CMPSC 448: Machine Learning

Lecture 3.1 Linear Algebra

Rui Zhang

Fall 2021

slides credit: Dr. Jing Liu

Why Linear Algebra for Machine Learning?

Linear Algebra is a key foundation to machine learning:

- **Representation**: An object model to represent data
- **Operation**: To describe the operation of algorithms
- **Implementation**: and the implementation of algorithms in code
- **Deeper Intuition**: If you can understand machine learning methods at the level of vectors and matrices, you will improve your intuition for how and when they work

What is Linear Algebra

Definition

More generally, we define a linear equation of n variables

$$x_1, x_2, \ldots, x_n$$

to be one that can be expressed in the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

where $a_1, a_2, \ldots, a_n$ and b are constants, and the a's are not all zero.

What is Linear Algebra

Definition

More generally, we define a linear equation of n variables

$$x_1, x_2, \ldots, x_n$$

to be one that can be expressed in the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

where a_1, a_2, ..., a_n and b are constants, and the a's are not all zero.

- Systems of linear equations, briefly linear systems, can be found in the earliest writings of many ancient civilizations.
- So linear algebra, traditionally, is the art of solving linear equations and systems of linear equations.

Solving Linear Systems

$$\begin{array}{rcrcrcr} x & + & y & + & 2z & = & 9 \\ 2x & + & 4y & - & 3z & = & 1 \\ 3x & + & 6y & - & 5z & = & 0 \end{array}$$

Solving Linear Systems

- Write down the corresponding augmented matrix

$$\begin{array}{rcrcrcr} x & + & y & + & 2z & = & 9 \\ 2x & + & 4y & - & 3z & = & 1 \\ 3x & + & 6y & - & 5z & = & 0 \end{array} \qquad \left[\begin{array}{rrr|r} 1 & 1 & 2 & 9 \\ 2 & 4 & -3 & 1 \\ 3 & 6 & -5 & 0 \end{array}\right]$$

Solving Linear Systems

E2.1 In the following example we will illustrate how to use elementary row operations and an augmented matrix to solve a linear system.

Solution

- Write down the corresponding augmented matrix

$$\begin{array}{rcrcrcr} x & + & y & + & 2z & = & 9 \\ 2x & + & 4y & - & 3z & = & 1 \\ 3x & + & 6y & - & 5z & = & 0 \end{array} \qquad \left[\begin{array}{rrr|r} 1 & 1 & 2 & 9 \\ 2 & 4 & -3 & 1 \\ 3 & 6 & -5 & 0 \end{array}\right]$$

- Add (-2) times the first row to the second row

Solution

- Add (-2) times the first row to the second row

$$\begin{array}{rcrcrcr} x & + & y & + & 2z & = & 9 \\ & & 2y & - & 7z & = & -17 \\ 3x & + & 6y & - & 5z & = & 0 \end{array} \qquad \left[\begin{array}{rrr|r} 1 & 1 & 2 & 9 \\ 0 & 2 & -7 & -17 \\ 3 & 6 & -5 & 0 \end{array}\right]$$

- Add (-3) times the first row to the third row

$$\begin{array}{rcrcrcr} x & + & y & + & 2z & = & 9 \\ & & 2y & - & 7z & = & -17 \\ & & 3y & - & 11z & = & -27 \end{array} \qquad \left[\begin{array}{rrr|r} 1 & 1 & 2 & 9 \\ 0 & 2 & -7 & -17 \\ 0 & 3 & -11 & -27 \end{array}\right]$$

- Multiply the second row by $\left(\frac{1}{2}\right)$

$$\begin{array}{rcrcrcr} x & + & y & + & 2z & = & 9 \\ & & y & - & \frac{7}{2}z & = & -\frac{17}{2} \\ & & 3y & - & 11z & = & -27 \end{array} \qquad \left[\begin{array}{rrr|r} 1 & 1 & 2 & 9 \\ 0 & 1 & -\frac{7}{2} & -\frac{17}{2} \\ 0 & 3 & -11 & -27 \end{array}\right]$$

- Add (-3) times the second row to the third row

Solution

- Add (-3) times the second row to the third row

$$\begin{array}{rcrcrcr} x & + & y & + & 2z & = & 9 \\ & & y & - & \frac{7}{2}z & = & -\frac{17}{2} \\ & & & - & \frac{1}{2}z & = & -\frac{3}{2} \end{array} \qquad \left[\begin{array}{ccr|r} 1 & 1 & 2 & 9 \\ 0 & 1 & -\frac{7}{2} & -\frac{17}{2} \\ 0 & 0 & -\frac{1}{2} & -\frac{3}{2} \end{array}\right]$$

- Multiply the third row by (-2)

$$\begin{array}{rcrcrcr} x & + & y & + & 2z & = & 9 \\ & & y & - & \frac{7}{2}z & = & -\frac{17}{2} \\ & & & & z & = & 3 \end{array} \qquad \left[\begin{array}{ccr|r} 1 & 1 & 2 & 9 \\ 0 & 1 & -\frac{7}{2} & -\frac{17}{2} \\ 0 & 0 & 1 & 3 \end{array}\right]$$

- Add (-1) times the second row to the first row

$$\begin{array}{rcrcrcr} x & & & + & \frac{11}{2}z & = & \frac{35}{2} \\ & & y & - & \frac{7}{2}z & = & -\frac{17}{2} \\ & & & & z & = & 3 \end{array} \qquad \left[\begin{array}{ccr|r} 1 & 0 & \frac{11}{2} & \frac{35}{2} \\ 0 & 1 & -\frac{7}{2} & -\frac{17}{2} \\ 0 & 0 & 1 & 3 \end{array}\right]$$

Solution

- Add $\left(-\frac{11}{2}\right)$ times the third row to the first row

$$\begin{array}{rcrcrcr} x & & & & & = & 1 \\ & & y & - & \frac{7}{2}z & = & -\frac{17}{2} \\ & & & & z & = & 3 \end{array} \qquad \left[\begin{array}{rrr|r} 1 & 0 & 0 & 1 \\ 0 & 1 & -\frac{7}{2} & -\frac{17}{2} \\ 0 & 0 & 1 & 3 \end{array}\right]$$

- Add $\left(\frac{7}{2}\right)$ times the third row to the second row

$$\begin{array}{rcrcrcr} x & & & & & = & 1 \\ & & y & & & = & 2 \\ & & & & z & = & 3 \end{array} \qquad \left[\begin{array}{rrr|r} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{array}\right]$$

- The solution $x = 1$, $y = 2$, $z = 3$ is now evident.

- The initial augmented matrix is reduced to $\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{bmatrix}$.

- This is an example of a matrix known to be in reduced row echelon form.

Three ways of looking at a linear system

- Three different ways of asking the same question

  Row (Intersection)

- It becomes harder to vision as the dimension increase

$$
\begin{aligned}
4y + z &= 5 \\
3x + 8y + z &= 9 \\
x + 2y + z &= 3
\end{aligned}
$$

Three ways of looking at a linear system

- Three different ways of asking the same question

Row (Intersection) Column (Combination)

- It becomes harder to vision as the dimension increase

$$\begin{aligned} 4y + z &= 5 \\ 3x + 8y + z &= 9 \\ x + 2y + z &= 3 \end{aligned}$$

$$x \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix} + y \begin{bmatrix} 4 \\ 8 \\ 2 \end{bmatrix} + z \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 9 \\ 3 \end{bmatrix}$$

Three ways of looking at a linear system

- Three different ways of asking the same question

Row (Intersection) Column (Combination) Matrix (Inverse image)

- It becomes harder to vision as the dimension increase

$$\begin{aligned} 4y + z &= 5 \\ 3x + 8y + z &= 9 \\ x + 2y + z &= 3 \end{aligned} \qquad x\begin{bmatrix}0\\3\\1\end{bmatrix} + y\begin{bmatrix}4\\8\\2\end{bmatrix} + z\begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}5\\9\\3\end{bmatrix} \qquad \begin{bmatrix}0 & 4 & 1\\3 & 8 & 1\\1 & 2 & 1\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix} = \begin{bmatrix}5\\9\\3\end{bmatrix}$$

Three ways of looking at a linear system

- Three different ways of asking the same question

Row (Intersection) Column (Combination) Matrix (Inverse image)

- It becomes harder to vision as the dimension increase

$$\begin{aligned} 4y + z &= 5 \\ 3x + 8y + z &= 9 \\ x + 2y + z &= 3 \end{aligned} \qquad x\begin{bmatrix}0\\3\\1\end{bmatrix} + y\begin{bmatrix}4\\8\\2\end{bmatrix} + z\begin{bmatrix}1\\1\\1\end{bmatrix} = \begin{bmatrix}5\\9\\3\end{bmatrix} \qquad \begin{bmatrix}0 & 4 & 1\\3 & 8 & 1\\1 & 2 & 1\end{bmatrix}\begin{bmatrix}x\\y\\z\end{bmatrix} = \begin{bmatrix}5\\9\\3\end{bmatrix}$$

- Given an augmented matrix, how many solutions will we have?
- What is the solution if we change $\mathbf{b}$, for example, change to

$$\mathbf{b} = \begin{bmatrix}2\\12\\2\end{bmatrix}$$

- Can we solve $\mathbf{Ax} = \mathbf{b}$ for all $\mathbf{b}$?

Number of solutions in a linear system

Definition

In general, we say that a linear system is consistent if it has at least one solution and inconsistent if it has no solutions.

Number of solutions in a linear system

Definition

In general, we say that a linear system is consistent if it has at least one solution and inconsistent if it has no solutions.

Theorem

Every system of linear equations has zero, one, or infinitely many solutions. There are no other possibilities.

Homogeneous Systems

Definition

A system of linear equations is said to be homogeneous if the constant terms are all zero; that is, the system has the form

$$\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0\\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0\\
\vdots \qquad\quad \vdots \qquad\qquad \vdots \quad &\quad \vdots\\
a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{mn}x_n &= 0
\end{aligned}$$

Homogeneous Systems

Definition

A system of linear equations is said to be homogeneous if the constant terms are all zero; that is, the system has the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0 \\ \vdots \qquad\quad \vdots \qquad\qquad\quad \vdots \quad & \quad \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{mn}x_n &= 0 \end{aligned}$$

Q: Why is every homogeneous system of linear equations consistent?

Because $x_1 = x_2 = \cdots = x_n = 0$ is always a solution for it

- This solution is called the trivial solution; if there are other nonzero solutions, they are called nontrivial solutions.

Homogeneous Systems

Definition

A system of linear equations is said to be homogeneous if the constant terms are all zero; that is, the system has the form

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0 \\
\vdots \qquad\quad \vdots \qquad\qquad\quad \vdots \qquad & \quad \vdots \\
a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{mn}x_n &= 0
\end{aligned}
$$

Q: Why is every homogeneous system of linear equations consistent?

Because $x_1 = x_2 = \cdots = x_n = 0$ is always a solution for it

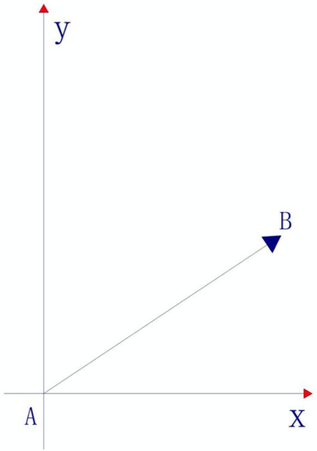
- This solution is called the trivial solution; if there are other nonzero solutions, they are called nontrivial solutions.

- Because a homogeneous linear system always has the trivial solution, there are only two possibilities for its solutions:

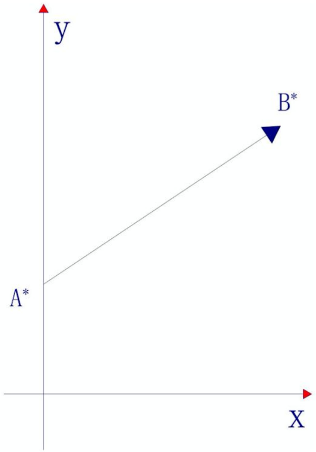
  1. The system has only the trivial solution.
  2. The system has infinitely many solutions including the trivial solution.

Vectors

- The earliest notion of vector comes from physics, where the term vector is used to indicate a quantity that has both magnitude and direction.



- Two vectors are equal if and only if they have the same magnitude and direction.
- In writing, we specify vectors by enclosing the components in square brackets, for example,

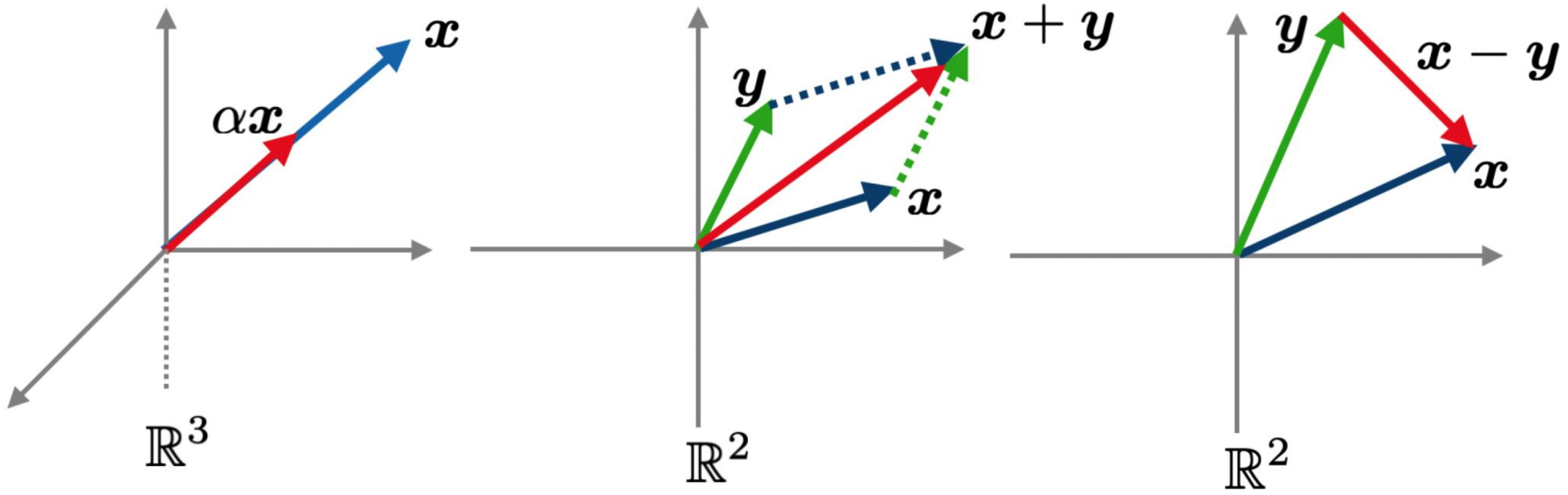


$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \in \mathbb{R}^2 \quad \text{and} \quad \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \in \mathbb{R}^3.$$

In general we are going to consider $\mathbb{R}^n$.

- We denote a vector by printing a letter in boldface $\mathbf{v}$.

Vector Operations



$$\alpha \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix} = \begin{bmatrix} \alpha x_1 \\ \vdots \\ \alpha x_d \end{bmatrix}, \quad \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix} + \begin{bmatrix} y_1 \\ \vdots \\ y_d \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_d + y_d \end{bmatrix}, \quad \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix} - \begin{bmatrix} y_1 \\ \vdots \\ y_d \end{bmatrix} = \begin{bmatrix} x_1 - y_1 \\ \vdots \\ x_d - y_d \end{bmatrix}$$

Vector Spaces

- In general, a vector space involves four things,

  two sets $\mathcal{F}$ and $\mathcal{V}$, two operations called addition and scalar multiplication.

1. $\mathcal{F}$ is a scalar field.

- For us $\mathcal{F}$ is either the field $\mathbb{R}$ of real numbers or $\mathbb{C}$ of complex numbers.

2. $\mathcal{V}$ is a non-empty set of mathematical objects called vectors.

- So in the general sense, matrices, polynomials, continuous and differentiable functions are all known as vectors.

Vector Spaces

- In general, a vector space involves four things,

  two sets $\mathcal{F}$ and $\mathcal{V}$, two operations called addition and scalar multiplication.

1. $\mathcal{F}$ is a scalar field.

- For us $\mathcal{F}$ is either the field $\mathbb{R}$ of real numbers or $\mathbb{C}$ of complex numbers.

2. $\mathcal{V}$ is a non-empty set of mathematical objects called vectors.

- So in the general sense, matrices, polynomials, continuous and differentiable functions are all known as vectors.

3. Addition is an operation between elements of $\mathcal{V}$.

- By addition we mean a rule for associating with each pair of objects $\mathbf{u}$ and $\mathbf{v}$ in $\mathcal{V}$ an object $\mathbf{u} + \mathbf{v}$, called the sum of $\mathbf{u}$ and $\mathbf{v}$;

4. Scalar multiplication is an operation between elements of $\mathcal{F}$ and $\mathcal{V}$.

- By scalar multiplication we mean a rule for associating with each object $\mathbf{u}$ in $\mathcal{V}$ and each scalar α in $\mathcal{F}$ an object $\alpha\mathbf{u}$, called the scalar multiple of $\mathbf{u}$ by α.

Vector Spaces

Definition

If the addition and scalar multiplication operations satisfy the following properties by all objects $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ in $\mathcal{V}$ and all scalars α and β in $\mathcal{F}$, then we call

$$\mathcal{V} \text{ a vector space over } \mathcal{F}.$$

1. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
2. $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$
3. $\alpha(\mathbf{u} + \mathbf{v}) = \alpha\mathbf{u} + \alpha\mathbf{v}$
4. $(\alpha + \beta)\mathbf{u} = \alpha\mathbf{u} + \beta\mathbf{u}$
5. $(\alpha\beta)\mathbf{u} = \alpha(\beta\mathbf{u})$
6. $1\ \mathbf{u} = \mathbf{u}$
7. $\mathbf{u} + \mathbf{0} = \mathbf{u}$
8. $\mathbf{u} + (-\mathbf{u}) = \mathbf{0}$
9. $\mathbf{u} + \mathbf{v}$ is in $\mathcal{V}$
10. $\alpha\mathbf{u}$ is in $\mathcal{V}$

- It is important to keep in mind that one does not prove axioms; rather, we check if they are satisfied or we take them as assumptions that serve as the starting point for proving theorems.

Examples of Vector Space

- $\mathbb{R}^n$ with the usual addition and scalar multiplication is a vector space over $\mathbb{R}$.
- $\mathbb{R}^{m\times n}$ of $m \times n$ matrices with the usual operations is a vector space over $\mathbb{R}$.
- $\mathbb{C}^{m\times n}$ of $m \times n$ matrices with the usual operations is a vector space over $\mathbb{C}$.
- The field itself is a vector space with vector addition begin the field addition and scalar multiplication being the field multiplication.
- With the usual function addition and scalar multiplication defined by

$$\big(f+g\big)(x) = f(x) + g(x) \qquad \text{and} \qquad \big(\alpha f\big)(x) = \alpha f(x)$$

Subspace

- It is often the case that we are interested in a subset of a known vector space.

Definition

A non-empty subset $\mathcal{H}$ of a vector space $\mathcal{V}$ over $\mathcal{F}$ is called a subspace of $\mathcal{V}$ if $\mathcal{H}$ is itself a vector space over $\mathcal{F}$ under the same addition and scalar multiplication.

Possible subspaces of $\mathbb{R}^3$:

- The $\mathbf{0}$ vector (smallest subspace and in all subspaces)
- Any line or plane through origin
- All of $\mathbb{R}^3$

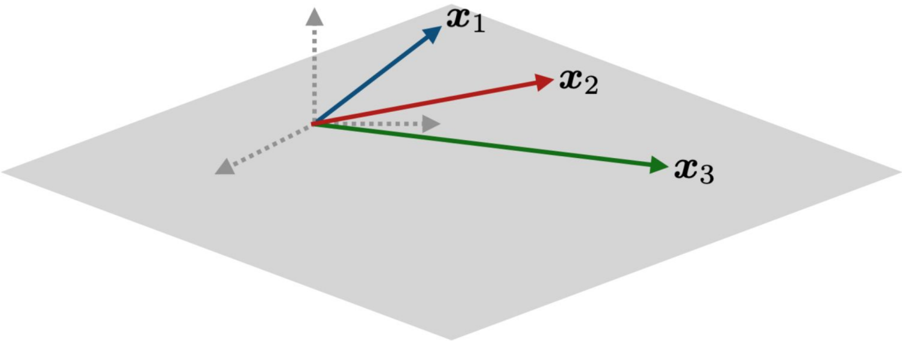
Linear Combination

Definition

If $\mathbf{u}$ is a vector in a vector space $\mathcal{V}$, then $\mathbf{u}$ is said to be a linear combination of the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, . . ., $\mathbf{v}_r$ in $\mathcal{V}$ if $\mathbf{u}$ can be expressed in the form

$$\mathbf{u} = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_r\mathbf{v}_r$$

where α_1, α_2 , . . ., α_r are scalars in $\mathcal{F}$. These scalars are called the coefficients.



Span

Theorem

If $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is a non-empty set of vectors in a vector space $\mathcal{V}$, then the set $\mathcal{H}$ of all possible linear combinations of the vectors in $\mathcal{S}$ is a subspace of $\mathcal{V}$.

Span

Theorem

If $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is a non-empty set of vectors in a vector space $\mathcal{V}$, then the set $\mathcal{H}$ of all possible linear combinations of the vectors in $\mathcal{S}$ is a subspace of $\mathcal{V}$.

- This subspace $\mathcal{H}$ is called the subspace of $\mathcal{V}$ generated by $\mathcal{S}$, and we say the set $\mathcal{S}$ spans $\mathcal{H}$, or $\mathcal{H}$ is the subspace spanned by $\mathcal{S}$.
- We denote this subspace $\mathcal{H}$ as

$$\mathcal{H} = \text{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\} \qquad \text{or} \qquad \mathcal{H} = \text{span}(\mathcal{S})$$

The set $\mathcal{S}$ is known as the spanning set for $\mathcal{H}$.

Span

- Recall that the standard unit vectors in $\mathbb{R}^n$ are

$$\mathbf{e}_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^\mathsf{T}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 & 1 & \cdots & 0 \end{bmatrix}^\mathsf{T}, \quad \ldots \quad \mathbf{e}_n = \begin{bmatrix} 0 & 0 & \cdots & 1 \end{bmatrix}^\mathsf{T}$$

- The subspace $\mathcal{H} = \text{span}\{\mathbf{e}_1\}$ is the subspace consisting of all scalar multiples of $\mathbf{e}_1$, so geometrically it is simply the line through origin determined by $\mathbf{e}_1$.

Span

- Recall that the standard unit vectors in $\mathbb{R}^n$ are

$$\mathbf{e}_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^\mathsf{T}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 & 1 & \cdots & 0 \end{bmatrix}^\mathsf{T}, \quad \ldots \quad \mathbf{e}_n = \begin{bmatrix} 0 & 0 & \cdots & 1 \end{bmatrix}^\mathsf{T}$$

- The subspace $\mathcal{H} = \text{span}\{\mathbf{e}_1\}$ is the subspace consisting of all scalar multiples of $\mathbf{e}_1$, so geometrically it is simply the line through origin determined by $\mathbf{e}_1$.
- The subspace

$$\mathcal{H} = \text{span}\{\mathbf{e}_1, \mathbf{e}_2\}$$

is the subspace containing all linear combinations $\mathbf{e}_1$ and $\mathbf{e}_2$,

$$\alpha\mathbf{e}_1 + \beta\mathbf{e}_2, \qquad \text{where } \alpha \text{ and } \beta \text{ are in } \mathcal{F}.$$

geometrically, it is simply the plane through origin determined by $\mathbf{e}_1$ and $\mathbf{e}_2$.

Span

- Recall that the standard unit vectors in $\mathbb{R}^n$ are

$$\mathbf{e}_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^\mathsf{T}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 & 1 & \cdots & 0 \end{bmatrix}^\mathsf{T}, \quad \ldots \quad \mathbf{e}_n = \begin{bmatrix} 0 & 0 & \cdots & 1 \end{bmatrix}^\mathsf{T}$$

- The subspace $\mathcal{H} = \text{span}\{\mathbf{e}_1\}$ is the subspace consisting of all scalar multiples of $\mathbf{e}_1$, so geometrically it is simply the line through origin determined by $\mathbf{e}_1$.
- The subspace

$$\mathcal{H} = \text{span}\{\mathbf{e}_1, \mathbf{e}_2\}$$

is the subspace containing all linear combinations $\mathbf{e}_1$ and $\mathbf{e}_2$,

$$\alpha\mathbf{e}_1 + \beta\mathbf{e}_2, \qquad \text{where } \alpha \text{ and } \beta \text{ are in } \mathcal{F}.$$

geometrically, it is simply the plane through origin determined by $\mathbf{e}_1$ and $\mathbf{e}_2$.
- Of course, we can go on adding more $\mathbf{e}_i$ into the set $\mathcal{S}$, we will just end up with subspaces that are hyperplanes determined by those vectors in $\mathcal{S}$.
- If we put all n of those standard unit vectors in the $\mathcal{S}$, then $\text{span}(\mathcal{S}) = \mathbb{R}^n$.
- Thus the n standard unit vectors span $\mathbb{R}^n$ since every vector $\mathbf{u}$ in $\mathbb{R}^n$ is a linear combination of those $\mathbf{e}_i$s, and the set $\{\mathbf{e}_1 \cdots \mathbf{e}_n\}$ is a spanning set for $\mathbb{R}^n$.

Linear Independence

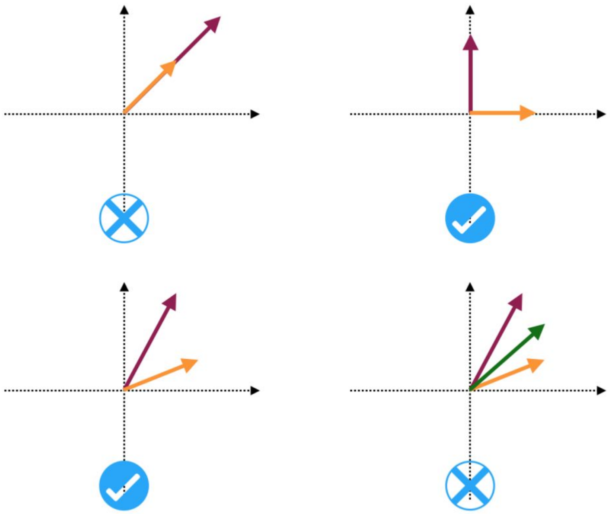
Defintion

If $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is a set of two or more vectors in a vector space $\mathcal{V}$, then $\mathcal{S}$ is said to be a linearly independent set if no vector in $\mathcal{S}$ can be expressed as a linear combination of the others. Otherwise it is said to be linearly dependent.

Linear Independence

Defintion

If $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is a set of two or more vectors in a vector space $\mathcal{V}$, then $\mathcal{S}$ is said to be a linearly independent set if no vector in $\mathcal{S}$ can be expressed as a linear combination of the others. Otherwise it is said to be linearly dependent.



Linear Independence

Defintion

If $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is a set of two or more vectors in a vector space $\mathcal{V}$, then $\mathcal{S}$ is said to be a linearly independent set if no vector in $\mathcal{S}$ can be expressed as a linear combination of the others. Otherwise it is said to be linearly dependent.

Theorem

A nonempty set $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ in a vector space $\mathcal{V}$ is linearly independent if and only if the only coefficients satisfying the vector equation

$$\alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_r \mathbf{v}_r = \mathbf{0}$$

are $\alpha_1 = \alpha_2 = \cdots = \alpha_r = 0$.

Linear Independence

Q: Determine whether the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 5 \\ 6 \\ -1 \end{bmatrix}, \mathbf{v}_1 = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$$

are linearly independent or linearly dependent in $\mathbb{R}^3$.

It depends whether $\alpha_1 \underset{\sim}{v_1} + \alpha_2 \underset{\sim}{v_2} + \alpha_3 \underset{\sim}{v_3} = \underset{\sim}{0}$

has non-trivial solution.

The matrix equation $\begin{bmatrix} 1 & 5 & 3 \\ -2 & 6 & 2 \\ 3 & -1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

$$\begin{bmatrix} 1 & 5 & 3 \\ -2 & 6 & 2 \\ 3 & -1 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1/2 \\ 0 & 1 & 1/2 \\ 0 & 0 & 0 \end{bmatrix} \neq I_3, \text{ Not L. I.}$$

Basis

Definition (Basis)

The basis of a vector space, $\mathcal{V}$, is a set of vectors $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k\}$ that satisfy:

(a) $\mathcal{V} = \texttt{span}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k\}$

(b) $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k\}$ are linearly independent.

Basis

- The standard basis for $\mathbb{R}^n$

$$\mathbf{e}_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^\mathsf{T}, \mathbf{e}_2 = \begin{bmatrix} 0 & 1 & \cdots & 0 \end{bmatrix}^\mathsf{T}, \cdots \mathbf{e}_n = \begin{bmatrix} 0 & 0 & \cdots & 1 \end{bmatrix}^\mathsf{T}$$

Q Do the vectors $\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$, $\mathbf{v}_2 = \begin{bmatrix} 2 \\ 9 \\ 0 \end{bmatrix}$, and $\mathbf{v}_3 = \begin{bmatrix} 3 \\ 3 \\ 4 \end{bmatrix}$ form a basis for $\mathbb{R}^3$?

A Yes. Because these three vectors are linearly independent and span $\mathbb{R}^3$.

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 9 & 3 \\ 1 & 0 & 4 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \implies \begin{array}{l} \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \alpha_3\mathbf{v}_3 = \mathbf{0}; \\ \beta_1\mathbf{v}_1 + \beta_2\mathbf{v}_2 + \beta_3\mathbf{v}_3, = \mathbf{b} \qquad \text{for all } \mathbf{b} \in \mathbb{R}^3. \end{array}$$

Basis

Any vector in the subspace can be represented **uniquely** as a linear combination of the vectors in basis:

$$\boldsymbol{x} = \sum_{i=1}^{k} \alpha_i \boldsymbol{x}_i, \text{ for all } \boldsymbol{x} \in \mathcal{V}$$

Example

For example consider the whole 3 dimensional space, $\mathcal{V} = \mathbb{R}^3$. The set of vectors $[1,1,1]^\top, [1,2,0]^\top\}$, and $\{[1,3,1]^\top$ constitute a basis for $\mathcal{V}$. For instance, $[1,2,3]^\top$ can be expressed as linear combination of these vectors:

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = (1.5) \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + (-2) \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} + (1.5) \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}$$

Note that finding these coefficients requires solving a system of linear equations (will be discussed later).

Dimension

Recall that a subspace can have infinitely many bases, but the number of vectors in any basis is fixed! The number of elements in the basis is a very important characteristic of the subspace as denoted below.

Dimension

Recall that a subspace can have infinitely many bases, but the number of vectors in any basis is fixed! The number of elements in the basis is a very important characteristic of the subspace as denoted below.

Definition

The dimension of a vector space $\mathcal{V}$ is denoted by $\dim(\mathcal{V})$ and is defined to be the number of vectors in a basis for $\mathcal{V}$.

Dimension

Recall that a subspace can have infinitely many bases, but the number of vectors in any basis is fixed! The number of elements in the basis is a very important characteristic of the subspace as denoted below.

Definition

The dimension of a vector space $\mathcal{V}$ is denoted by $\dim(\mathcal{V})$ and is defined to be the number of vectors in a basis for $\mathcal{V}$.

For possible subspaces of $\mathbb{R}^3$:

- The dimension of single point $\mathbf{0}$ vector (smallest subspace) is 0!
- The dimension of any line is 1
- The dimension of plane through origin is 2
- The dimension of all of $\mathbb{R}^3$ is 3

Dot product

Definition

The inner product of any two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ is defined as:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^\top \boldsymbol{y} = \sum_{i=1}^{d} x_i y_i \in \mathbb{R}$$

Inner Product and Inner Product space

- To generalize the of concept of the dot product to other vector spaces $\mathcal{V}$,

Definition

An inner product on a vector space $\mathcal{V}$ is an operation on $\mathcal{V}$ that assigns, to each pair of vectors $\mathbf{u}$ and $\mathbf{v} \in \mathcal{V}$, a scalar $\langle \mathbf{u}, \mathbf{v} \rangle$ in $\mathcal{F}$, satisfying the followings:

1. Symmetry property
$$\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$$
2. Distributive property
$$\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle, \qquad \text{where} \quad \mathbf{w} \in \mathcal{V}$$
3. Homogeneity property
$$\langle \alpha \mathbf{u}, \mathbf{v} \rangle = \alpha \langle \mathbf{u}, \mathbf{v} \rangle, \qquad \text{where } \alpha \text{ is a scalar}$$
4. Positivity property
$$\langle \mathbf{u}, \mathbf{u} \rangle \geq 0 \qquad \text{and} \quad \langle \mathbf{u}, \mathbf{u} \rangle = 0 \quad \text{if and only if} \quad \mathbf{u} = \mathbf{0}$$

- A vector space $\mathcal{V}$ with an inner product is called an inner product space.

Inner Product and Inner Product space

- A vector space $\mathcal{V}$ with an inner product is called an inner product space.
- The standard inner product for $\mathbb{R}^n$ is the dot product.

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

- Inner product of a vector space is NOT unique. e.g.

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^{n} u_i v_i w_i \qquad \text{where} \quad \mathbf{w} \in \mathbb{R}^n \quad \text{and} \quad w_i > 0 \quad \text{for} \quad \forall i.$$

- Given $\mathbf{A}$ and $\mathbf{B}$ in $\mathbb{R}^{m \times n}$, we can define an inner product by

$$\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ij}$$

- For the vector space $C[a, b]$, we may use following inner product

$$\langle f, g \rangle = \int_a^b f(x) g(x)\, dx$$

Linear functions

The function. $f : \mathbb{R}^d \to \mathbb{R}$ means that f is a function that maps real d dimensional vectors to real numbers (i.e., it is scalar valued functions of d-vectors). We can also interpret f as a function of d scalar variables:

$$f(\boldsymbol{x}) = f(x_1, x_2, \ldots, x_d)$$

Definition (Linear function)

A function f is linear if

$$f(\alpha \boldsymbol{x} + \beta \boldsymbol{y}) = \alpha f(\boldsymbol{x}) + \beta f(\boldsymbol{y})$$

For example consider the inner product function:

$$f(\boldsymbol{x}) = \boldsymbol{w}^\top \boldsymbol{x}$$

We have

$$\begin{aligned} f(\alpha \boldsymbol{x} + \beta \boldsymbol{y}) &= \boldsymbol{w}^\top (\alpha \boldsymbol{x} + \beta \boldsymbol{y}) \\ &= \boldsymbol{w}^\top (\alpha \boldsymbol{x}) + \boldsymbol{w}^\top (\beta \boldsymbol{y}) \\ &= \alpha (\boldsymbol{w}^\top \boldsymbol{x}) + \beta (\boldsymbol{w}^\top \boldsymbol{y}) \\ &= \alpha f(\boldsymbol{x}) + \beta f(\boldsymbol{y}) \end{aligned}$$

Example: polynomial regression

$$f(x) = w_0 + w_1 x + w_2 x^2 + \ldots + w_p x^p$$

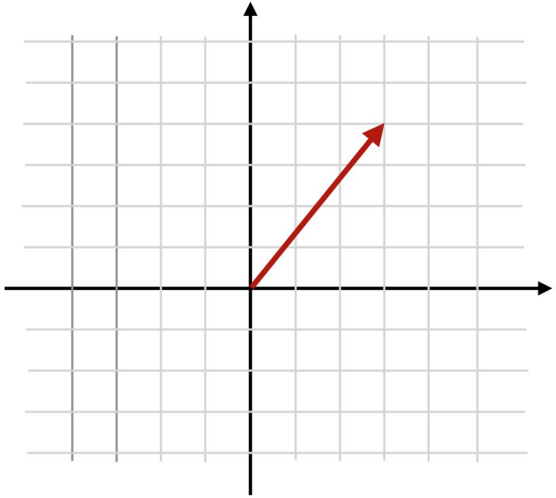
$$\boldsymbol{x} = \Phi(x) = \begin{bmatrix} 1 \\ x^1 \\ x^2 \\ \vdots \\ x^p \end{bmatrix} \qquad \boldsymbol{w} = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix}$$

Length

Definition (Length)

The length $\|\boldsymbol{x}\|_2$ of a vector $\boldsymbol{x} \in \mathbb{R}^d$ is the square root of $\boldsymbol{x}^\top \boldsymbol{x}$, i.e.,

$$\|\boldsymbol{x}\|_2 = \sqrt{\boldsymbol{x}^\top \boldsymbol{x}} = \left(x_1^2 + x_2^2 + \ldots + x_d^2\right)^{\frac{1}{2}}$$



$$\sqrt{3^2 + 4^2} = 5$$

Vector norms

Definition 3.1 (Norm). A *norm* on a vector space V is a function

$$\|\cdot\| : V \to \mathbb{R}\,, \tag{3.1}$$

$$\boldsymbol{x} \mapsto \|\boldsymbol{x}\|\,, \tag{3.2}$$

which assigns each vector $\boldsymbol{x}$ its *length* $\|\boldsymbol{x}\| \in \mathbb{R}$, such that for all $\lambda \in \mathbb{R}$ and $\boldsymbol{x}, \boldsymbol{y} \in V$ the following hold:

- *Absolutely homogeneous:* $\|\lambda\boldsymbol{x}\| = |\lambda|\|\boldsymbol{x}\|$
- *Triangle inequality:* $\|\boldsymbol{x} + \boldsymbol{y}\| \leqslant \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$
- *Positive definite:* $\|\boldsymbol{x}\| \geqslant 0$ and $\|\boldsymbol{x}\| = 0 \iff \boldsymbol{x} = \boldsymbol{0}$

Vector norms

Definition 3.1 (Norm). A *norm* on a vector space V is a function

$$\|\cdot\| : V \to \mathbb{R}, \tag{3.1}$$

$$\boldsymbol{x} \mapsto \|\boldsymbol{x}\|, \tag{3.2}$$

which assigns each vector $\boldsymbol{x}$ its *length* $\|\boldsymbol{x}\| \in \mathbb{R}$, such that for all $\lambda \in \mathbb{R}$ and $\boldsymbol{x}, \boldsymbol{y} \in V$ the following hold:

- *Absolutely homogeneous:* $\|\lambda \boldsymbol{x}\| = |\lambda| \|\boldsymbol{x}\|$
- *Triangle inequality:* $\|\boldsymbol{x} + \boldsymbol{y}\| \leqslant \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$
- *Positive definite:* $\|\boldsymbol{x}\| \geqslant 0$ and $\|\boldsymbol{x}\| = 0 \iff \boldsymbol{x} = \boldsymbol{0}$

There is a family of ℓ_p-norms parametrized by $p \in [1, \infty)$ defined by:

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{d} |x_i|^p \right)^{1/p}$$

Examples

(1) ℓ_1-norm: $\|\boldsymbol{x}\|_1 = \left(\sum_{i=1}^{d} |x_i| \right)$

(2) ℓ_2-norm: $\|\boldsymbol{x}\|_2 = \sqrt{\sum_{i=1}^{d} x_i^2}$ (Euclidean norm)

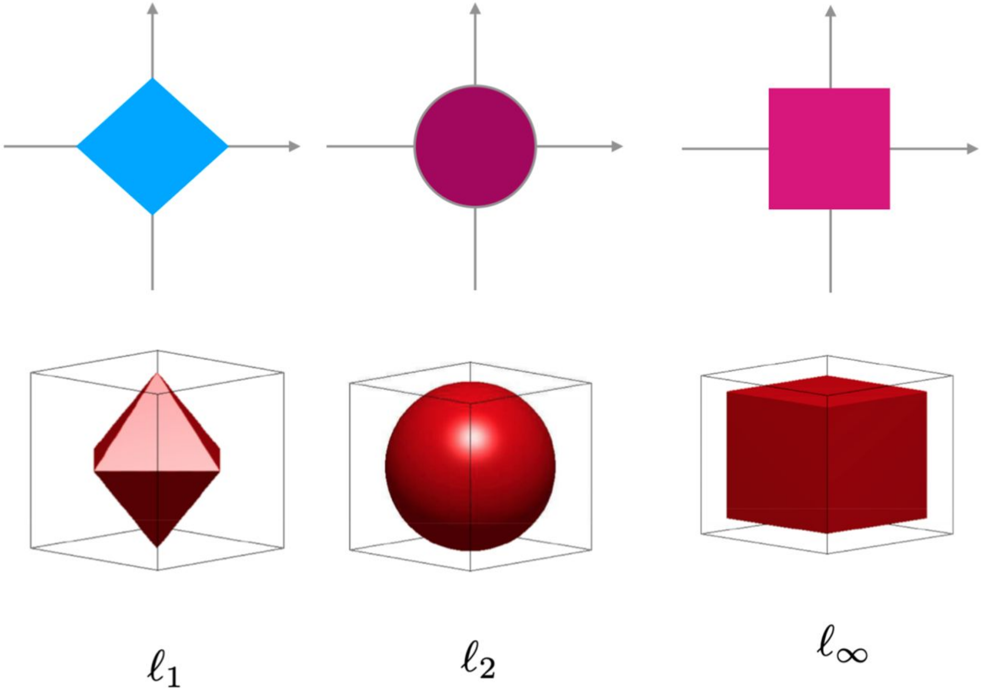
(3) ℓ_∞-norm: $\|\boldsymbol{x}\|_\infty = \max_{i=1,\ldots,d} |x_i|$

Norm balls

For every ℓ_p norm, we can define its corresponding ball as:

$$\mathbb{B}_p^d = \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_p \leq r\}$$

which is the set of all vectors with ℓ_p norm less than equal to a constant.



Angle

The standard inner product in $\mathbb{R}^d$ is related to the notion of angle between two vectors.

Definition (Angle)

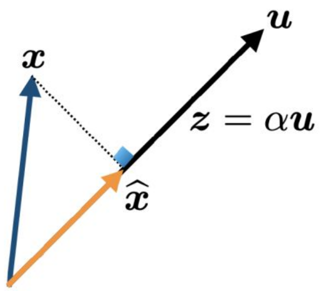
The angle between two vectors is defined by

$$\angle(\boldsymbol{x}, \boldsymbol{y}) = \arccos\left(\frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{x}\|_2 \|\boldsymbol{y}\|_2}\right)$$

Projection onto a line

Let $\boldsymbol{x}$ and $\boldsymbol{u}$ be two vectors in $\mathbb{R}^d$. We would like to find the projection of $\boldsymbol{x}$ onto $\boldsymbol{u}$ (a point on $\boldsymbol{u}$ with minimum distance to $\boldsymbol{x}$):

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{z}} \frac{1}{2}||\boldsymbol{x} - \boldsymbol{z}||_2^2, \text{ such that } \boldsymbol{z} = \alpha\boldsymbol{u}$$

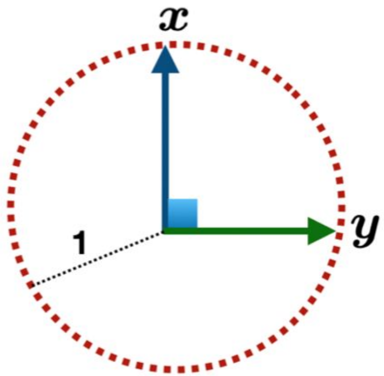


The answer is:

$$\hat{\boldsymbol{x}} = ||\hat{\boldsymbol{x}}||_2 \frac{\boldsymbol{u}}{||\boldsymbol{u}||_2} = ||\boldsymbol{x}||_2 \cos\theta \frac{\boldsymbol{u}}{||\boldsymbol{u}||_2} = \frac{<\boldsymbol{x}, \boldsymbol{u}>}{||\boldsymbol{u}||_2} \frac{\boldsymbol{u}}{||\boldsymbol{u}||_2} = \frac{(\boldsymbol{u}\boldsymbol{u}^\top)\boldsymbol{x}}{||\boldsymbol{u}||_2^2}$$

Orthogonality

Two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ in an inner product space are orthonormal if they are orthogonal and unit vectors, i.e., $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^\top \boldsymbol{y} = 0$ and $\|\boldsymbol{x}\|_2 = \|\boldsymbol{y}\|_2 = 1$.



Matrices

- A matrix is a rectangular array of numbers. For example,

$$\mathbf{A} = \begin{bmatrix} 0.3 & 1 & -5 \\ 0 & -0.2 & 16 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} r_1 & r_2 & r_3 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 1 \\ \frac{1}{2} \end{bmatrix}$$

- We denote a matrix by upper-case letters in boldface, i.e. $\mathbf{A}$
- The numbers in the brackets are called entries or elements of the matrix.

$$a_{ij}$$

- By an $m \times n$ matrix (read m by n matrix) we mean a matrix with m rows and n columns, $m \times n$ is called the size of the matrix.
- The second matrix is a square matrix, which means that it has as many rows as columns. For a square matrix, the diagonal containing the entries

$$a_{11}, a_{22}, \ldots, a_{nn}$$

is called the main diagonal of the square matrix.

- Matrices having just a single row or column are simply vectors.

Matrix Addition and Scalar Multiplication

Definition

- Addition of Matrices

  The sum of two matrices $\mathbf{A} = [a_{ij}]$ and $\mathbf{B} = [b_{ij}]$ of the same size

$$\mathbf{A} + \mathbf{B} = [a_{ij} + b_{ij}]$$

  is obtained by adding the corresponding entries of **A** and **B** together.

- Scalar Multiplication

  The product of any matrix $\mathbf{A} = [a_{ij}]$ and any scalar α is written as

$$\alpha\mathbf{A} = [\alpha a_{ij}]$$

  and is obtained by multiplying each entry of **A** by α.

Matrix Transpose

Definition

If $\mathbf{A}$ is any $m \times n$ matrix, then the transpose of $\mathbf{A}$, denoted by

$$\mathbf{A}^\top$$

is defined to be the $n \times m$ matrix that results by interchanging the columns and rows of $\mathbf{A}$; that is, the first column of $\mathbf{A}^\top$ is the first row of $\mathbf{A}$, the second column of $\mathbf{A}^\top$ is the second row of $\mathbf{A}$, and so forth.

The transpose of the following 3×4 matrix is a 4×3 matrix:

$$\boldsymbol{X} = \begin{bmatrix} 1 & 0 & 0.5 & 0 \\ 2 & 3 & 2 & 0 \\ -1 & 0 & 0 & 5 \end{bmatrix} \Rightarrow \boldsymbol{X}^\top = \begin{bmatrix} 1 & 2 & -1 \\ 0 & 3 & 0 \\ 0.5 & 2 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

Diagonal matrices

Diagonal

A matrix $\boldsymbol{D} \in \mathbb{R}^{n \times n}$ is diagonal if its entries outside the main diagonal are all zero.

$$\boldsymbol{D} = \texttt{diag}(d_1, \ldots, d_n) = \begin{bmatrix} d_1 & 0 & 0 & \ldots & 0 \\ 0 & d_2 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & d_n \end{bmatrix}$$

Symmetric matrices

Symmetric

A symmetric matrix is a square matrix that is equal to its transpose. Formally, $\boldsymbol{X} = \boldsymbol{X}^\top$. The entries of a symmetric matrix are symmetric with respect to the main diagonal, meaning $X_{i,j} = X_{j,i}$.

Dyad matrices

Dyad

A matrix $\boldsymbol{X} \in \mathbb{R}^{n\times d}$ is a dyad if it is of the form $\boldsymbol{X} = \boldsymbol{u}\boldsymbol{v}^\top$, for some vectors $\boldsymbol{u} \in \mathbb{R}^n$ and $\boldsymbol{v} \in \mathbb{R}^d$:

$$\boldsymbol{u}\boldsymbol{v}^\top = \begin{bmatrix} u_1v_1 & u_1v_2 & u_1v_3 & \dots & u_1v_d \\ u_2v_1 & u_2v_2 & u_2v_3 & \dots & u_2v_d \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ u_nv_1 & u_nv_2 & u_nv_3 & \dots & u_nv_d \end{bmatrix}$$

Matrix Multiplication

The product $\mathbf{C} = \mathbf{AB}$ of a matrix $\mathbf{A}_{m\times r} = [a_{ij}]$ and $\mathbf{B}_{p\times n} = [b_{ij}]$ is defined,

if and only if $r = p$,

to be the $m \times n$ matrix $\mathbf{C} = [c_{ij}]$ with entries

$$c_{ij} = \sum_{k=1}^{r} a_{ik}b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ir}b_{rj} \quad \text{for} \quad \begin{cases} i = 1, \cdots, m \\ j = 1, \cdots, n \end{cases}$$

- For instance,

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ a_{41} & a_{42} & a_{43} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \\ c_{41} & c_{42} \end{bmatrix}$$

- This is done by multiplying rows into columns,

$$c_{21} = a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31}$$

Matrix Multiplication

Properties of matrix multiplication

- Suppose **A**, **B** and **C** are matrices of the right size for which the indicated sums and products are defined.

  1. $\mathbf{A}(\mathbf{B}\mathbf{C}) = (\mathbf{A}\mathbf{B})\mathbf{C}$
  2. $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{A}\mathbf{B}+\mathbf{A}\mathbf{C}$
  3. $(\mathbf{B}+\mathbf{C})\mathbf{A} = \mathbf{B}\mathbf{A}+\mathbf{C}\mathbf{A}$
  4. $\alpha(\mathbf{A}\mathbf{B}) = (\alpha\mathbf{A})\mathbf{B} = \mathbf{A}(\alpha\mathbf{B})$
  5. $\mathbf{I}_m\mathbf{A} = \mathbf{A} = \mathbf{A}\mathbf{I}_n$

  where **I** is known as identity matrix, they are square matrices of a given size with ones on the main diagonal and zeros everywhere else.

$$\mathbf{I}_1 = [1], \mathbf{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \cdots \mathbf{I}_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

Matrix Multiplication

- Common mistakes regarding matrix multiplication.
  1. In general, you cannot change the order

$$\mathbf{AB} \neq \mathbf{BA}$$

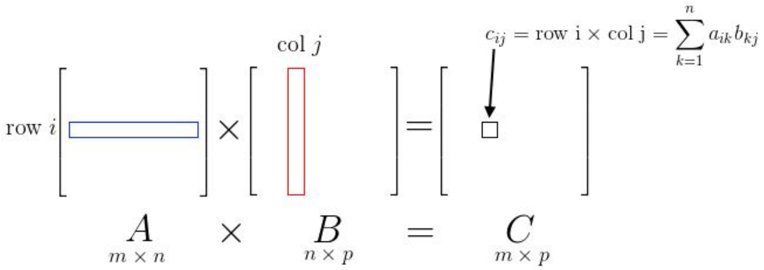
  2. In general, the cancellation laws do not hold

$$\mathbf{AB} = \mathbf{AC} \qquad \not\Rightarrow \qquad \mathbf{B} = \mathbf{C}$$

  3. In general, you cannot conclude that

$$\mathbf{AB} = \mathbf{0} \qquad \not\Rightarrow \qquad \mathbf{A} = \mathbf{0} \quad \text{or} \quad \mathbf{B} = \mathbf{0}$$

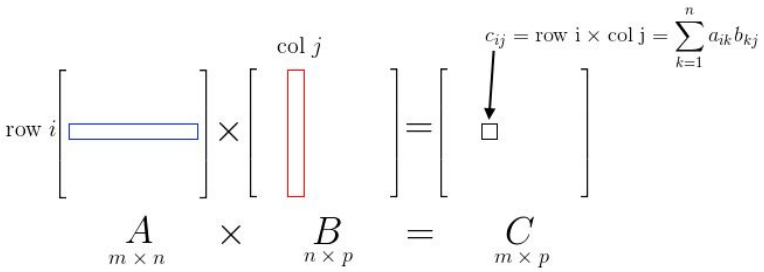
Matrix Multiplication - four different views



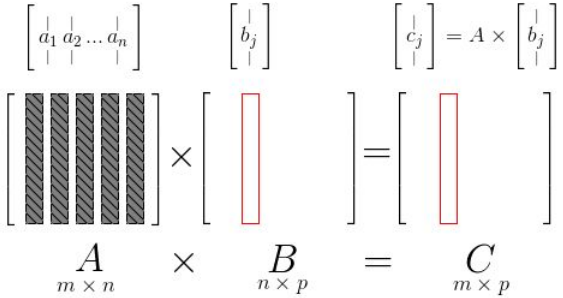
Row By Columns

http://mlwiki.org/index.php/Matrix-Matrix_Multiplication

Matrix Multiplication - four different views



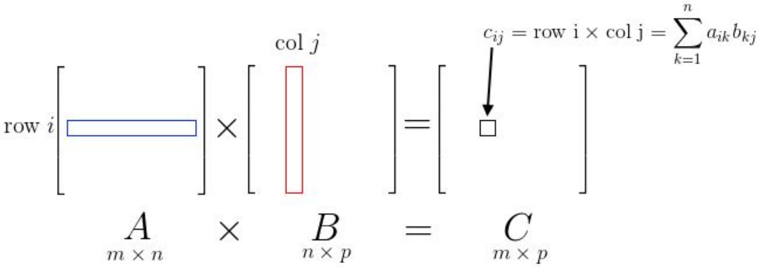
Row By Columns



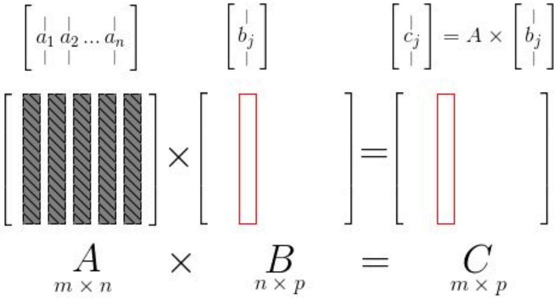
Column at a Time

http://mlwiki.org/index.php/Matrix-Matrix_Multiplication

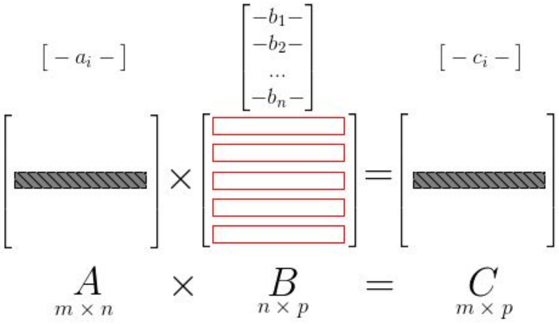
Matrix Multiplication - four different views



Row By Columns



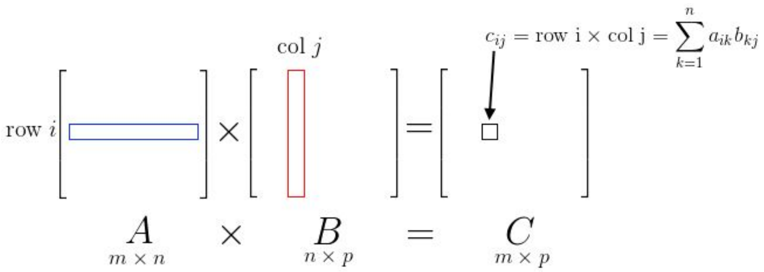
Column at a Time



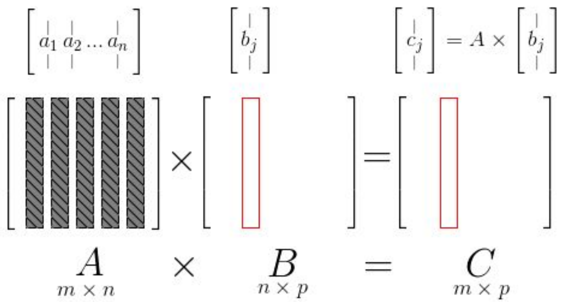
Row at a Time

http://mlwiki.org/index.php/Matrix-Matrix_Multiplication

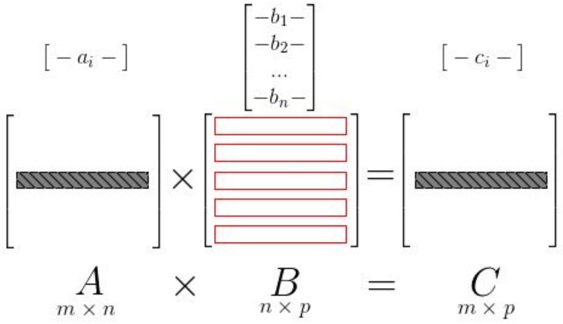
Matrix Multiplication - four different views



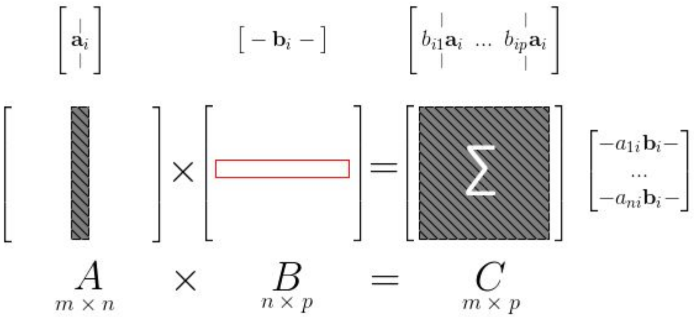
Row By Columns



Column at a Time



Row at a Time



Sum of Outer Products $C = AB = \sum_{i=1}^{n} \mathbf{a}_i \mathbf{b}_i^T$

http://mlwiki.org/index.php/Matrix-Matrix_Multiplication

Inverse Matrix

Definition

If $\mathbf{A}$ is a square matrix, and if a matrix $\mathbf{B}$ of the same size can be found such that

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I},$$

then $\mathbf{A}$ is said to be invertible (or nonsingular) and $\mathbf{B}$ is called an inverse of $\mathbf{A}$.

If no such matrix $\mathbf{B}$ can be found, then $\mathbf{A}$ is said to be singular.

Theorem

If $\mathbf{B}$ and $\mathbf{C}$ are both inverses of the matrix $\mathbf{A}$, then $\mathbf{B} = \mathbf{C}$.

Inverse Properties

For nonsingular matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$, the following properties hold:

- $\left(\boldsymbol{X}^{-1}\right)^{-1} = \boldsymbol{X}$
- The product $\boldsymbol{XY}$ is also nonsingular
- $(\boldsymbol{XY})^{-1} = \boldsymbol{Y}^{-1}\boldsymbol{X}^{-1}$ (the reverse order law of inversion)
- $\left(\boldsymbol{X}^{-1}\right)^{\top} = \left(\boldsymbol{X}^{\top}\right)^{-1}$

Determinant

- With each $n \times n$ matrix $\mathbf{A}$, it is possible to associate a scalar, denoted

  $\det(\mathbf{A})$, whose value will tell us whether the matrix is invertible.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

- Thus, if $\mathbf{A}$ is any 2×2 matrix and we define

$$\det(\mathbf{A}) = a_{11}a_{22} - a_{12}a_{21}$$

then $\mathbf{A}$ is invertible if and only if $\det(\mathbf{A}) \neq 0$.

Determinant

Definition

The determinant of an $n \times n$ matrix $\mathbf{A}$, denoted

$$\det(\mathbf{A}),$$

is a scalar associated with the matrix $\mathbf{A}$ that is defined successively as

$$\det(\mathbf{A}) = \begin{cases} a_{11} & \text{if} \quad n = 1 \\ a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n} & \text{if} \quad n > 1 \end{cases}$$

where C_{ij} is known as the cofactor for a_{ij},

$$C_{ij} = (-1)^{i+j} \det(\mathbf{M}_{ij})$$

where $\mathbf{M}_{ij}$ is known as the minor of a_{ij}, which

is the submatrix formed by deleting the i-th row and j-th column

Theorem

An $n \times n$ matrix $\mathbf{A}$ is singular if and only if

$$\det(\mathbf{A}) = 0$$

Column Space and Row Space

Definition

Given a matrix $\mathbf{A}$ of $m \times n$,

1. The column space is the set of all linear combinations of the columns of $\mathbf{A}$,

$$\mathrm{col}(\mathbf{A}) = \mathrm{span}\{\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_n\}$$

2. The row space is the set of all linear combinations of the rows of $\mathbf{A}$, denoted,

$$\mathrm{row}(\mathbf{A}) = \mathrm{span}\{\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_m\}$$

   where $\mathbf{c}_j$ and $\mathbf{r}_i$ are columns and rows of $\mathbf{A}$ respectively.

Theorem

The column space and the row space of an $m \times n$ matrix $\mathbf{A}$ are subspaces.

Rank of Matrices

Theorem

The row space and the column space of a matrix $\mathbf{A}$ have the same dimension.

Definition

The common dimension of the row space and column space of a matrix $\mathbf{A}$ is called the rank of $\mathbf{A}$ and is denoted by

$$\text{rank}(\mathbf{A})$$

Theorem

For any matrix $\mathbf{A}$,

$$\text{rank}(\mathbf{A})$$

is equal to the maximum number of linearly independent columns of $\mathbf{A}$.

Null Space and Nullity

Theorem

Given a matrix $\mathbf{A}$ of $m \times n$, the solutions of the homogeneous linear system

$$\mathbf{Ax} = \mathbf{0}$$

is a subspace of $\mathbb{R}^n$, and it is called the null space of $\mathbf{A}$, denoted

$$\text{null}(\mathbf{A})$$

Definition

The dimension of the null space of $\mathbf{A}$ is called the nullity of $\mathbf{A}$ and is denoted by

$$\text{nullity}(\mathbf{A})$$

Rank and Nullity

Dimension theorem for matrices

If $\mathbf{A}$ is a matrix with n columns, then

$$\operatorname{rank}(\mathbf{A}) + \operatorname{nullity}(\mathbf{A}) = n$$

Equivalence Theorem

If $\mathbf{A}$ is an $n \times n$ matrix, then the following statements are equivalent,

1. $\mathbf{A}$ is invertible.

Equivalence Theorem

If $\mathbf{A}$ is an $n \times n$ matrix, then the following statements are equivalent,

1. $\mathbf{A}$ is invertible.
2. $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution.

Equivalence Theorem

If $\mathbf{A}$ is an $n \times n$ matrix, then the following statements are equivalent,

1. $\mathbf{A}$ is invertible.
2. $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution.
3. The reduced echelon form of $\mathbf{A}$ is $\mathbf{I}_n$.
4. $\mathbf{A}$ is expressible as a product of elementary matrices.

Equivalence Theorem

If $\mathbf{A}$ is an $n \times n$ matrix, then the following statements are equivalent,

1. $\mathbf{A}$ is invertible.
2. $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution.
3. The reduced echelon form of $\mathbf{A}$ is $\mathbf{I}_n$.
4. $\mathbf{A}$ is expressible as a product of elementary matrices.
5. $\mathbf{Ax} = \mathbf{b}$ is consistent for every $n \times 1$ matrix $\mathbf{b}$.
6. $\mathbf{Ax} = \mathbf{b}$ has exactly one solution for every $n \times 1$ matrix $\mathbf{b}$.

Equivalence Theorem

If $\mathbf{A}$ is an $n \times n$ matrix, then the following statements are equivalent,

1. $\mathbf{A}$ is invertible.
2. $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution.
3. The reduced echelon form of $\mathbf{A}$ is $\mathbf{I}_n$.
4. $\mathbf{A}$ is expressible as a product of elementary matrices.
5. $\mathbf{Ax} = \mathbf{b}$ is consistent for every $n \times 1$ matrix $\mathbf{b}$.
6. $\mathbf{Ax} = \mathbf{b}$ has exactly one solution for every $n \times 1$ matrix $\mathbf{b}$.
7. $\det(\mathbf{A}) \neq 0$.

Equivalence Theorem

If $\mathbf{A}$ is an $n \times n$ matrix, then the following statements are equivalent,

1. $\mathbf{A}$ is invertible.
2. $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution.
3. The reduced echelon form of $\mathbf{A}$ is $\mathbf{I}_n$.
4. $\mathbf{A}$ is expressible as a product of elementary matrices.
5. $\mathbf{Ax} = \mathbf{b}$ is consistent for every $n \times 1$ matrix $\mathbf{b}$.
6. $\mathbf{Ax} = \mathbf{b}$ has exactly one solution for every $n \times 1$ matrix $\mathbf{b}$.
7. $\det(\mathbf{A}) \neq 0$.
8. The column vectors of $\mathbf{A}$ are linearly independent.
9. The row vectors of $\mathbf{A}$ are linearly independent.

Equivalence Theorem

If $\mathbf{A}$ is an $n \times n$ matrix, then the following statements are equivalent,

1. $\mathbf{A}$ is invertible.
2. $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution.
3. The reduced echelon form of $\mathbf{A}$ is $\mathbf{I}_n$.
4. $\mathbf{A}$ is expressible as a product of elementary matrices.
5. $\mathbf{Ax} = \mathbf{b}$ is consistent for every $n \times 1$ matrix $\mathbf{b}$.
6. $\mathbf{Ax} = \mathbf{b}$ has exactly one solution for every $n \times 1$ matrix $\mathbf{b}$.
7. $\det(\mathbf{A}) \neq 0$.
8. The column vectors of $\mathbf{A}$ are linearly independent.
9. The row vectors of $\mathbf{A}$ are linearly independent.
10. $\mathbf{A}$ has rank n
11. $\mathbf{A}$ has nullity 0.

Eigenvalues and Eigenvectors

Definition

Suppose $\mathbf{A}$ is an $n \times n$ matrix. A scalar λ is said to be an eigenvalue of $\mathbf{A}$ if there exists a nonzero vector $\mathbf{x}$ such that

$$\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$$

The vector $\mathbf{x}$ is said to be an eigenvector corresponding to λ.

Eigenvalues and Eigenvectors

- For example,

$$\mathbf{A} = \begin{bmatrix} 4 & -2 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

- Since

$$\mathbf{A}\mathbf{x} = \begin{bmatrix} 4 & -2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \end{bmatrix} = 3 \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 3\mathbf{x}$$

- It follows that $\lambda = 3$ is an eigenvalue of $\mathbf{A}$, and $\mathbf{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ is an eigenvector to

$$\lambda = 3$$

Eigenvalues and Eigenvectors

- For example,

$$\mathbf{A} = \begin{bmatrix} 4 & -2 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

- Since

$$\mathbf{Ax} = \begin{bmatrix} 4 & -2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \end{bmatrix} = 3 \begin{bmatrix} 2 \\ 1 \end{bmatrix} = 3\mathbf{x}$$

- It follows that $\lambda = 3$ is an eigenvalue of $\mathbf{A}$, and $\mathbf{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ is an eigenvector to

$$\lambda = 3$$

- Actually, any nonzero multiple of $\mathbf{x}$ will be a corresponding eigenvector, because

$$\mathbf{A}(\alpha\mathbf{x}) = \alpha\mathbf{Ax} = \alpha\lambda\mathbf{x} = \lambda(\alpha\mathbf{x})$$

- For example, $\begin{bmatrix} 4 \\ 2 \end{bmatrix}$ is also an eigenvector belonging to $\lambda = 3$:

$$\begin{bmatrix} 4 & -2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 12 \\ 6 \end{bmatrix} = 3 \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

Eigenvalues and Eigenvectors

- An eigenvalue problem is about finding the eigenvalue λ and the corresponding eigenvectors that satisfy the equation

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

Eigenvalues and Eigenvectors

- An eigenvalue problem is about finding the eigenvalue λ and the corresponding eigenvectors that satisfy the equation

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$

- The equation can be written in the form

$$(\mathbf{A} - \lambda\mathbf{I})\,\mathbf{x} = \mathbf{0}$$

Eigenvalues and Eigenvectors

- An eigenvalue problem is about finding the eigenvalue λ and the corresponding eigenvectors that satisfy the equation

$$\mathbf{Ax} = \lambda\mathbf{x}$$

- The equation can be written in the form

$$(\mathbf{A} - \lambda\mathbf{I})\,\mathbf{x} = \mathbf{0}$$

- Since we are looking for nonzero $\mathbf{x}$, λ is an eigenvalue of $\mathbf{A}$ if and only if

$$(\mathbf{A} - \lambda\mathbf{I})\,\mathbf{x} = \mathbf{0}$$

  has a nontrivial solution, that is

$$\mathrm{null}(\mathbf{A} - \lambda\mathbf{I}) \neq \{\mathbf{0}\}$$

- Any nonzero vector in $\mathrm{null}(\mathbf{A} - \lambda\mathbf{I})$ is an eigenvector belonging to λ.
- The subspace $\mathrm{null}(\mathbf{A} - \lambda\mathbf{I})$ is called the eigenspace corresponding to λ.

Eigenvalues and Eigenvectors

- The equation $(\mathbf{A} - \lambda \mathbf{I})\,\mathbf{x} = \mathbf{0}$ will have a nontrivial solution if and only if

$$\det(\mathbf{A} - \lambda \mathbf{I}) = 0$$

- This is known as the characteristic equation for the matrix $\mathbf{A}$.
- If the determinant is expanded, we obtain an nth-degree polynomial:

$$p(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I})$$

- This polynomial of λ is called the characteristic polynomial for the matrix $\mathbf{A}$.

Eigenvalues and Eigenvectors

- The equation $(\mathbf{A} - \lambda \mathbf{I})\,\mathbf{x} = \mathbf{0}$ will have a nontrivial solution if and only if

$$\det(\mathbf{A} - \lambda \mathbf{I}) = 0$$

- This is known as the characteristic equation for the matrix **A**.
- If the determinant is expanded, we obtain an *n*th-degree polynomial:

$$p(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I})$$

- This polynomial of λ is called the characteristic polynomial for the matrix **A**.

Ex Find the eigenvalues and the corresponding eigenvectors of the matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 2 \\ 3 & -2 \end{bmatrix}$$

Solution

$$\det(\mathbf{A} - \lambda \mathbf{I}) = 0 \implies \det \begin{bmatrix} 3-\lambda & 2 \\ 3 & -2-\lambda \end{bmatrix} \implies \lambda^2 - \lambda - 12 = 0$$

Eigenvalues and Eigenvectors

Thus, the eigenvalues of $\mathbf{A}$ are $\lambda_1 = 4$ and $\lambda_2 = -3$. To find the eigenvectors, we must determine the null space of $\mathbf{A} - \lambda\mathbf{I}$ for $\lambda_1 = 4$ and $\lambda_2 = -3$.

For $\lambda = 4$, $\mathbf{A} - 4\mathbf{I} = \begin{bmatrix} -1 & 2 \\ 3 & -6 \end{bmatrix}$
- Find a basis for the null space

$$\text{null}\left(\mathbf{A} - 4\mathbf{I}\right)$$

$$\begin{bmatrix} -1 & 2 \\ 3 & -6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$x_1 = 2x_2$$

The null space is spanned by

$$\mathbf{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix},$$

so any nonzero multiple of $\mathbf{x}$ is an eigenvector belonging to λ_1.

For $\lambda = -3$, $\mathbf{A} + 3\mathbf{I} = \begin{bmatrix} 6 & 2 \\ 3 & 1 \end{bmatrix}$
- Find a basis for the null space

$$\text{null}\left(\mathbf{A} + 3\mathbf{I}\right)$$

$$\begin{bmatrix} 6 & 2 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$-3x_1 = x_2$$

The null space is spanned by

$$\mathbf{x} = \begin{bmatrix} -1 \\ 3 \end{bmatrix},$$

so any nonzero multiple of $\mathbf{x}$ is an eigenvector belonging to λ_2.

Eigenvalues and Eigenvectors

- The characteristic polynomial of nth-degree will have exactly n roots, some of which may be repeated and some of which may be complex numbers.

Eigenvalues and Eigenvectors

- The characteristic polynomial of nth-degree will have exactly n roots, some of which may be repeated and some of which may be complex numbers.

Ex Let

$$\mathbf{A} = \begin{bmatrix} 2 & -3 & 1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix}$$

Find the eigenvalues and the corresponding eigenspaces.

Solution

$$\det \begin{bmatrix} 2-\lambda & -3 & 1 \\ 1 & -2-\lambda & 1 \\ 1 & -3 & 2-\lambda \end{bmatrix} = -\lambda(\lambda-1)^2 = 0 \implies \lambda_1 = 0, \quad \lambda_2 = \lambda_3 = 1$$

Eigenvalues and Eigenvectors

- The characteristic polynomial of nth-degree will have exactly n roots, some of which may be repeated and some of which may be complex numbers.

Ex Let

$$\mathbf{A} = \begin{bmatrix} 2 & -3 & 1 \\ 1 & -2 & 1 \\ 1 & -3 & 2 \end{bmatrix}$$

Find the eigenvalues and the corresponding eigenspaces.

Solution

$$\det \begin{bmatrix} 2-\lambda & -3 & 1 \\ 1 & -2-\lambda & 1 \\ 1 & -3 & 2-\lambda \end{bmatrix} = -\lambda(\lambda-1)^2 = 0 \implies \lambda_1 = 0, \quad \lambda_2 = \lambda_3 = 1$$

The eigenspace corresponding to $\lambda_1 = 0$ is null($\mathbf{A}$), solve $\mathbf{Ax} = \mathbf{0}$, we have

$$x_1 = x_2 = x_3$$

The eigenspace corresponding to $\lambda_1 = 0$ is the span of $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

Eigenvalues and Eigenvectors

- To find the eigenspace corresponding to $\lambda = 1$, we must solve $(\mathbf{A} - \mathbf{I})\mathbf{x} = \mathbf{0}$:

$$\begin{bmatrix} 1 & -3 & 1 & 0 \\ 1 & -3 & 1 & 0 \\ 1 & -3 & 1 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & -3 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

- Setting $x_2 = \alpha$ and $x_3 = \beta$, $x_1 = 3\alpha - \beta$,

$$\begin{bmatrix} 3\alpha - \beta \\ \alpha \\ \beta \end{bmatrix} = \alpha \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}.$$

- Thus, the eigenspace corresponding to $\lambda = 1$ is

$$\text{span}(\begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix})$$

- Hence, we have three eigenvalues and three eigenvectors for a 3×3 matrix.

Eigenvalues and Eigenvectors

Definition

The degree of a root (eigenvalue) of the characteristic polynomial of a matrix, that is the number of times the root is repeated, is called the

algebraic multiplicity of the eigenvalue.

The dimension of the eigenspace corresponding to a given λ, that is the number of linearly independent eigenvectors corresponding to the eigenvalue, is called the

geometric multiplicity of the eigenvalue.

Eigenvalues and Eigenvectors

Definition

The degree of a root (eigenvalue) of the characteristic polynomial of a matrix, that is the number of times the root is repeated, is called the

algebraic multiplicity of the eigenvalue.

The dimension of the eigenspace corresponding to a given λ, that is the number of linearly independent eigenvectors corresponding to the eigenvalue, is called the

geometric multiplicity of the eigenvalue.

- Consider $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, the characteristic polynomial of this matrix is $(1-\lambda)^2$,

$$\lambda_1 = \lambda_2 = 1 \implies \mathbf{A} - \lambda \mathbf{I} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

$$\implies \text{null}(\mathbf{A} - \lambda \mathbf{I}) = \text{span}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} \right)$$

- So the geometric multiplicity is 1 but the algebraic multiplicity is 2.

Eigenvalues and Eigenvectors

Theorem

If $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct eigenvalues of an $n \times n$ matrix $\mathbf{A}$ with corresponding eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$, then $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ are linearly independent.

Compute Determinant and Trace from Eigenvalues

Theorem 4.16. *The determinant of a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is the product of its eigenvalues, i.e.,*

$$\det(\boldsymbol{A}) = \prod_{i=1}^{n} \lambda_i \,, \tag{4.42}$$

where $\lambda_i \in \mathbb{C}$ are (possibly repeated) eigenvalues of $\boldsymbol{A}$.

Theorem 4.17. *The trace of a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is the sum of its eigenvalues, i.e.,*

$$tr(\boldsymbol{A}) = \sum_{i=1}^{n} \lambda_i \,, \tag{4.43}$$

where $\lambda_i \in \mathbb{C}$ are (possibly repeated) eigenvalues of $\boldsymbol{A}$.

Diagonalizable

Definition

An $n \times n$ matrix $\mathbf{A}$ is said to be diagonalizable if there exists a nonsingular matrix $\mathbf{P}$ and a diagonal matrix $\mathbf{D}$ such that

$$\mathbf{D} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$$

We say that $\mathbf{P}$ diagonalizes $\mathbf{A}$.

Eigen Decomposition

Theorem

An $n \times n$ matrix $\mathbf{A}$ is diagonalizable if and only if

$$\mathbf{A} \text{ has } n \text{ linearly independent eigenvectors.}$$

- If $\mathbf{A}$ is $n \times n$ and $\mathbf{A}$ has n distinct eigenvalues, then $\mathbf{A}$ is diagonalizable.
- If $\mathbf{A}$ is diagonalizable, then
  1. The column vectors of the diagonalizing matrix $\mathbf{P}$ are eigenvectors of $\mathbf{A}$.
  2. The diagonal elements of D are the corresponding eigenvalues of $\mathbf{A}$.

$$\mathbf{A} = \mathbf{PDP}^{-1}$$

Eigen Decomposition

Ex Find a matrix $\mathbf{P}$ that diagonalizes $\mathbf{A} = \begin{bmatrix} 2 & -3 \\ 2 & -5 \end{bmatrix}$ and the corresponding diagonal matrix $\mathbf{D}$ that is similar to $\mathbf{A}$ if it is possible, explain otherwise.

Solution

The eigenvalues of $\mathbf{A}$ are $\lambda_1 = 1$ and $\lambda_2 = -4$. Corresponding to λ_1 and λ_2, we have the eigenvectors $\mathbf{x}_1 = \begin{bmatrix} 3 & 1 \end{bmatrix}^\mathsf{T}$ and $\mathbf{x}_2 = \begin{bmatrix} 1 & 2 \end{bmatrix}^\mathsf{T}$, so

$$\mathbf{P} = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} \qquad \text{and} \qquad \mathbf{D} = \begin{bmatrix} 1 & 0 \\ 0 & -4 \end{bmatrix}$$

You can easily verify that $\mathbf{A} = \mathbf{PDP}^{-1}$.

- The diagonalizing matrix $\mathbf{P}$ is not unique.

Eigen Decomposition

- If the eigenvalues are not distinct, then **A** may or may not be diagonalizable, depending on whether **A** has n linearly independent eigenvectors.

Eigen Decomposition

- If the eigenvalues are not distinct, then $\mathbf{A}$ may or may not be diagonalizable, depending on whether $\mathbf{A}$ has n linearly independent eigenvectors.

Ex Consider $\mathbf{A} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 1 & 0 & 2 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 2 & 0 & 0 \\ -1 & 4 & 0 \\ -3 & 6 & 2 \end{bmatrix}$, are they diagonalizable?

Solution

$\mathbf{A}$ and $\mathbf{B}$ both have the same eigenvalues, $\lambda_1 = 4$, $\lambda_2 = \lambda_3 = 2$. The eigenspace of $\mathbf{A}$ corresponding to $\lambda_1 = 4$ is spanned by $\mathbf{e}_2$ and the eigenspace corresponding to $\lambda = 2$ is spanned by $\mathbf{e}_3$. So $\mathbf{A}$ is not diagonalizable because $\mathbf{A}$ has only two linearly independent eigenvectors instead of three. On the other hand, $\mathbf{B}$ has three

$$\mathbf{x}_1 = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}, \qquad \mathbf{x}_2 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \qquad \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Therefore $\mathbf{B}$ is diagonalizable.

Eigen Decomposition for Symmetric Matrix (Spectral Theorem)

☞ Every symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{d\times d}$ guaranteed to have decomposition

with real **eigenvalues**: $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^\top = \sum_{i=1}^{d} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^\top$

$$\boxed{\boldsymbol{A} \in \mathbb{R}^{d\times d}} = \boxed{\boldsymbol{U} \in \mathbb{R}^{d\times d}} \times \boxed{\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}} \times \boxed{\boldsymbol{U}^\top \in \mathbb{R}^{d\times d}}$$

- **Real eigenvalues:** $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d$ $\quad\quad \boldsymbol{\Sigma} = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$
- the columns of $\boldsymbol{U}$, i.e., $\boldsymbol{u}_i$ are eigenvectors of $\boldsymbol{A}$
  $\boldsymbol{U} \in \mathbb{R}^{n\times n}$ is an unitary matrix $(\boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I})$
- **Correspondence of eigenvalues/eigenvectors:**

$$\boldsymbol{A}\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i, i = 1, 2, \ldots, d$$

Eigen Decomposition for Symmetric Matrix

Ex Find a matrix $\mathbf{P}$ that orthogonally diagonalizes $\mathbf{A} = \begin{bmatrix} 0 & 2 & -1 \\ 2 & 3 & -2 \\ -1 & -2 & 0 \end{bmatrix}$.

Solution

- The eigenvalues of $\mathbf{A}$ are $\lambda_1 = \lambda_2 = -1$ and $\lambda_3 = 5$.
- Computing eigenvectors in the usual way, for $\lambda = 1$, we have

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{x}_2 = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$$

- The Gram–Schmidt gives the following orthonormal basis for the eigenspace corresponding to $\lambda = -1$.

$$\mathbf{u}_1 = \frac{1}{\|\mathbf{x}_1\|}\mathbf{x}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & 1 \end{bmatrix}^\mathsf{T}$$

$$\mathbf{u}_2 = \frac{1}{\left\|\mathbf{x}_2 - \left(\mathbf{x}_2^\mathsf{T}\mathbf{u}_1\right)\mathbf{u}_1\right\|}\left(\mathbf{x}_2 - \left(\mathbf{x}_2^\mathsf{T}\mathbf{u}_1\right)\mathbf{u}_1\right) = \frac{1}{\sqrt{3}} \begin{bmatrix} -1 & 1 & 1 \end{bmatrix}^\mathsf{T}$$

Eigen Decomposition for Symmetric Matrix

Solution

- The eigenspace corresponding to $\lambda_3 = 5$ is spanned by $\mathbf{x}_3 = \begin{bmatrix} -1 & -2 & 1 \end{bmatrix}^\mathsf{T}$.
- $\mathbf{x}_3$ must be orthogonal to $\mathbf{u}_1$ and $\mathbf{u}_2$ since $\mathbf{A}$ is symmetric (Hermitian), so we only need normalize

$$\mathbf{u}_3 = \frac{1}{\|\mathbf{x}_3\|}\mathbf{x}_3 = \frac{1}{\sqrt{6}} \begin{bmatrix} -1 & -2 & 1 \end{bmatrix}^\mathsf{T}$$

- Thus, $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthonormal set and

$$\mathbf{P} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{3}} & -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \end{bmatrix}$$

orthogonally diagonalizes $\mathbf{A}$,

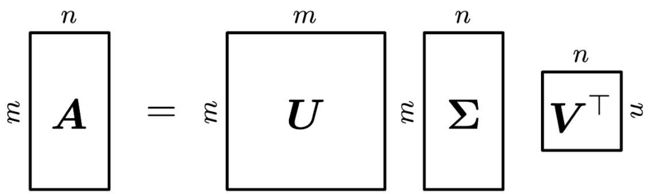
$$\mathbf{D} = \mathbf{P}^\mathsf{T}\mathbf{A}\mathbf{P}$$

What about any non-square matrix? SVD!

It has been referred to as the “fundamental theorem of linear algebra” (Strang, 1993) because it can be applied to all matrices, not only to square matrices, and it always exists.

What about any non-square matrix? SVD!

Theorem 4.22 (SVD Theorem)**.** *Let* $\boldsymbol{A}^{m\times n}$ *be a rectangular matrix of rank* $r \in [0, \min(m,n)]$*. The SVD of* $\boldsymbol{A}$ *is a decomposition of the form*

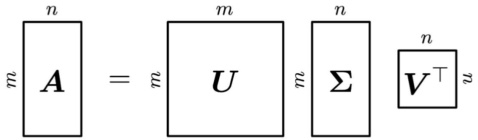


(4.64)

with an orthogonal matrix $\boldsymbol{U} \in \mathbb{R}^{m\times m}$ *with column vectors* $\boldsymbol{u}_i$*,* $i = 1, \ldots, m$*, and an orthogonal matrix* $\boldsymbol{V} \in \mathbb{R}^{n\times n}$ *with column vectors* $\boldsymbol{v}_j$*,* $j = 1, \ldots, n$*. Moreover,* $\boldsymbol{\Sigma}$ *is an* $m \times n$ *matrix with* $\Sigma_{ii} = \sigma_i \geqslant 0$ *and* $\Sigma_{ij} = 0,\ i \neq j$*.*

It has been referred to as the “fundamental theorem of linear algebra” (Strang, 1993) because it can be applied to all matrices, not only to square matrices, and it always exists.

Theorem 4.22 (SVD Theorem). *Let $A^{m \times n}$ be a rectangular matrix of rank $r \in [0, \min(m, n)]$. The SVD of A is a decomposition of the form*



(4.64)

with an orthogonal matrix $U \in \mathbb{R}^{m \times m}$ with column vectors u_i, $i = 1, \ldots, m$, and an orthogonal matrix $V \in \mathbb{R}^{n \times n}$ with column vectors v_j, $j = 1, \ldots, n$. Moreover, Σ is an $m \times n$ matrix with $\Sigma_{ii} = \sigma_i \geqslant 0$ and $\Sigma_{ij} = 0,\ i \neq j$.

The diagonal entries σ_i, $i = 1, \ldots, r$, of Σ are called the *singular values*, u_i are called the *left-singular vectors*, and v_j are called the *right-singular vectors*. By convention, the singular values are ordered, i.e., $\sigma_1 \geqslant \sigma_2 \geqslant \sigma_r \geqslant 0$.

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n \\ 0 & \ldots & 0 \\ \vdots & & \vdots \\ 0 & \ldots & 0 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} \sigma_1 & 0 & 0 & 0 & \ldots & 0 \\ 0 & \ddots & 0 & 0 & & 0 \\ 0 & 0 & \sigma_m & 0 & \ldots & 0 \end{bmatrix}$$

SVD from Eigen Decomposition of $\boldsymbol{A}^\top \boldsymbol{A}$ and $\boldsymbol{A}\boldsymbol{A}^\top$

$$\boldsymbol{A}^\top \boldsymbol{A} = (\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top)^\top(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top) = \boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{U}^\top\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$$
$$= \boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{\Sigma}\boldsymbol{V}^\top = \boldsymbol{V}\begin{bmatrix}\sigma_1^2 & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \sigma_n^2\end{bmatrix}\boldsymbol{V}^\top$$

$$\boldsymbol{A}\boldsymbol{A}^\top = (\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top)(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top)^\top = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top\boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{U}^\top$$
$$= \boldsymbol{U}\begin{bmatrix}\sigma_1^2 & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \sigma_m^2\end{bmatrix}\boldsymbol{U}^\top .$$

Positive (semi-)definite matrices

Let A be a real, symmetric $d \times d$ matrix. We say A is *positive semi-definite* (PSD) if, for all $\mathbf{x} \in \mathbb{R}^d$, $\mathbf{x}^T A\mathbf{x} \geq 0$. We say A is *positive definite* (PD) if, for all $\mathbf{x} \neq \mathbf{0}$, $\mathbf{x}^T A\mathbf{x} > 0$. We write $A \succeq 0$ when A is PSD, and $A \succ 0$ when A is PD.

$\boldsymbol{A}^\top \boldsymbol{A}$ is positive semi-definite

Theorem 4.14. *Given a matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, we can always obtain a symmetric, positive semidefinite matrix $\boldsymbol{S} \in \mathbb{R}^{n\times n}$ by defining*

$$\boldsymbol{S} := \boldsymbol{A}^\top \boldsymbol{A}\,. \tag{4.36}$$

Remark. If $\mathrm{rk}(\boldsymbol{A}) = n$, then $\boldsymbol{S} := \boldsymbol{A}^\top \boldsymbol{A}$ is symmetric, positive definite. $\diamondsuit$

Understanding why Theorem 4.14 holds is insightful for how we can use symmetrized matrices: Symmetry requires $\boldsymbol{S} = \boldsymbol{S}^\top$, and by inserting (4.36) we obtain $\boldsymbol{S} = \boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{A}^\top (\boldsymbol{A}^\top)^\top = (\boldsymbol{A}^\top \boldsymbol{A})^\top = \boldsymbol{S}^\top$. Moreover, positive semidefiniteness (Section 3.2.3) requires that $\boldsymbol{x}^\top \boldsymbol{S}\boldsymbol{x} \geqslant 0$ and inserting (4.36) we obtain $\boldsymbol{x}^\top \boldsymbol{S}\boldsymbol{x} = \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} = (\boldsymbol{x}^\top \boldsymbol{A}^\top)(\boldsymbol{A}\boldsymbol{x}) = (\boldsymbol{A}\boldsymbol{x})^\top (\boldsymbol{A}\boldsymbol{x}) \geqslant 0$, because the dot product computes a sum of squares (which are themselves non-negative).

Positive (semi-)definite matrices

Let A be a real, symmetric $d \times d$ matrix. We say A is *positive semi-definite* (PSD) if, for all $\mathbf{x} \in \mathbb{R}^d$, $\mathbf{x}^T A\mathbf{x} \geq 0$. We say A is *positive definite* (PD) if, for all $\mathbf{x} \neq \mathbf{0}$, $\mathbf{x}^T A\mathbf{x} > 0$. We write $A \succeq 0$ when A is PSD, and $A \succ 0$ when A is PD.

Theorem

(a) A is PSD iff $\lambda_i \geq 0$ for each i.

(b) A is PD iff $\lambda_i > 0$ for each i.

$\boldsymbol{A}^\top \boldsymbol{A}$ is positive semi-definite

Theorem 4.14. *Given a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, we can always obtain a symmetric, positive semidefinite matrix $\boldsymbol{S} \in \mathbb{R}^{n \times n}$ by defining*

$$\boldsymbol{S} := \boldsymbol{A}^\top \boldsymbol{A}\,. \tag{4.36}$$

Remark. If $\operatorname{rk}(\boldsymbol{A}) = n$, then $\boldsymbol{S} := \boldsymbol{A}^\top \boldsymbol{A}$ is symmetric, positive definite.

Because $\boldsymbol{A}^\top \boldsymbol{A}$ has non-negative eigenvalues!

CMPSC 448: Machine Learning

Lecture 3.2 Vector Calculus

Rui Zhang

Fall 2020

PennState

Scalar valued functions of scalars

We are all familiar with basic calculus, and functions of the form $f : \mathbb{R} \to \mathbb{R}$ and their derivatives:

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$

The derivative of a function of a real variable measures the sensitivity to change of the function value (output value) with respect to a change in its argument (input value).

Scalar valued functions of scalars

We are all familiar with basic calculus, and functions of the form $f : \mathbb{R} \to \mathbb{R}$ and their derivatives:

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$

The derivative of a function of a real variable measures the sensitivity to change of the function value (output value) with respect to a change in its argument (input value).

- $f(x) = x^r$, then $f'(x) = rx^{r-1}$,
- $\frac{d}{dx} e^x = e^x$.

Scalar valued functions of scalars

We are all familiar with basic calculus, and functions of the form $f : \mathbb{R} \to \mathbb{R}$ and their derivatives:

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$

The derivative of a function of a real variable measures the sensitivity to change of the function value (output value) with respect to a change in its argument (input value).

- $f(x) = x^r$, then $f'(x) = rx^{r-1}$,
- $\frac{d}{dx} e^x = e^x$.
- Sums rule: $(\alpha f + \beta g)' = \alpha f' + \beta g'$ for all functions f and g and all real numbers α and β
- Product rule: $(fg)' = f'g + fg'$ for all functions f and g.
- Chain rule: If $f(x) = h(g(x))$, then

$$f'(x) = h'(g(x)) \cdot g'(x).$$

Scalar valued functions of vectors

Multivariable calculus (also known as multivariate calculus) is the extension of calculus in one variable to calculus with functions of several variables, $f : \mathbb{R}^d \to \mathbb{R}$, e.g.,

$$f(x, y) = \frac{x^2 y}{x^4 + y^2}$$

In many machine learning applications, our model can be modeled as a function f that takes d features as inputs and maps it to a real number (regression) or binary variable (classification), e.g.,

Scalar valued functions of vectors

Multivariable calculus (also known as multivariate calculus) is the extension of calculus in one variable to calculus with functions of several variables, $f : \mathbb{R}^d \to \mathbb{R}$, e.g.,

$$f(x, y) = \frac{x^2 y}{x^4 + y^2}$$

In many machine learning applications, our model can be modeled as a function f that takes d features as inputs and maps it to a real number (regression) or binary variable (classification), e.g.,

- Linear regressions: let's $\boldsymbol{x} \in \mathbb{R}^d$ be d features of a samples and $\boldsymbol{w} \in \mathbb{R}^d$ be parameter vector of of a linear model f, then the prediction is:

$$f(\boldsymbol{x}) = \boldsymbol{w}^\top \boldsymbol{x} = \sum_{i=1}^{d} w_i x_i$$

So, we need to have a basic understanding of multivariable calculus.

Gradient (first order)

Definition

Let $f : \mathcal{C} \subseteq \mathbb{R}^d \to \mathbb{R}$ be a differentiable function. Then, the gradient of f at $\boldsymbol{x} \in \mathcal{C}$ is the vector in $\mathbb{R}^d$ denoted by $\nabla f(\boldsymbol{x})$ and defined by

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\boldsymbol{x}) \\ \vdots \\ \frac{\partial f}{\partial x_d}(\boldsymbol{x}) \end{bmatrix}$$

Gradient (first order)

Definition

Let $f : \mathcal{C} \subseteq \mathbb{R}^d \to \mathbb{R}$ be a differentiable function. Then, the gradient of f at $\boldsymbol{x} \in \mathcal{C}$ is the vector in $\mathbb{R}^d$ denoted by $\nabla f(\boldsymbol{x})$ and defined by

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\boldsymbol{x}) \\ \vdots \\ \frac{\partial f}{\partial x_d}(\boldsymbol{x}) \end{bmatrix}$$

An an example, let's consider the function $f(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^\top \boldsymbol{y} = \sum_{i=1}^d x_i y_i$, then

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial \sum_{i=1}^d x_i y_i}{\partial x_1} \\ \vdots \\ \frac{\partial \sum_{i=1}^d x_i y_i}{\partial x_d} \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_d \end{bmatrix} = \boldsymbol{y}$$

Hessian (second order)

Definition

Let $f : \mathcal{C} \subseteq \mathbb{R}^d \to \mathbb{R}$ be a twice differentiable function. Then, the Hessian of f at $\boldsymbol{x} \in \mathcal{C}$ is the vector in $\mathbb{R}^d$ denoted by $\nabla^2 f(\boldsymbol{x})$ and defined by

$$\nabla^2 f(\boldsymbol{x}) = \left[\frac{\partial^2 f}{\partial x_i x_j}(\boldsymbol{x})\right]_{1\le i,j\le d} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\boldsymbol{x}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_d} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_2^2}(\boldsymbol{x}) & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2 f}{\partial x_d \partial x_1}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_d \partial x_2}(\boldsymbol{x}) & \cdots & \frac{\partial^2 f}{\partial x_d^2} \end{bmatrix}$$

Hessian (second order)

Definition

Let $f : \mathcal{C} \subseteq \mathbb{R}^d \to \mathbb{R}$ be a twice differentiable function. Then, the Hessian of f at $\boldsymbol{x} \in \mathcal{C}$ is the vector in $\mathbb{R}^d$ denoted by $\nabla^2 f(\boldsymbol{x})$ and defined by

$$\nabla^2 f(\boldsymbol{x}) = \left[\frac{\partial^2 f}{\partial x_i x_j}(\boldsymbol{x})\right]_{1 \le i,j \le d} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\boldsymbol{x}) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_d} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_2^2}(\boldsymbol{x}) & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial^2 f}{\partial x_d \partial x_1}(\boldsymbol{x}) & \frac{\partial^2 f}{\partial x_d \partial x_2}(\boldsymbol{x}) & \dots & \frac{\partial^2 f}{\partial x_d^2} \end{bmatrix}$$

$$f(x,y) = x^3 - 2xy - y^6 \xrightarrow{\textbf{Hessian}} \begin{bmatrix} 6x & -2 \\ -2 & -30y^4 \end{bmatrix}$$