Vector and Matrix Linear Algebra

Introduction to Optimization for Machine Learning
M1 MLSD/AMSD

October 17, 2023

Roadmap

Roadmap

(1) Systems of Linear Equations

(2) Matrices

(3) Solving Systems of Linear Equations

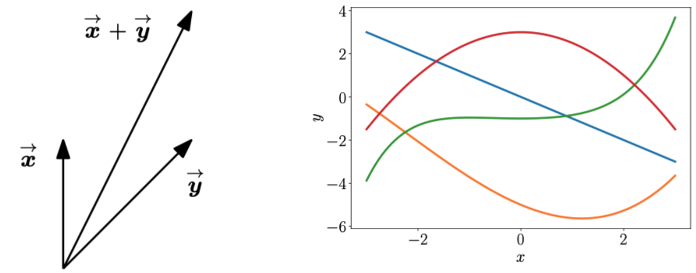
(4) Vector Spaces

(5) Linear Independence

(6) Basis and Rank

Linear Algebra

- Algebra: a set of objects and a set of rules or operations to manipulate those objects
- Linear algebra
  - Object: vectors $\boldsymbol{v}$
  - Operations: their additions ($\boldsymbol{v} + \boldsymbol{w}$) and scalar multiplication ($k\boldsymbol{v}$)
- Examples
  - Geometric vectors
  - Polynomials
  - Audio signals
  - Elements of $\mathbb{R}^n$



(a) Geometric vectors. (b) Polynomials.

System of Linear Equations

- For unknown variables $(x_1, \cdots, x_n) \in \mathbb{R}^n$,

$$a_{11}x_1 + \cdots + a_{1n}x_n = b_1$$
$$\vdots$$
$$a_{m1}x_1 + \cdots + a_{mn}x_n = b_m$$

- Three cases of solutions

- No solution

$$\begin{aligned} x_1 + x_2 + x_3 &= 3 \\ x_1 - x_2 + 2x_3 &= 2 \\ 2x_1 \phantom{{}- x_2} + 3x_3 &= 1 \end{aligned}$$

- Unique solution

$$\begin{aligned} x_1 + x_2 + x_3 &= 3 \\ x_1 - x_2 + 2x_3 &= 2 \\ x_2 + 3x_3 &= 1 \end{aligned}$$

- Infinitely many solutions

$$\begin{aligned} x_1 + x_2 + x_3 &= 3 \\ x_1 - x_2 + 2x_3 &= 2 \\ 2x_1 \phantom{{}- x_2} + 3x_3 &= 5 \end{aligned}$$

Matrix Representation

- A collection of linear equations

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

- Matrix representations:

$$\begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} x_1 + \cdots + \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} x_n = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \iff \underbrace{\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}}_{\boldsymbol{A}} \underbrace{\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}}_{\boldsymbol{x}} = \underbrace{\begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}}_{\boldsymbol{b}}$$

- Understanding $\boldsymbol{A}$ is the key to answering various questions about this linear system $\boldsymbol{Ax} = \boldsymbol{b}$.

Roadmap

(1) Systems of Linear Equations

(2) Matrices

(3) Solving Systems of Linear Equations

(4) Vector Spaces

(5) Linear Independence

(6) Basis and Rank

Matrix: Addition and Multiplication

- For two matrices $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{B} \in \mathbb{R}^{m \times n}$,

$$\boldsymbol{A} + \boldsymbol{B} := \begin{pmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{pmatrix} \in \mathbb{R}^{m \times n}$$

- For two matrices $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{B} \in \mathbb{R}^{n \times k}$, the elements c_{ij} of the product $\boldsymbol{C} = \boldsymbol{AB} \in \mathbb{R}^{m \times k}$ is:

$$c_{ij} = \sum_{l=1}^{n} a_{il} b_{lj}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, k.$$

- Example. $\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \end{pmatrix}$, compute $\boldsymbol{AB}$ and $\boldsymbol{BA}$.

Identity Matrix and Matrix Properties

- A square matrix[1] $\boldsymbol{I}_n$ with $I_{ii} = 1$ and $I_{ij=0}$ for $i \neq j$, where n is the number of rows and columns. For example,

$$\boldsymbol{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \boldsymbol{I}_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

- Associativity: For $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{B} \in \mathbb{R}^{n \times p}$, $\boldsymbol{C} \in \mathbb{R}^{p \times q}$, $(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$
- Distributivity: For $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m \times n}$, and $\boldsymbol{C}, \boldsymbol{D} \in \mathbb{R}^{n \times p}$,

  (i) $(\boldsymbol{A} + \boldsymbol{B})\boldsymbol{C} = \boldsymbol{AC} + \boldsymbol{BC}$ and (ii) $\boldsymbol{A}(\boldsymbol{C} + \boldsymbol{D}) = \boldsymbol{AC} + \boldsymbol{AD}$
- Multiplication with the identity matrix: For $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{I}_m\boldsymbol{A} = \boldsymbol{A}\boldsymbol{I}_n = \boldsymbol{A}$

[1] # of rows = # of cols

Inverse and Transpose

- For a square matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\boldsymbol{B}$ is the inverse of A, denoted by $\boldsymbol{A}^{-1}$, if

$$\boldsymbol{AB} = \boldsymbol{I}_n = \boldsymbol{BA}.$$

- Called regular/invertible/nonsingular, if it exists.
- If it exists, it is unique.
- How to compute? For 2×2 matrix,

$$\boldsymbol{A}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$$

- For a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{B} \in \mathbb{R}^{n\times m}$ with $b_{ij} = a_{ji}$ is the transpose of $\boldsymbol{A}$, which we denote by $\boldsymbol{A}^{\mathsf{T}}$.
- Example. For $\boldsymbol{A} = \begin{pmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \end{pmatrix}$,

$$\boldsymbol{A}^{\mathsf{T}} = \begin{pmatrix} 0 & 1 & 0 \\ 2 & -1 & 1 \end{pmatrix}$$

- If $\boldsymbol{A} = \boldsymbol{A}^{\mathsf{T}}$, $\boldsymbol{A}$ is called symmetric.

Inverse and Transpose: More Properties

- $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I} = \boldsymbol{A}^{-1}\boldsymbol{A}$
- $(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$
- $(\boldsymbol{A} + \boldsymbol{B})^{-1} \neq \boldsymbol{A}^{-1} + \boldsymbol{B}^{-1}$
- $(\boldsymbol{A}^{\mathsf{T}})^{\mathsf{T}} = \boldsymbol{A}$
- $(\boldsymbol{A} + \boldsymbol{B})^{\mathsf{T}} = \boldsymbol{A}^{\mathsf{T}} + \boldsymbol{B}^{\mathsf{T}}$
- $(\boldsymbol{A}\boldsymbol{B})^{\mathsf{T}} = \boldsymbol{B}^{\mathsf{T}}\boldsymbol{A}^{\mathsf{T}}$
- If $\boldsymbol{A}$ is invertible, so is $\boldsymbol{A}^{\mathsf{T}}$.

Scalar Multiplication

- Multiplication by a scalar $\lambda \in \mathbb{R}$ to $\boldsymbol{A} \in \mathbb{R}^{m \times n}$
- Example. For $\boldsymbol{A} = \begin{pmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \end{pmatrix}$, $3 \times \boldsymbol{A} = \begin{pmatrix} 0 & 6 \\ 3 & -3 \\ 0 & 3 \end{pmatrix}$
- Associativity
  - $(\lambda\psi)\boldsymbol{C} = \lambda(\psi\boldsymbol{C})$
  - $\lambda(\boldsymbol{BC}) = (\lambda\boldsymbol{B})\boldsymbol{C} = \boldsymbol{B}(\lambda\boldsymbol{C}) = (\boldsymbol{BC})\lambda$
  - $(\lambda\boldsymbol{C})^{\mathsf{T}} = \boldsymbol{C}^{\mathsf{T}}\lambda^{\mathsf{T}} = \boldsymbol{C}^{\mathsf{T}}\lambda = \lambda\boldsymbol{C}^{\mathsf{T}}$
- Distributivity
  - $(\lambda + \psi)\boldsymbol{C} = \lambda\boldsymbol{C} + \psi\boldsymbol{C}$
  - $\lambda(\boldsymbol{B} + \boldsymbol{C}) = \lambda\boldsymbol{B} + \lambda\boldsymbol{C}$

Roadmap

(1) Systems of Linear Equations

(2) Matrices

(3) Solving Systems of Linear Equations

(4) Vector Spaces

(5) Linear Independence

(6) Basis and Rank

Algorithms for Solving System of Linear Equations

1. Pseudo-inverse

   $$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \iff \boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^\top\boldsymbol{b} \iff \boldsymbol{x} = (\boldsymbol{A}^\top\boldsymbol{A})^{-1}\boldsymbol{A}^\top\boldsymbol{b}$$

   - $(\boldsymbol{A}^\top\boldsymbol{A})^{-1}\boldsymbol{A}^\top$: *Moore-Penrose pseudo-inverse*
   - many computations: matrix product, inverse, etc
2. Gaussian elimination
   - intuitive and constructive way
   - cubic complexity (in terms of # of simultaneous equations)
3. Iterative methods
   - practical ways to solve indirectly

   (a) stationary iterative methods: Richardson method, Jacobi method, Gaus-Seidel method, successive over-relaxation method

   (b) Krylov subspace methods: conjugate gradients, generalized minimal residual, biconjugate gradients

Roadmap

Vector Spaces

Definition. A real-valued vector space $V = (\mathcal{V}, +, \cdot)$ is a set $\mathcal{V}$ with two operations

$(a) \quad + : \mathcal{V} \times \mathcal{V} \mapsto \mathcal{V} \quad$ (vector addition)

$(b) \quad \cdot : \mathbb{R} \times \mathcal{V} \mapsto \mathcal{V} \quad$ (scalar multiplication),

where

1. Distributivity.
   - $\forall \lambda \in \mathbb{R}, \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V},\ \lambda \cdot (\boldsymbol{x} + \boldsymbol{y}) = \lambda \cdot \boldsymbol{x} + \lambda \boldsymbol{y}$
   - $\forall \lambda, \psi \in \mathbb{R}, \boldsymbol{x} \in \mathcal{V},\ (\lambda + \psi) \cdot \boldsymbol{x} = \lambda \cdot \boldsymbol{x} + \psi \cdot \boldsymbol{x}$
2. Associativity. $\forall \lambda, \psi \in \mathbb{R}, \boldsymbol{x} \in \boldsymbol{\mathcal{V}},\ \lambda \cdot (\psi \cdot \boldsymbol{x}) = (\lambda \psi) \cdot \boldsymbol{x}$
3. Neutral element. $\forall \boldsymbol{x} \in \mathcal{V},\ 1 \cdot \boldsymbol{x} = \boldsymbol{x}$

Example

- $\mathcal{V} = \mathbb{R}^n$ with
  - Vector addition: $\boldsymbol{x} + \boldsymbol{y} = (x_1 + y_1, \ldots, x_n + y_n)$
  - Scalar multiplication: $\lambda \boldsymbol{x} = (\lambda x_1, \ldots, \lambda x_n)$
- $\mathcal{V} = \mathbb{R}^{m \times n}$ with
  - Vector addition: $\boldsymbol{A} + \boldsymbol{B} = \begin{pmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{pmatrix}$
  - Scalar multiplication: $\lambda \boldsymbol{A} = \begin{pmatrix} \lambda a_{11} & \cdots & \lambda a_{1n} \\ \vdots & & \vdots \\ \lambda a_{m1} & \cdots & \lambda a_{mn} \end{pmatrix}$

Roadmap

Linear Independence

- Definition. For a vector space V and vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in V$, every $\boldsymbol{v} \in V$ of the form $\boldsymbol{v} = \lambda_1 \boldsymbol{x}_1 + \cdots + \lambda_k \boldsymbol{x}_k$ with $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ is a linear combination of the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n \in V$.
- Definition. If there is a non-trivial linear combination such that $\mathbf{0} = \sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i$ with at least one $\lambda_i \neq 0$, the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are linearly dependent. If only the trivial solution exists, i.e., $\lambda_1 = \ldots = \lambda_k = 0$, $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ are linearly independent.
- Meaning. A set of linearly independent vectors consists of vectors that have no redundancy.
- Useful fact. The vectors $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ are linearly dependent, iff (at least) one of them is a linear combination of the others.
    - $x - 2y = 2$ and $2x - 4y = 4$ are linearly dependent.

Rank (1)

- Definition. The rank of $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ denoted by $\text{rk}(\boldsymbol{A})$ is $\#$ of linearly independent columns
    - Same as the number of linearly independent rows
- $\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{pmatrix}$

Thus, $\text{rk}(\boldsymbol{A}) = 2$.

- $\text{rk}(\boldsymbol{A}) = \text{rk}(\boldsymbol{A}^{\mathsf{T}})$

Rank (2)

- The columns (resp. rows) of $\boldsymbol{A}$ span a subspace U (resp. W) with $\dim(U) = \text{rk}(\boldsymbol{A})$ (resp. $\dim(W) = \text{rk}(\boldsymbol{A})$), and a basis of U (resp. W) can be found by Gauss elimination of $\boldsymbol{A}$ (resp. $\boldsymbol{A}^{\mathsf{T}}$).

- For all $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, $\text{rk}(\boldsymbol{A}) = n$, iff $\boldsymbol{A}$ is regular (invertible).

- The linear system $\boldsymbol{Ax} = \boldsymbol{b}$ is solvable, iff $\text{rk}(\boldsymbol{A}) = \text{rk}(\boldsymbol{A}|\boldsymbol{b})$.

- For $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, the subspace of solutions for $\boldsymbol{Ax} = \boldsymbol{0}$ possesses dimension $n - \text{rk}(\boldsymbol{A})$.

- $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ has full rank if its rank equals the largest possible rank for a matrix of the same dimensions. The rank of the full-rank matrix $\boldsymbol{A}$ is min(# of cols, # of rows).