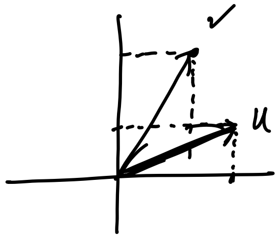
2

Vector $x \in \mathbb{R}^n$

x_i

- $x = (x_1, \ldots, x_n)$
- length and direction
- column vectors, row vectors

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

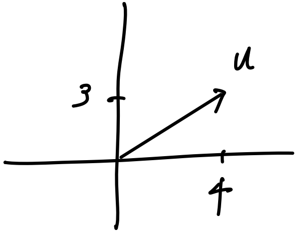


$$X^T = [x_1 \; x_2 \; \cdots \; x_n]$$

Linear algebra

p-norm of vector $x \in \mathbb{R}^n$ where $1 \leq p \leq \infty$:

$$\boxed{\|x\|_p} := \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}} .$$



$\|u\|_1 = 7$

$\|u\|_2 = 5$

$\|u\|_\infty = 4$

- $p = 1$ $\quad \|x\|_1 := |x_1| + \cdots + |x_n|$
- $p = 2$ $\quad \|x\|_2 := \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = (x^T x)^{\frac{1}{2}}$
- $p = \infty$

$\|x\|_\infty := \max_{i=\{1,\cdots,n\}} |x_i|$

Subspace

The subset $\mathcal{S} \subset \mathbb{R}^n$ is a subspace of $\mathbb{R}^n$ if the following property holds: If x and y are any two elements of $\mathcal{S}$, then

$$\boxed{\alpha x + \beta y} \in \mathcal{S}, \quad \forall \alpha, \beta \in \mathbb{R} \ .$$

(*i.e.*, set closed under addition and scaling)

Linear algebra

Span

$\{s_1, s_2, ..., s_k\}$ is a spanning set for $\mathcal{S}$ if any vector $s \in \mathcal{S}$ can be written as

$$s = \alpha_1 s_1 + \alpha_2 s_2 + ... + \alpha_k s_k \, ,$$

for some $\alpha_1, \alpha_2, ..., \alpha_k$.

Linear independence

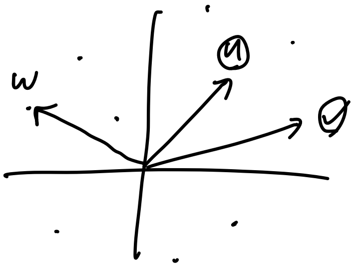
$\alpha_i x_i = -(\alpha_1 x_1 + \cdots + \alpha_k x_k)$

- A set of vectors $\{x_1, x_2, \ldots, x_k\} \in \mathbb{R}^n$ is called linearly independent if there are no $\alpha_1, \alpha_2, \ldots, \alpha_k \in \mathbb{R}$ such that

$$\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_k x_k = 0\,,$$

$\alpha u + \beta v = w \in \mathbb{R}^2$

except $\alpha_1 = \alpha_2 = \ldots = \alpha_k = 0$.



$\{u, v, w\}$ span $\mathbb{R}^2$

$\alpha_1 u + \alpha_2 v + \alpha_3 w = 0 \iff w = -\frac{1}{\alpha_3}(\alpha_1 u + \alpha_2 v)$

Linear algebra

Linear independence

A set of vectors $x_1, x_2, ..., x_k \in \mathbb{R}^n$ is called linearly independent if there are no $\alpha_1, \alpha_2, ..., \alpha_k \in \mathbb{R}$ such that

$$\alpha_1 x_1 + \alpha_2 x_2 + ... + \alpha_k x_k = 0 \ ,$$

except $\alpha_1 = \alpha_2 = ... = \alpha_k = 0$.

(*i.e.*, $x_1, x_2, ..., x_k$ are linearly independent if none of them can be written as a linear combination of the others.)

Linear algebra

$\{u, v, x\}$ span $\mathbb{R}^2$

$\{u, v\}$

Basis

If $\{x_1, \ldots, x_k\}$ are linearly independent & span $\mathcal{X}$, we call them a basis of $\mathcal{X}$

- k (the number of elements in the basis) is referred to as the dimension of $\mathcal{X}$, and denoted by $\dim(\mathcal{X})$.
- There are many ways to choose a basis of $\mathcal{X}$ in general, but that all bases contain the same nubmer of vectors.

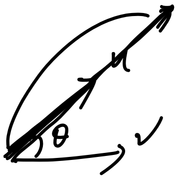
Linear algebra

$u, v \in \mathbb{R}^n$

$\begin{bmatrix} - & u^\top & - \end{bmatrix} \begin{bmatrix} | \\ v \\ | \end{bmatrix}$

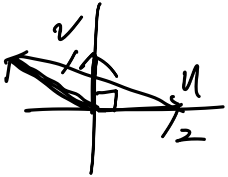
Inner product / dot product

$$u \cdot v = \langle u, v \rangle = u^\top v = \underline{\sum_{i=1}^{n} u_i v_i}$$



Angle between u and v

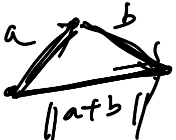
$$\cos\theta_{u,v} = \frac{\langle u, v\rangle}{\|u\|_2\|v\|_2}$$



- If they are perpendicular, $\langle u, v\rangle = 0$.

$$\|a+b\|_p \le \|a\|_p + \|b\|_p$$

$$u = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \qquad v = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$



$$\langle u, v\rangle = 0$$

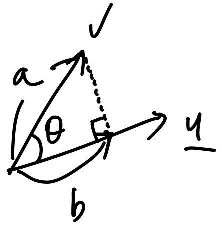
Linear algebra

$$\cos\theta = \frac{\langle u, v\rangle}{\|u\|_2 \|v\|_2} = \frac{b}{a}, \quad a = \|v\|_2$$

Projection of $\underline{v}$ onto $\underline{u}$

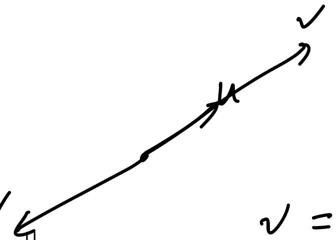
$$\boxed{v\ \cancel{u} = \frac{\langle v, u\rangle}{\|u\|_2^2} u} \quad v$$

$$b = \frac{\langle u, v\rangle}{\|u\|_2}$$



$$v = (b)\left(\frac{u}{\|u\|_2}\right)$$

Linear algebra

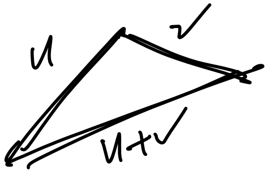


Cauchy-Schwarz inequality

$$|u^\top v| \leq \|u\|_2 \|v\|_2$$

$v = \alpha u$

- Two sides are equal iff u and v are linearly dependent.



Triangle inequality

$$\|u + v\|_2 \leq \|u\|_2 + \|v\|_2$$

Linear algebra

Outer product

$u, v \in \mathbb{R}^n$

$$u \otimes v = uv^\top = \left.\begin{array}{c} \overbrace{\hphantom{u_1v_1 \quad u_1v_2 \quad \cdots \quad u_1v_n}}^{n} \\ \begin{bmatrix} u_1 v_1 & u_1 v_2 & \dots & u_1 v_n \\ u_2 v_1 & u_2 v_2 & \dots & u_2 v_n \\ \vdots & \vdots & \ddots & \vdots \\ u_n v_1 & u_n v_2 & \dots & u_n v_n \end{bmatrix} \end{array}\right\} n$$

$\Leftrightarrow u^T v$

Matrix $A \in \underline{\mathbb{R}^{m\times n}}$

$A_{i,j}$

$$A = \begin{array}{cc} n & \\ \begin{bmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & \ddots & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{bmatrix} & m \end{array}$$

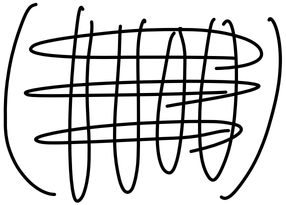
Some concepts to recall

- square matrix $\rightarrow$ $m = n$
- transpose of a matrix $\rightarrow$ $A^T \rightarrow$ $n \times m$ matrix where $(A^T)_{ij} = A_{ji}$
- symmetric matrix $\rightarrow$ $A_{ij} = A_{ji}$

Linear algebra

non square

- ✓ Null $(A) = \{x \in \mathbb{R}^n : Ax = 0\}$
- ✓ Range $(A) = \{y \in \mathbb{R}^n : Ax = y \text{ for some } x\}$ $\in \mathbb{R}^n$
- ✓ Rank $(A) =$ dimension of span of columns/rows of A

$A =$ 

Linear algebra

If A is $n \times n$, Rank $(A) = n$ iff

✓ ▶ det $(A) \neq 0$ $\prod_i \lambda_i$

✓ ▶ Null $(A) = \{0\}$

✓ ▶ Range $(A) = \mathbb{R}^n$

* For a matrix $A \in \mathbb{R}^{n\times n}$, an eigenvalue λ and eigenvector v are those that satisfy

$$\boxed{Av = \lambda v}.$$

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \qquad v = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \lambda = 2$$

Linear algebra

For a symmetric matrix A,

- All eigenvalues are real.
- All eigenvectors are perpendicular to each other.

If A is nonsingular, none of its eigenvalues are zero.

$\hookrightarrow \det A = 0$

singular

$\|v_i\| = 1$

For a symmetric matrix A, eigen or spectral decomposition

$$A = \sum_{i=1}^{n} \lambda_i v_i v_i^\top \,, \quad \in \mathbb{R}^n$$

or

$$A = Q\Lambda Q^\top \,,$$

$$Q = \begin{bmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{bmatrix}$$

using matrix forms.

$$\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$$

$$= \begin{bmatrix} \lambda_1 & & & \text{—} \\ & \lambda_2 & & \text{—} \\ | & & \ddots & \\ & \text{—} & & \lambda_n \end{bmatrix}$$

semi

Postive definite matrix

$\lambda_i \cancel{\geq} 0$

- A symmetric matrix A is called positive se~~mi~~definite, if all eigenvalues are greater than or equal to 0.
- Or

$$x^\top A x \cancel{\geq} 0 \ , \quad \forall x \in \mathbb{R}^n$$

$x \neq 0$

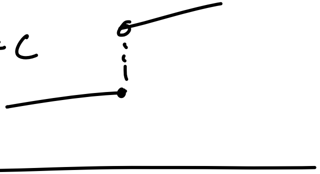
- $AA^\top$ and $A^\top A$ are always psd.

Calculus

$f(x) = x^TQx + q^Tx + c$

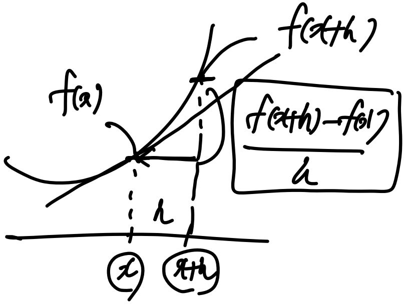
$\nabla f(x) = 2Qx + q$

$\nabla^2 f(x) = 2Q$



- Continuity
- Lipschitz continuity
- Derivative
- Gradient
- Hessian
- Quadratic function

$|f(u) - f(v)| \le L\|u - v\|_2$



$$\boxed{\frac{\partial f(x)}{\partial x}} := \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}$$

$$\frac{\partial f(x = x_1, \cdots, x_n)}{\partial x} = \begin{bmatrix} \vdots \\ \frac{\partial f(x)}{\partial x_i} \\ \vdots \end{bmatrix}$$

$\boxed{\nabla f(x)}$

$$\begin{bmatrix} \cdots & & \\ \vdots & \frac{\partial^2 f(x)}{\partial x_i \partial x_j} & \\ & & \end{bmatrix} \in \mathbb{R}^{n\times n}$$

Thank you

Any questions?

Credits

A lot of material in this course is borrowed or derived from the following:

- Numerical Optimization, Jorge Nocedal and Stephen J. Wright.
- Convex Optimization, Stephen Boyd and Lieven Vandenberghe.
- Convex Optimization, Ryan Tibshirani.
- Optimization for Machine Learning, Martin Jaggi and Nicolas Flammarion.
- Optimization Algorithms, Constantine Caramanis.
- Advanced Machine Learning, Mark Schmidt.