Linear Algebra

Mathematics for Machine Learning

San Diego Machine Learning
Liam Barstad

Systems of Linear Equations

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

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 2 \\ 2 & 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 2 & 5 \end{bmatrix}$$

$$\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} + \begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix} = \begin{bmatrix} a_1+b_1 & a_2+b_2 \\ a_3+b_3 & a_4+b_4 \end{bmatrix}$$

$$\lambda \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} = \begin{bmatrix} \lambda a_1 & \lambda a_2 \\ \lambda a_3 & \lambda a_4 \end{bmatrix}$$

Matrix Multiplication:

$A \in \boldsymbol{R}^{mxn}$ $i = 1,...m$
$B \in \boldsymbol{R}^{nxk}$ $j = 1,...k$

$AB := C \in \boldsymbol{R}^{mxk}$

$c_{ij} = \sum^{n}_{l=1} a_{il} b_{lj}$

$AB \neq BA$

Properties

Commutivity

$A*B = B*A$

Associativity

$(A*B)*C = A*(B*C)$

Distributivity

$A*(B \oplus C) = (A*B) \oplus (A*C)$

$(B \oplus C)*A = (B*A) \oplus (C*A)$

Closure

$\forall A, B \in S : A*B \in S$

Identity / Neutral Element

$\exists e \in S, \forall a \in S :$

$a*e = a$

Inverse

$\forall a \in S, \exists b \in S :$
$a*b = b*a = e$

Identity, Transpose, Inverse

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

A^T

A^{-1}

Identity Matrix

$\forall A \in \boldsymbol{R}^{mxn} : I_m A = A I_n = A$

Serves as "identity element" for matrices

Transpose

$A \in \boldsymbol{R}^{mxn}$, $B \in \boldsymbol{R}^{nxm}$,
$a_{ij} = b_{ji}$:
$B = A^T$

Flip a matrix, so its rows become its columns

$(A + B)^T = A^T + B^T$

$(AB)^T = B^T A^T$

To square each element in a matrix: AA^T

Inverse

$AA^{-1} = I = A^{-1}A$

A must be a square matrix

Not all matrices are invertible (aka regular, or nonsingular)

$(A + B)^{-1} \neq A^{-1} + B^{-1}$

$(AB)^{-1} = B^{-1}A^{-1}$

Solving Systems of Equations

Particular Solution

Rewrite system of equations in matrix form: Ax = b

(Inhomogeneous)

$$\left[\begin{array}{ccccc|c} -2 & 4 & -2 & -1 & 4 & -3 \\ 4 & -8 & 3 & -3 & 1 & 2 \\ 1 & -2 & 1 & -1 & 1 & 0 \\ 1 & -2 & 0 & -3 & 4 & -1 \end{array}\right]$$

Gaussian Elimination

Reduce to Row Echelon Form:

The first nonzero entry of each row is to the right (pivot)

Can swap, add, subtract, or multiply rows by scalar

$$\left[\begin{array}{ccccc|c} 1 & -2 & 1 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 3 & -2 \\ 0 & 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix}$$

$$x = \begin{bmatrix} 2 \\ 0 \\ -1 \\ 1 \\ 0 \end{bmatrix} + \lambda_1 \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 2 \\ 0 \\ -1 \\ 2 \\ 1 \end{bmatrix}$$

General Solution

Find Ax = 0 (homogeneous)

# of solutions = # of non-pivot columns

Basis vectors of null space

Groups

Consists of a set $\boldsymbol{G}$ and operation $\otimes$, denoted $(\boldsymbol{G}, \otimes)$

Attributes:

- Closure $\forall x, y \in \boldsymbol{G}: \quad x \otimes y \in \boldsymbol{G}$
- Associative $\forall x, y, z \in \boldsymbol{G}: \quad (x \otimes y) \otimes z = x \otimes (y \otimes z)$
- Neutral (Identity) $\exists e \in \boldsymbol{G}\ \forall x \in \boldsymbol{G}: \quad x \otimes e = e \otimes x = x$
- Inverse $\forall x \in \boldsymbol{G}\ \exists y \in \boldsymbol{G}: \quad x \otimes y = y \otimes x = e$

Abelian Groups:

- Commutative $\forall x, y \in \boldsymbol{G}: \quad x \otimes y = y \otimes x$

Vector Spaces

Vector space $\mathbf{V} = (V, +, \otimes)$

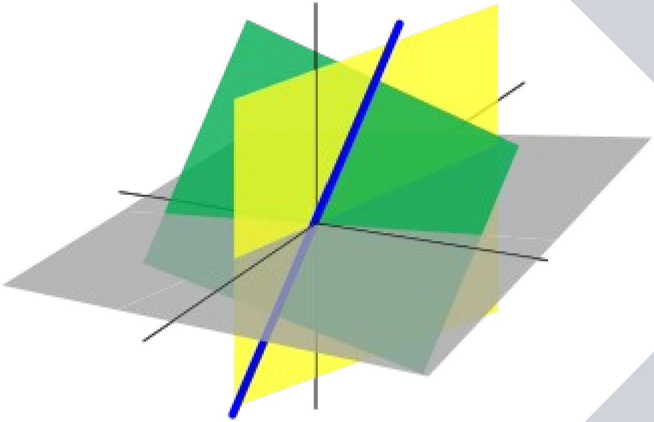
$v \in \mathbf{R}^n : v = [x_1, x_2, \ldots x_n]$ $\mathbf{R}^{nxm} = \mathbf{R}^{nm}$

$+ \rightarrow$ addition

$\otimes \rightarrow$ scalar multiplication (λv)

$(V, +) \rightarrow$ Abelian group (Closure, Associative, Neutral (0), Inverse, Commutative)

$\otimes$ is distributive over $+$: $\lambda(x + y) = \lambda x + \lambda y$



Multiplication:

- Array multiplication $c = ab,\ c_j = a_j b_j$
- Outer product $ab^T \in \mathbf{R}^{nxn}$
- Inner/Scalar/Dot product $a^T b \in \mathbf{R}$

W is a subspace of **V** if:

- $(W, +, \otimes)$ has closure under addition
- $0 \in W$

Vector Space Examples

- Euclidean space → $\mathbf{R}^n$
- Matrix space → $\mathbf{R}^{nxm}$
- Polynomial spaces:
  $p(x) = a_0 + a_1 x + a_2 x^2 + \ldots a_n x^n$
- Function spaces:
  - Continuous function spaces
  - Hilbert spaces
  - Fourier spaces
  - Sobolev spaces
  - Banach spaces
- Null space $Ax = 0$ is a subspace of $\mathbf{R}^n$, but the particular solutions for $Ax = b$ is not necessarily

Linear Independence

Linear combination:

$\lambda \in \mathbf{R}, v \in \mathbf{V}: \quad v = \sum^{k}_{i=1} \lambda_i x_i$

Linear independence:

$\sum^{k}_{i=1} \lambda_i x_i = 0 : \quad \lambda = \{0\}$

Find through Gaussian Elimination:

- Reduce to REF
- Linear independent IFF no nonpivot columns

$$\begin{bmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{bmatrix}$$

Basis

Generating set: $\mathbf{A} = \{x_1 \ldots x_n\} \subseteq \mathbf{V}$

Each vector in **A** is linearly independent, and every vector in **V** can be expressed as a linear combination of vectors in **A**

B is a basis of V IFF:

- $b \in \mathbf{B} : \mathbf{B} \neq 0$
- **B** is smallest generating set of **V**
- Adding any more vectors to **B** will make it linearly dependent

The canonical/standard basis of $\mathbf{R}^3$ is: $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Rank

$A \in \mathbf{R}^{nxm}$: rk(A) = Number of linearly independent columns (or rows) in A

Tells us the number of independent equations in a system

LoRa – Low-Rank Adaptation of Large Language Models

If a square matrix has full rank, $\det(A) \neq 0$

Linear Mappings

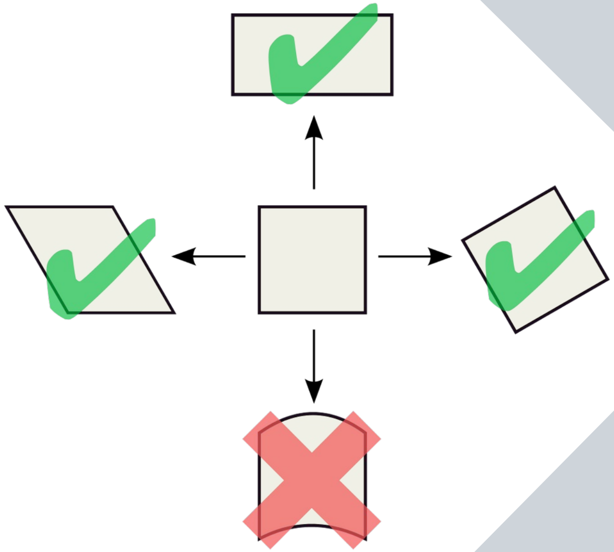
A mapping ϕ is linear (homomorphism) IFF it preserves:

- Vector addition (+) $\phi(x + y) = \phi(x) + \phi(y)$
- Scalar multiplication (⊗) $\phi(\lambda x) = \phi\lambda(x)$

In other words:

$\forall x, y \in \mathbf{V}\ \forall \lambda, \psi \in \mathbf{R} : \ \phi(\lambda x + \psi y) = \phi\lambda(x) + \phi\psi(y)$

- Injective → $\forall x, y \in \mathbf{V}, x = y : \ \phi(x) = \phi(y)$
  Each element has a 1 to 1 mapping from V to W
- Surjective → $\phi(\mathbf{V}) = \mathbf{W}$
  Every element in W can be reached by a mapping from V
- Bijective → Injective + Surjective
  Can be undone, has inverse ϕ^{-1}



Linear Mappings (cont.)

Types of Homomorphisms:

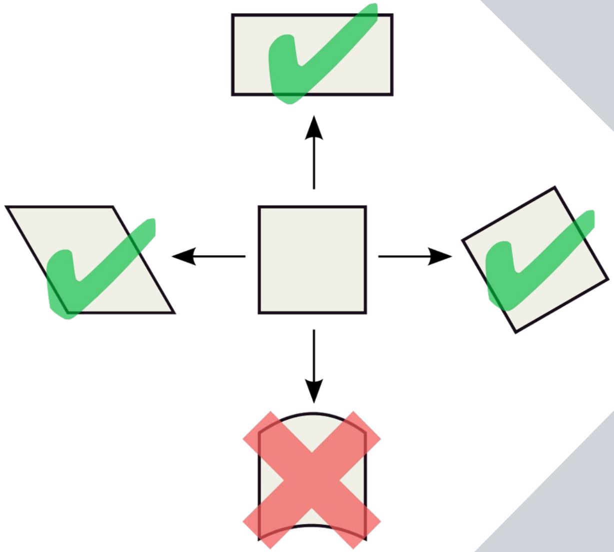
- Isomorphism: linear + bijective
  Theorem: vector spaces V and W are isomorphic IFF dim(V) = dim(W)
- Endomorphism: linear, **V** → **V**
- Automorphism: linear + bijective, **V** → **V**

Transformation matrix:

$A_\phi \in \mathbf{R}^{mxn}$, $B = (b_1 \ldots b_n)$, $x \in \mathbf{V}$, $y \in \mathbf{W}$:

$y = A_\phi x$

$$\begin{bmatrix} | & & | \\ \phi(b_1) & \ldots & \phi(b_n) \\ | & & | \end{bmatrix}$$



Basis Change

What happens to a linear mapping when you change the basis of V or W?

$\mathbf{V} \;\rightarrow\; \mathbf{B}_1 = (b_{11},\ldots b_{1n}) \;\rightarrow\; \mathbf{B}_2 = (b_{21},\ldots b_{2n})$

$S\mathbf{B}_1 = \mathbf{B}_2$

$\mathbf{W} \;\rightarrow\; \mathbf{C}_1 = (c_{11},\ldots c_{1n}) \;\rightarrow\; \mathbf{C}_2 = (c_{12},\ldots c_{2n})$

$T\mathbf{C}_1 = \mathbf{C}_2$

$A^1_\phi = T^{-1}A_\phi S$ $\quad A^1_\phi$ and A_ϕ are *equivalent*

$A_\phi, A^1_\phi, S \in \mathbf{R}^{nxn}$: $A^1_\phi = S^{-1}A_\phi S$ $\quad A^1_\phi$ and A_ϕ are *similar*

Image & Kernel

Kernel (Null Space)

Solutions to Ax = 0

$\phi(x) = 0_W$

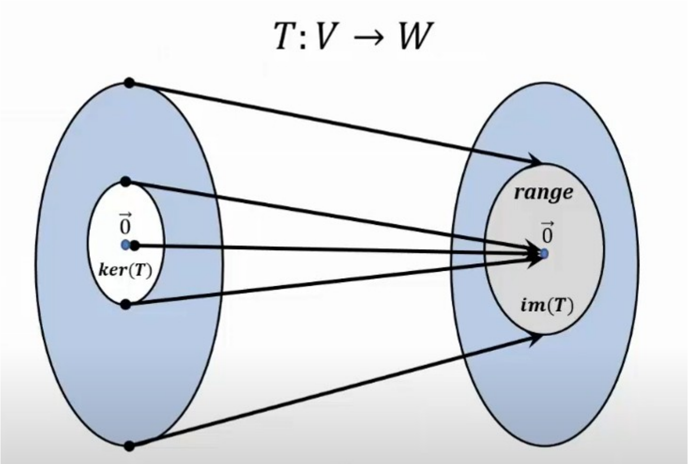


Image (Range)

Vectors in W that are reachable by the transformation

$\mathrm{Im}(\phi) = \phi(\mathbf{V})$

Rank-Nullity Theorem

$\dim(\ker(\phi)) + \dim(\mathrm{Im}(\phi)) = \dim(\mathbf{V})$

If $\dim(\mathbf{V}) = \dim(\mathbf{W})$, bijective

Affine Subspaces & Mappings

$\mathbf{V} = (V, +, \otimes)$, $x_0 \in \mathbf{V}$, $\mathbf{U} \subseteq \mathbf{V}$

$\mathbf{L} = x_0 + \mathbf{U} := \{ x_0 + u : u \in \mathbf{U} \}$

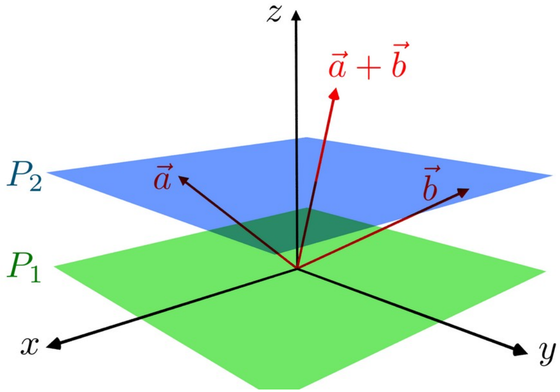
$= \{ v \in \mathbf{V} | \exists u \in \mathbf{U} : v = x_0 + u \} \subseteq \mathbf{V}$

x0 = support/translation point

U = direction/direction space

$\forall x \in \mathbf{L} : x = x_0 + \sum^{k}_{i=1} \lambda_k b_k$ Parametric equation of **L** : b_k = directional vectors (basis of **U**), λ_k = parameters

Solution to Ax = b is affine subspace of $\mathbf{R}^n$ of dimension n – rk(A)



Affine Mapping:

$\phi : V \rightarrow W$, $a \in W$

$a + \phi(x)$