UNIVERSITY OF MARYLAND

MSML610: Advanced Machine Learning

Linear Algebra

Instructor: GP Saggese, PhD - gsaggese@umd.edu

References:

Linear algebra

- **Linear algebra**
  - Vector and vector spaces
  - Affine spaces
  - Vectors and matrices
  - Linear functions
  - Connections between Machine Learning and Linear Algebra

Vector and vector spaces

- Linear algebra
    - **Vector and vector spaces**
    - Affine spaces
    - Vectors and matrices
    - Linear functions
    - Connections between Machine Learning and Linear Algebra

Field: definition

A field $\mathbb{F} = (X, +, *)$ is a set X with two binary operations + and *, satisfying the following 6 axioms:

1. Closed with respect to + and *:
$$a, b \in X \implies a + b \in X$$
$$a, b \in X \implies a * b \in X$$
2. Commutativity of + and *:
$$a + b = b + a$$
$$a * b = b * a$$
3. Associativity of + and *:
$$a + (b + c) = (a + b) + c = a + b + c$$
$$a * (b * c) = (a * b) * c = a * b * c$$
4. Distributivity of multiplication over addition:
$$a * (b + c) = a * b + a * c$$
5. Existence of + and * identity elements, 0 and 1:
$$a + 0 = a$$

Field: examples

- Examples
  - The set of $\mathbb{R}, \mathbb{C}$, GF(2)
  - The set of rational numbers $\mathbb{Q}$, i.e., numbers that can be written as fraction $\frac{a}{b}$ with $a, b \in \mathbb{Z}$ and $b \neq 0$
- Non-examples
  - The set of positive integers $\mathbb{N} = 1, 2, 3, \ldots$ is not a field
  - The set of integers $\mathbb{Z} =$ . . . , -2, -1, 0, 1, 2, . . . $ is not a field

Vector space: definition

- A "vector space $\mathcal{V}$ over a field $\mathbb{F}$" is a triple $(\mathcal{V}, \mathbb{F}, +, \cdot)$ where:
  - $\mathcal{V}$ is a set of vectors
  - $\mathbb{F}$ is a field of scalars
  - $+$ is a sum operation between vectors
  - $\cdot$ is a scalar multiplication
- A vector space needs to satisfy the following 2 properties:
  1. Closed with respect to to scalar multiplication: if $\underline{\boldsymbol{x}} \in \mathcal{V}$, then $\alpha \cdot \underline{\boldsymbol{x}} \in \mathcal{V}$
  2. Closed with respect to to vector addition: if $\underline{\boldsymbol{x}}, \underline{\boldsymbol{y}} \in \mathcal{V}$, then $\underline{\boldsymbol{x}} + \underline{\boldsymbol{y}} \in \mathcal{V}$

Linear combination of vectors

- The vector
$$\alpha_1 \underline{\boldsymbol{v}}_1 + \alpha_2 \underline{\boldsymbol{v}}_2 + ... + \alpha_n \underline{\boldsymbol{v}}_n$$
is a linear combination of vectors $\underline{\boldsymbol{v}}_1, ..., \underline{\boldsymbol{v}}_n$ with coefficients $\alpha_1, ..., \alpha_n$
- A linear combination can be written in matrix form:
$$\underline{\underline{\boldsymbol{V}}} \cdot \underline{\boldsymbol{\alpha}} \text{ or } \underline{\boldsymbol{\alpha}}^T \cdot \underline{\underline{\boldsymbol{V}}}^T$$
where $\underline{\underline{\boldsymbol{V}}} = (\underline{\boldsymbol{v}}_1 | ... | \underline{\boldsymbol{v}}_n)$

Span of vectors

- The span of n m-dimensional vectors is the set of all linear combinations of the n vectors:

$$\text{Span}(\underline{\boldsymbol{v}}_1, ..., \underline{\boldsymbol{v}}_n) = \{\underline{\underline{\boldsymbol{V}}} \cdot \underline{\boldsymbol{\alpha}} \text{ with } \underline{\boldsymbol{\alpha}} \in \mathbb{F}^n\} = \{\underline{\boldsymbol{v}} \in \mathbb{F}^m : \underline{\boldsymbol{v}} = \sum_{i=1}^{n} \alpha_i \underline{\boldsymbol{v}}_i\}$$

- E.g., the span of vectors is a vector space

Null space of a matrix

- Null space of the columns of a matrix $\underline{\underline{\boldsymbol{A}}}$ is defined as the set:

$$\text{Null}(\underline{\underline{\boldsymbol{A}}}) = \{\underline{\boldsymbol{v}} : \underline{\underline{\boldsymbol{A}}} \cdot \underline{\boldsymbol{v}} = \underline{\boldsymbol{0}}\}$$

  - In words, all the vectors that are coefficients of linear combinations of columns of $\underline{\underline{\boldsymbol{A}}}$ yielding the zero vector
- Null space is a vector space

Homogeneous linear system associated with null space

- From the definition of matrix-vector multiplication the vector $\underline{\boldsymbol{v}}$ is in $\text{Null}(\underline{\underline{\boldsymbol{A}}}) \iff \underline{\boldsymbol{v}}$ is a solution of the homogeneous linear system involving the columns of $\underline{\underline{\boldsymbol{A}}}$:

$$\underline{\boldsymbol{a}}_1^T \cdot \underline{\boldsymbol{v}} = 0$$
$$\underline{\boldsymbol{a}}_2^T \cdot \underline{\boldsymbol{v}} = 0$$
$$\dots$$
$$\underline{\boldsymbol{a}}_m^T \cdot \underline{\boldsymbol{v}} = 0$$

- Note that the notation is a bit confusing since we mean the transpose of the columns $\underline{\boldsymbol{a}}_i$ of $\underline{\underline{\boldsymbol{A}}}$ and not the rows of $\underline{\underline{\boldsymbol{A}}}$

Dot product on a vector space: definition

- Given a field of scalars $\mathbb{F}$ and a vector space $\mathcal{V}$ over $\mathbb{F}$, an inner product is a mapping:

$$\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \to \mathbb{F}$$

that satisfies the 3 axioms:

1. Conjugate symmetry (almost commutativity):

$$\langle \underline{\boldsymbol{x}}, \underline{\boldsymbol{y}} \rangle = \overline{\langle \underline{\boldsymbol{y}}, \underline{\boldsymbol{x}} \rangle}$$

2. Linearity in the first argument:

$$\langle a\underline{\boldsymbol{x}}, \underline{\boldsymbol{y}} \rangle = a\langle \underline{\boldsymbol{x}}, \underline{\boldsymbol{y}} \rangle$$

$$\langle \underline{\boldsymbol{x}} + \underline{\boldsymbol{y}}, \underline{\boldsymbol{z}} \rangle = \langle \underline{\boldsymbol{x}}, \underline{\boldsymbol{z}} \rangle + \langle \underline{\boldsymbol{y}}, \underline{\boldsymbol{z}} \rangle$$

3. Positive definitiveness:

$$\langle \underline{\boldsymbol{x}}, \underline{\boldsymbol{x}} \rangle \geq 0$$
$$\langle \underline{\boldsymbol{x}}, \underline{\boldsymbol{x}} \rangle = 0 \implies \underline{\boldsymbol{x}} = \underline{\mathbf{0}}$$

Vector inner product

- Aka "dot product", "scalar product"
- Given $\underline{\boldsymbol{x}}, \underline{\boldsymbol{y}} \in \mathbb{F}^n$ (i.e., same number of components and also same "label" for each element), the inner product of $\underline{\boldsymbol{x}}$ and $\underline{\boldsymbol{y}}$ is defined as:

$$\langle \underline{\boldsymbol{x}}, \underline{\boldsymbol{y}} \rangle = \underline{\boldsymbol{x}}^T \cdot \underline{\boldsymbol{y}} = \sum_{i=1}^{n} x_i y_i \in \mathbb{F}$$

- $\underline{\boldsymbol{x}}^T \cdot \underline{\boldsymbol{y}}$ is read "x dotted y" or "x transposed y"

Affine spaces

- Linear algebra
  - Vector and vector spaces
  - **Affine spaces**
  - Vectors and matrices
  - Linear functions
  - Connections between Machine Learning and Linear Algebra

Affine space: definition

- If $\underline{\boldsymbol{c}}$ is a vector and $\mathcal{V}$ is a vector space then

$$\mathcal{A} = \underline{\boldsymbol{c}} + \mathcal{V} = \{\underline{\boldsymbol{c}} + \underline{\boldsymbol{v}} : \underline{\boldsymbol{v}} \in \mathcal{V}\}$$

  is called an affine space
- An affine space is a vector space translated by a point represented by a vector
  - E.g., a plane or a line that do not contain the origin

Affine space: example of plane passing through 3 points

- Given 3 not collinear vectors: $\underline{\boldsymbol{u}}_1$, $\underline{\boldsymbol{u}}_2$, and $\underline{\boldsymbol{u}}_3$, the plane containing the endpoints of the 3 vectors can be represented as $\mathcal{A} = \underline{\boldsymbol{u}}_1 + \mathcal{V}$ where $\mathcal{V} = \mathrm{Span}(\underline{\boldsymbol{u}}_2 - \underline{\boldsymbol{u}}_1, \underline{\boldsymbol{u}}_3 - \underline{\boldsymbol{u}}_1)$
- **Note**: the span of the 3 vectors has dimension 3, but an affine space with dimension 2

Affine combination: definition

- An affine combination is a linear combination of vectors where the sum of the (positive or negative) coefficients is 1:

$$\alpha_1 \underline{\boldsymbol{u}}_1 + \ldots + \alpha_n \underline{\boldsymbol{u}}_n \text{ where } \sum_i \alpha_i = 1$$

- In matrix form: $\underline{\underline{\boldsymbol{U}}} \cdot \underline{\boldsymbol{\alpha}}$ where $\underline{\mathbf{1}}^T \underline{\boldsymbol{\alpha}} = 1$

Affine hull of vectors

- Given vectors $\underline{\boldsymbol{u}}_1, ..., \underline{\boldsymbol{u}}_n$, the set of all affine combinations is the affine hull:

$$\mathcal{A} = \{\underline{\boldsymbol{v}} = \sum_{i}^{n} \alpha_i \underline{\boldsymbol{u}}_i : \sum_{i} \alpha_i = 1\} = \{\underline{\boldsymbol{v}} = \underline{\underline{\boldsymbol{U}}}\underline{\boldsymbol{\alpha}} : \underline{\mathbf{1}}^T \underline{\boldsymbol{\alpha}} = 1\}$$

- The affine hull includes each point because if $\underline{\boldsymbol{\alpha}}$ has a single 1 in position i and all others 0, we get $\underline{\boldsymbol{u}}_i$

Affine hull of vectors is an affine space

- We can write the affine hull of $\underline{\boldsymbol{u}}_1, \ldots, \underline{\boldsymbol{u}}_n$ as an affine space:

$$\underline{\boldsymbol{u}}_i + \mathrm{Span}(\underline{\boldsymbol{u}}_1 - \underline{\boldsymbol{u}}_i, \ldots, \underline{\boldsymbol{u}}_n - \underline{\boldsymbol{u}}_i)$$

  - Thus an affine space is an affine hull and vice versa
- This is the dual of "the span of vectors is a vector space"

The solution set of non-homogeneous linear system is empty or affine space

- Consider the solution of a system of non-homogeneous linear equations
$$\{\underline{\boldsymbol{x}} : \underline{\boldsymbol{a}}_1^T \underline{\boldsymbol{x}} = \beta_1, ..., \underline{\boldsymbol{a}}_m^T \underline{\boldsymbol{x}} = \beta_m\} \text{ or in matrix form } \underline{\underline{\boldsymbol{A}}} \cdot \underline{\boldsymbol{x}} = \underline{\boldsymbol{\beta}}$$
- The solution set is either empty or an affine space
- There are two cases: either the non-homogeneous system has solution, or not
- Consider the case where the system of equations has no solutions (e.g., it is contradictory, e.g., $x = 1, x = 2$), then the solution set is empty
- Consider the case where there is a solution
- Each linear system $\underline{\underline{\boldsymbol{A}}}\underline{\boldsymbol{x}} = \underline{\boldsymbol{\beta}}$ has an associated homogeneous linear system $\underline{\underline{\boldsymbol{A}}}\underline{\boldsymbol{x}} = \underline{\boldsymbol{0}}$
- If $\underline{\boldsymbol{u}}_1$ is a solution of the non-homogeneous system (i.e., $\underline{\underline{\boldsymbol{A}}}\underline{\boldsymbol{u}}_1 = \underline{\boldsymbol{\beta}}$), then any other solution $\underline{\boldsymbol{u}}_2$ is a solution (i.e., $\underline{\underline{\boldsymbol{A}}}\underline{\boldsymbol{u}}_2 = \underline{\boldsymbol{\beta}}$) $\iff$ $\underline{\boldsymbol{u}}_2 - \underline{\boldsymbol{u}}_1$ is in the vector space which is the solution of the homogeneous linear system (i.e., $\underline{\underline{\boldsymbol{A}}}(\underline{\boldsymbol{u}}_1 - \underline{\boldsymbol{u}}_2) = \underline{\boldsymbol{0}}$)

Vector space vs affine space: summary

- Linear combination vs affine combination
- Span of vectors vs affine hull of vectors
    - Span (affine hull) is the set of all linear (affine) combinations
- Vector space vs affine space

Vectors and matrices

- Linear algebra
  - Vector and vector spaces
  - Affine spaces
  - **Vectors and matrices**
  - Linear functions
  - Connections between Machine Learning and Linear Algebra

Matrices and Matrix Operations

- A matrix $\underline{\underline{\boldsymbol{A}}} \in \mathbb{F}^{m \times n}$ is a two dimensional array with dimensions $m \times n$ of elements from a field $\mathbb{F}$
- Matrix notation
  - $\underline{\underline{\boldsymbol{A}}} \in \mathbb{R}^{m \times n}$ has m rows and n columns
  - A_{ij} is the element on i-th row and j-th column

$$\underline{\underline{\boldsymbol{A}}} = \begin{pmatrix} A_{1,1} & \dots & A_{1,n} \\ \dots & A_{i,j} & \dots \\ A_{m,1} & \dots & A_{m,n} \end{pmatrix}$$

  - The convention is first rows and then columns (i.e., y-x instead of the more usual x-y) for both elements and dimensions of a matrix
- Notation for rows and columns in a matrix
  - j-th column is $\underline{\boldsymbol{a}}_j$ or $\underline{\boldsymbol{a}}_{:,j}$ (using `numpy` notation)
  - i-th row is $\underline{\boldsymbol{a}}_i^T$ or $\underline{\boldsymbol{a}}_{i,:}$
  - Fixing a coordinate (e.g., row) one gets the orthogonal indices (e.g., column)

Linear functions

- Linear algebra
  - Vector and vector spaces
  - Affine spaces
  - Vectors and matrices
  - **Linear functions**
  - Connections between Machine Learning and Linear Algebra

Linear functions over vector spaces: definition

- Consider two vector spaces $\mathcal{V}$ and $\mathcal{W}$ over the same field $\mathbb{F}$
- A linear function $f : \mathcal{V} \to \mathcal{W}$ satisfies two properties:
  1. $f(\alpha \underline{\boldsymbol{v}}) = \alpha f(\underline{\boldsymbol{v}})$
  2. $f(\underline{\boldsymbol{u}} + \underline{\boldsymbol{v}}) = f(\underline{\boldsymbol{u}}) + f(\underline{\boldsymbol{v}})$
- Linear functions "push linear combination through":

$$f(\alpha_1 \underline{\boldsymbol{v}}_1 + ... + \alpha_n \underline{\boldsymbol{v}}_n) = \alpha_1 f(\underline{\boldsymbol{v}}_1) + ... + \alpha_n f(\underline{\boldsymbol{v}}_n)$$

  - This is equivalent to the 2 properties of linear functions

From matrix to linear function

- Given a matrix $\underline{\underline{\boldsymbol{A}}} \in \mathbb{F}^{n \times m}$ we can define the function:

$$f(\underline{\boldsymbol{x}}) = \underline{\underline{\boldsymbol{A}}} \cdot \underline{\boldsymbol{x}}$$

- The function $f()$ maps m-vectors into n-vectors
  - The domain is $\mathbb{F}^m$ for the matrix-vector product to be defined
  - The co-domain is $\mathbb{F}^n$
  - MEM: the inputs are from the top and the outputs from the side
- $f(\underline{\boldsymbol{x}})$ is a linear function because of the properties of matrix-vector product

From linear function to matrix

- Consider a linear function $f : \mathbb{F}^m \rightarrow \mathbb{F}^n$
- We want to find a matrix $\underline{\underline{\boldsymbol{A}}}$ such that $f(\underline{\boldsymbol{x}}) = \underline{\underline{\boldsymbol{A}}} \cdot \underline{\boldsymbol{x}}$
- ***Solution***
- We know that $\underline{\underline{\boldsymbol{A}}} \in \mathbb{F}^{n \times m}$ from matrix-vector product definition, but what are the elements of $\underline{\underline{\boldsymbol{A}}}$?
- Note that if we compute $\underline{\underline{\boldsymbol{A}}} \cdot \underline{\boldsymbol{e_i}}$, where $e_i = (0, ..., 0, 1, ...0)$ is the i-th standard generator, we obtain $\underline{\boldsymbol{a}}_i$ (i.e., the i-th column of $\underline{\underline{\boldsymbol{A}}}$)
- Thus $\underline{\underline{\boldsymbol{A}}}$ is the matrix with columns equal to the standard generators transformed by $f()$

Linear functions: examples and non-examples

- The identity function corresponds to the identity matrix
- A diagonal matrix corresponds to a transformation scaling each coordinate independently
- Rotation and scaling are linear transformations
- Translation is not a linear function, since it does not satisfy either of the two linearity properties

A linear function maps zero vector to zero vector

- It can be proved that a linear function $f : \mathcal{V} \to \mathcal{W}$ maps the zero vector of $\mathcal{V}$ to the zero vector of $\mathcal{W}$

Kernel of a function

- The kernel of a (linear) function f is the set of vectors that are transformed by f into the $\underline{\mathbf{0}}$ vector

$$\mathrm{Ker}(f) = \{\underline{\boldsymbol{v}} : f(\underline{\boldsymbol{v}}) = \underline{\mathbf{0}}\}$$

Kernel of a linear function and null space of matrix

- If f is written in terms of a matrix as $f(\underline{\boldsymbol{x}}) = \underline{\underline{\boldsymbol{A}}} \cdot \underline{\boldsymbol{x}}$ then:

$$\mathrm{Ker}(f) = \mathrm{Null}(\underline{\underline{\boldsymbol{A}}})$$

- The kernel of a linear function is the null space of the columns of the associated matrix $\underline{\underline{\boldsymbol{A}}}$

Domain

- Consider a function $f : \mathcal{V} \to \mathcal{W}$
- $\mathcal{V}$ "domain": the set of all values where the function is defined

Image of domain

- Consider a function $f : \mathcal{V} \to \mathcal{W}$
- $f(\mathcal{V})$ "image of domain": the set of all values that the function can assume

Codomain

- $\mathcal{W}$ "co-domain": the set where the function assumes its value (e.g., $\mathbb{R}^2$)
- Note that often it is not easy to describe $f(\mathcal{V})$ in a compact way so we use $\mathcal{W} \supseteq f(\mathcal{V})$

One-to-one function: definition

- Aka injective
- Consider a function $f : \mathcal{V} \to \mathcal{W}$
- A function is one-to-one $\iff$ two different elements $v_1 \neq v_2 \in \mathcal{V}$ have different images $f(v_1) \neq f(v_2)$
- Equivalent definition using contrapositive
  - One-to-one function $\iff$ two elements v_1 and v_2 have the same image $f(v_1) = f(v_2)$, then they are equal $v_1 = v_2$
- Equivalent definition in terms of set cardinality
  - One-to-one function $\iff$ $|f(\mathcal{V})| = |\mathcal{V}|$, i.e., the image of the domain has the same number of elements as the domain

Onto function

- Aka surjective
- Consider a function $f : \mathcal{V} \to \mathcal{W}$
- A function is onto $\iff$ for any element of its co-domain $w \in \mathcal{W}$, there exists an element of the domain $v \in \mathcal{V}$ that is transformed into it, i.e., $f(v) = w$
- Equivalent definition in terms of set cardinality
    - Onto function $\iff$ $f(\mathcal{V}) = \mathcal{W}$, i.e., the image of the domain is equal to the co-domain
    - Note that any function can be made surjective by restricting $\mathcal{W}$ to $f(\mathcal{V})$

Invertible function

- A function $f : \mathcal{V} \to \mathcal{W}$ is invertible $\iff$ it is both one-to-one (injective) and onto (surjective)
- Equivalently: $\forall w \in W \exists! v \in V : f(v) = w$
- Equivalent definition in terms of set cardinality
  - Invertible function $\iff |\mathcal{V}| = |\mathcal{W}|$, i.e., the co-domain and the domain have the same number of elements

Inverse of a function

- The inverse of f is $f^{-1} : \mathcal{W} \to \mathcal{V}$ and $f \circ f^{-1}$ is the identity function

One-to-one lemma for linear functions

- A linear function is one-to-one $\iff$ its kernel is the trivial vector space, i.e., $\mathrm{Ker}(f) = \{\underline{\mathbf{0}}\}$
- Equivalently the associated matrix $\underline{\underline{\boldsymbol{A}}}$ has $\mathrm{Null}(\underline{\underline{\boldsymbol{A}}}) = \{\underline{\mathbf{0}}\}$

Matrix-matrix multiplication and linear function composition

- We have seen that there is a correspondence between linear functions and matrices
- What does composition of linear functions correspond to?
- Consider two matrices $\underline{\underline{\boldsymbol{A}}}$ and $\underline{\underline{\boldsymbol{B}}}$ and the two associated functions: $f(\underline{\boldsymbol{y}}) = \underline{\underline{\boldsymbol{A}}} \cdot \underline{\boldsymbol{y}}$ and $g(\underline{\boldsymbol{x}}) = \underline{\underline{\boldsymbol{B}}} \cdot \underline{\boldsymbol{x}}$
- The composed function is $h(\underline{\boldsymbol{x}}) = (f \circ g)(\underline{\boldsymbol{x}}) = f(g(\underline{\boldsymbol{x}}))$
- It can be shown that: $h(\underline{\boldsymbol{x}}) = \underline{\underline{\boldsymbol{A}}} \cdot \underline{\underline{\boldsymbol{B}}} \cdot \underline{\boldsymbol{x}}$

Matrix inverse

- Using the definition of inverse functions, two square matrices $\underline{\underline{\boldsymbol{A}}}$ and $\underline{\underline{\boldsymbol{B}}}$ are inverses $\iff$ $\underline{\underline{\boldsymbol{A}}} \cdot \underline{\underline{\boldsymbol{B}}} = \underline{\underline{\boldsymbol{I}}}$
- We indicate the (unique) inverse of $\underline{\underline{\boldsymbol{A}}}$ with $\underline{\underline{\boldsymbol{A}}}^{-1}$

Invertible matrix and solution of matrix-vector equation

- Given an invertible (square) matrix $\underline{\underline{\boldsymbol{A}}}$, then $f(\underline{\boldsymbol{x}}) = \underline{\underline{\boldsymbol{A}}} \cdot \underline{\boldsymbol{x}}$ is an invertible function, i.e., $f()$ is one-to-one and onto
- Thus the matrix-vector equation $\underline{\underline{\boldsymbol{A}}} \cdot \underline{\boldsymbol{x}} = \underline{\boldsymbol{b}}$ has one and only one solution $\underline{\boldsymbol{x}}$ for any $\underline{\boldsymbol{b}}$, i.e., $\underline{\underline{\boldsymbol{A}}}^{-1}\underline{\boldsymbol{b}}$

Invertibility of matrix-matrix product

- If $\underline{\underline{\boldsymbol{A}}}$ and $\underline{\underline{\boldsymbol{B}}}$ are invertible and can be multiplied, then $\underline{\underline{\boldsymbol{A}}} \cdot \underline{\underline{\boldsymbol{B}}}$ is invertible

Systems of Linear Equations

- Represented as $Ax = b$
- Solutions using Gaussian elimination
- Conditions for existence and uniqueness

Rank of a Matrix

- Number of linearly independent rows or columns
- Rank-nullity theorem: $\mathrm{rank}(A) + \mathrm{nullity}(A) = n$
- Implications for solution spaces

Inverse of a Matrix

- Matrix A is invertible if $AA^{-1} = I$
- Not all matrices are invertible (singular matrices)
- Importance for solving linear systems and optimization

Determinants

- Scalar value summarizing matrix properties
- Zero determinant implies matrix is singular
- Used in volume interpretation and invertibility

Orthogonality and Orthonormal Bases

- Two vectors are orthogonal if $x^T y = 0$
- Orthonormal set: orthogonal unit vectors
- Used in projections, decomposition, and optimization

Projections and Least Squares

- Projection of b onto column space of A: $\hat{b} = A(A^T A)^{-1} A^T b$
- Basis for linear regression
- Solves inconsistent systems approximately

Eigenvalues and Eigenvectors

- For $Ax = \lambda x$, x is an eigenvector and λ an eigenvalue
- Capture invariant directions of transformations
- Used in PCA, spectral clustering, and stability analysis

Diagonalization

- Expressing matrix as $A = PDP^{-1}$ where D is diagonal
- Simplifies powers and exponentials of matrices
- Requires linearly independent eigenvectors

Symmetric Matrices

- $A = A^T$
- All eigenvalues are real
- Basis of many ML methods (e.g., covariance matrices, kernels)

Positive Definite and Semidefinite Matrices

- $x^T Ax > 0$ for all $x \neq 0$: positive definite
- Ensures convexity in optimization
- Common in loss functions and kernels

Singular Value Decomposition (SVD)

- Factorization: $A = U\Sigma V^T$
- Generalizes eigen-decomposition to all matrices
- Used in PCA, matrix compression, and pseudoinverse

Principal Component Analysis (PCA)

- Dimensionality reduction via eigen-decomposition of covariance matrix
- Projects data onto directions of maximum variance
- Computed using SVD in practice

Matrix Norms

- Measures of matrix size: Frobenius, spectral, etc.
- Important for measuring errors and convergence
- Used in regularization and optimization

Vector Norms

- L^p norms (e.g., L^1, L^2)
- Measure magnitude of vectors
- Used in loss functions and sparsity enforcement

Trace of a Matrix

- Sum of diagonal entries: $\mathrm{Tr}(A)$
- Invariant under cyclic permutations: $\mathrm{Tr}(ABC) = \mathrm{Tr}(CAB)$
- Appears in matrix derivatives and expectation identities

Kronecker and Hadamard Products

- Kronecker: tensor product of matrices
- Hadamard: element-wise multiplication
- Important in multi-dimensional ML models and tensor operations

Matrix Calculus

- Derivatives of scalar functions w.r.t. vectors/matrices
- Gradient and Hessian computation
- Crucial for backpropagation and optimization algorithms

Connections between Machine Learning and Linear Algebra

- Linear algebra
  - Vector and vector spaces
  - Affine spaces
  - Vectors and matrices
  - Linear functions
  - **Connections between Machine Learning and Linear Algebra**