

Lecture 02. Linear Algebra

Xin Chen

These slides are based on slides from Mahdi Roozbahani

Some logistics

- Please make your decision ASAP if you want to drop
- Creating team for your project.
- Two TA: Wendi Ren and Hua Jiang.
- Please be nice to our TA team.
- Office hours will be started from next week.
- A schedule is posted on the website.

Outline

- Linear Algebra Basics ←
- Norms
- Multiplications
- Matrix Inversion
- Trace and Determinant
- Eigen Values and Eigen Vectors
- Singular Value Decomposition
- Matrix Calculus

Why Linear Algebra?

- Linear algebra provides a way of compactly representing and operating on sets of linear equations

$$4x_1 - 5x_2 = -13 \quad -2x_1 + 3x_2 = 9$$

can be written in the form of $Ax = b$

Data

$$\underbrace{A}_{\begin{matrix} x_1 \\ x_2 \end{matrix}} = \begin{bmatrix} 4 & -5 \\ -2 & 3 \end{bmatrix} \quad b = \begin{bmatrix} -13 \\ 9 \end{bmatrix}$$

.1

- $A \in \mathbb{R}^{m \times n}$ denotes a matrix with m rows and n columns, where elements belong to real numbers.
- $x \in \mathbb{R}^n$ denotes a vector with n real entries. By convention an n dimensional vector is often thought as a matrix with n rows and 1 column.

Linear Algebra Basics

- Transpose of a matrix results from flipping the rows and columns. Given $A \in \mathbb{R}^{m \times n}$, transpose is $A^T \in \mathbb{R}^{n \times m}$

- For each element of the matrix, the transpose can be written as $\rightarrow A^T_{ij} = A_{ji}$
- The following properties of the transposes are easily verified
 - $(A^T)^T = A$
 - $(AB)^T = B^T A^T$
 - $(A + B)^T = A^T + B^T$
- A square matrix $A \in \mathbb{R}^{n \times n}$ is symmetric if $A = A^T$ and it is anti-symmetric if $A = -A^T$. Thus each matrix can be written as a sum of symmetric and anti-symmetric matrices.

$$A = \frac{1}{2} (A + A^T) + \frac{1}{2} (A - A^T)$$

$$B = A + A^T$$

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

$$C = A - A^T$$

$$= A^T - A$$

$$C^T = (A - A^T)^T = A^T - A$$

$$= -C$$

Outline

- Linear Algebra Basics
- Norms 
- Multiplications
- Matrix Inversion
- Trace and Determinant
- Eigen Values and Eigen Vectors
- Singular Value Decomposition

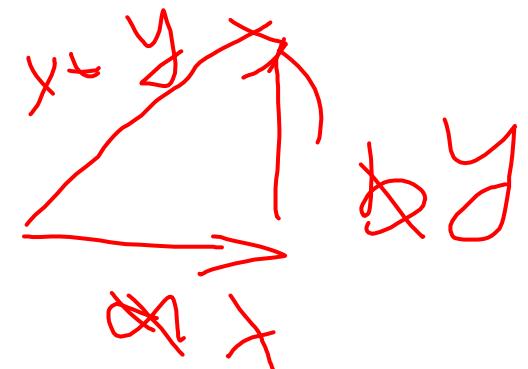
Norms

- Norm of a vector $\|x\|$ is informally a measure of the “length” of a vector
- More formally, a norm is any function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ that satisfies
 - For all $x \in \mathbb{R}^n$, $f(x) \geq 0$ (non-negativity)
 - $f(x) = 0$ if and only if $x = 0$ (definiteness)
 - For $x \in \mathbb{R}^n$, $t \in \mathbb{R}$, $f(tx) = |t|f(x)$ (homogeneity)
 - For all $x, y \in \mathbb{R}^n$, $f(x + y) \leq f(x) + f(y)$ (triangle inequality)

- Common norms used in machine learning are

- ℓ_2 norm

- $$\|\boldsymbol{x}\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$$

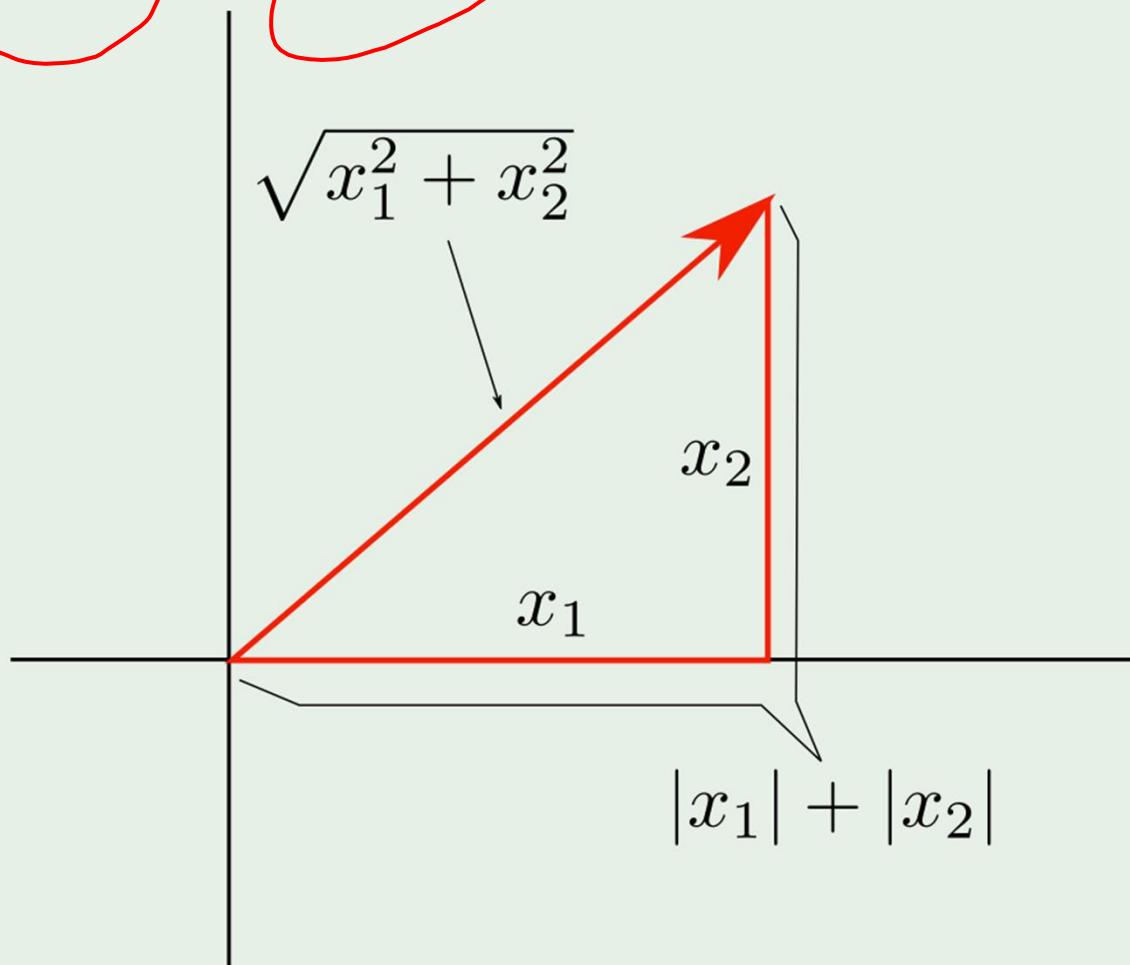


Norms

- ℓ_1 norm
 - $\|x\|_1 = \sum_{i=1}^n |x_i|$
- ℓ_∞ norm
 - $\|x\|_\infty = \max_i |x_i|$
- All norms presented so far are examples of the family of ℓ_p norms, which are parameterized by a real number $p \geq 1$
 - $\|x\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}$ P \in \mathbb{L}
- Norms can be defined for matrices, such as the Frobenius norm.
 - $\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |A_{ij}|^2} = \sqrt{\text{tr}(A^\top A)}$

Vector Norm Examples

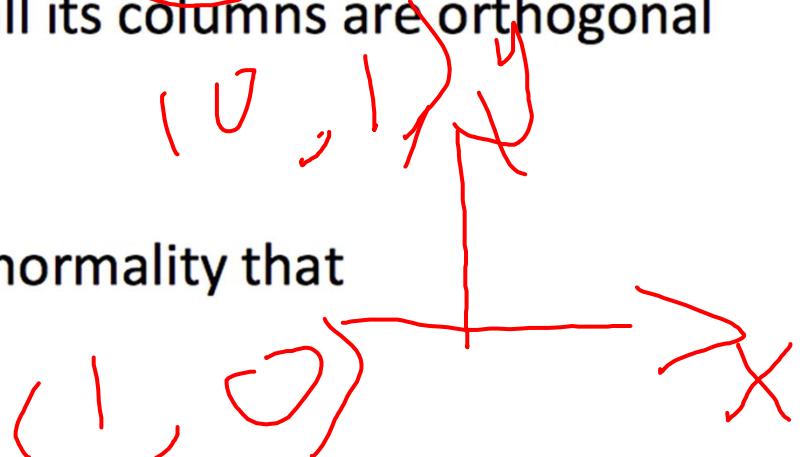
Example ℓ_1 -norm and ℓ_2 -norm



Special Matrices

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

- The identity matrix, denoted by $I \in \mathbb{R}^{n \times n}$ is a square matrix with ones on the diagonal and zeros everywhere else
- A diagonal matrix is matrix where all non-diagonal matrices are 0. This is typically denoted as $D = \text{diag}(d_1, d_2, d_3, \dots, d_n)$
- Two vectors $x, y \in \mathbb{R}^n$ are orthogonal if $x \cdot y = 0$. A square matrix $U \in \mathbb{R}^{n \times n}$ is orthogonal if all its columns are orthogonal to each other and are normalized
 - It follows from orthogonality and normality that
 - $\underline{U^T U = I = U U^T}$
 - $\|Ux\|_2 = \|x\|_2$



Outline

- Linear Algebra Basics
- Norms
- Multiplications 
- Matrix Inversion
- Trace and Determinant
- Eigen Values and Eigen Vectors
- Singular Value Decomposition

Multiplications

- The product of two matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$ is given by $C \in \mathbb{R}^{m \times p}$, where $C_{ij} = \sum_{k=1}^n A_{ik}B_{kj}$
- Given two vectors $x, y \in \mathbb{R}^n$, the term $x^\top y$ (also $x \cdot y$) is called the **inner product** or **dot product** of the vectors, and is a real number given by $\sum_{i=1}^n x_i y_i$. For example,

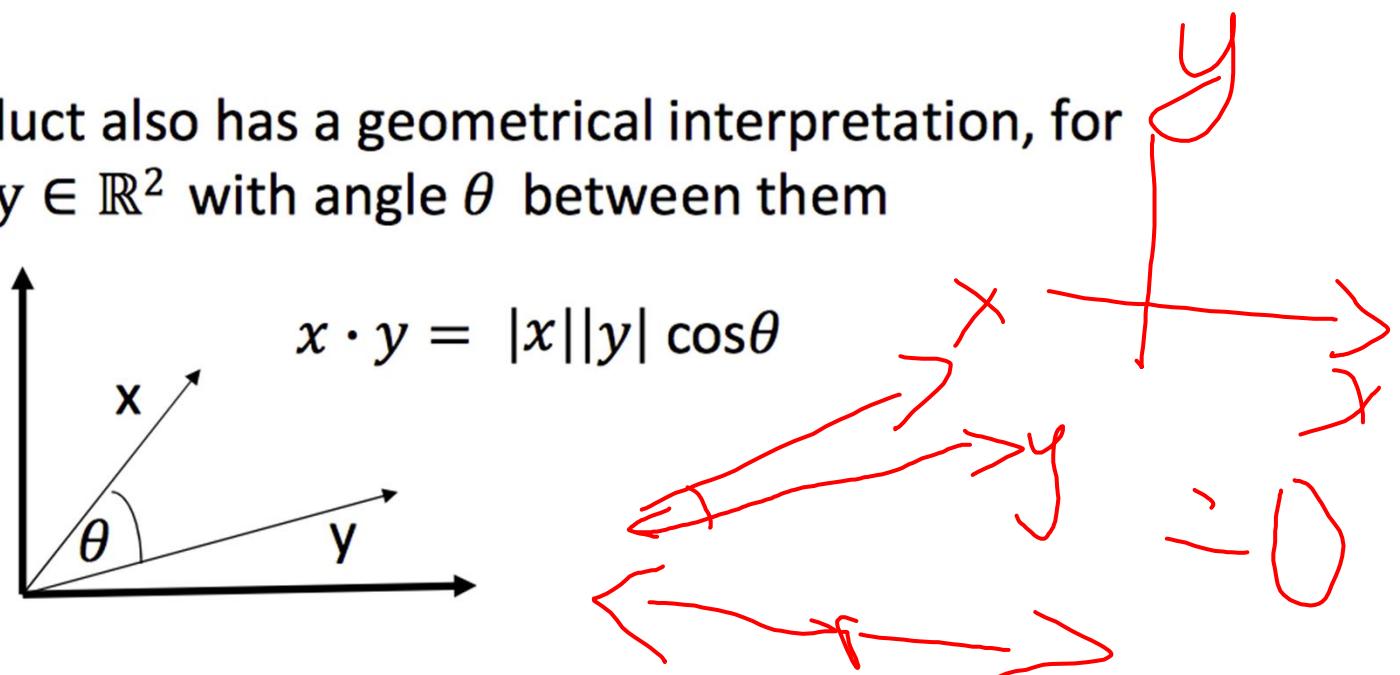
$$x^\top y = [x_1 \quad x_2 \quad x_3] \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \sum_{i=1}^3 x_i y_i$$

- Given two vectors $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, the term xy^\top is called the **outer product** of the vectors, and is a matrix given by $(x_i y_j)^\top = x_i y_j$. For example,

Multiplications

$$xy^T = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} [y_1 \quad y_2 \quad y_3] = \begin{bmatrix} x_1y_1 & x_1y_2 & x_1y_3 \\ x_2y_1 & x_2y_2 & x_2y_3 \\ x_3y_1 & x_3y_2 & x_3y_3 \end{bmatrix}$$

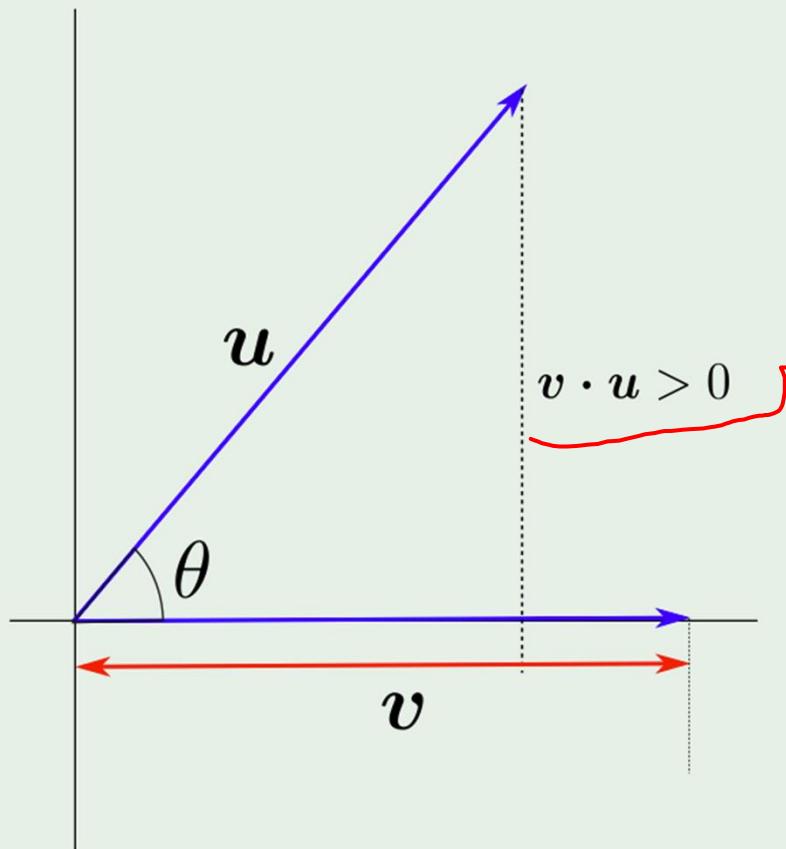
- The dot product also has a geometrical interpretation, for vectors in $x, y \in \mathbb{R}^2$ with angle θ between them



which leads to use of dot product for testing orthogonality, getting the Euclidean norm of a vector, and scalar projections.

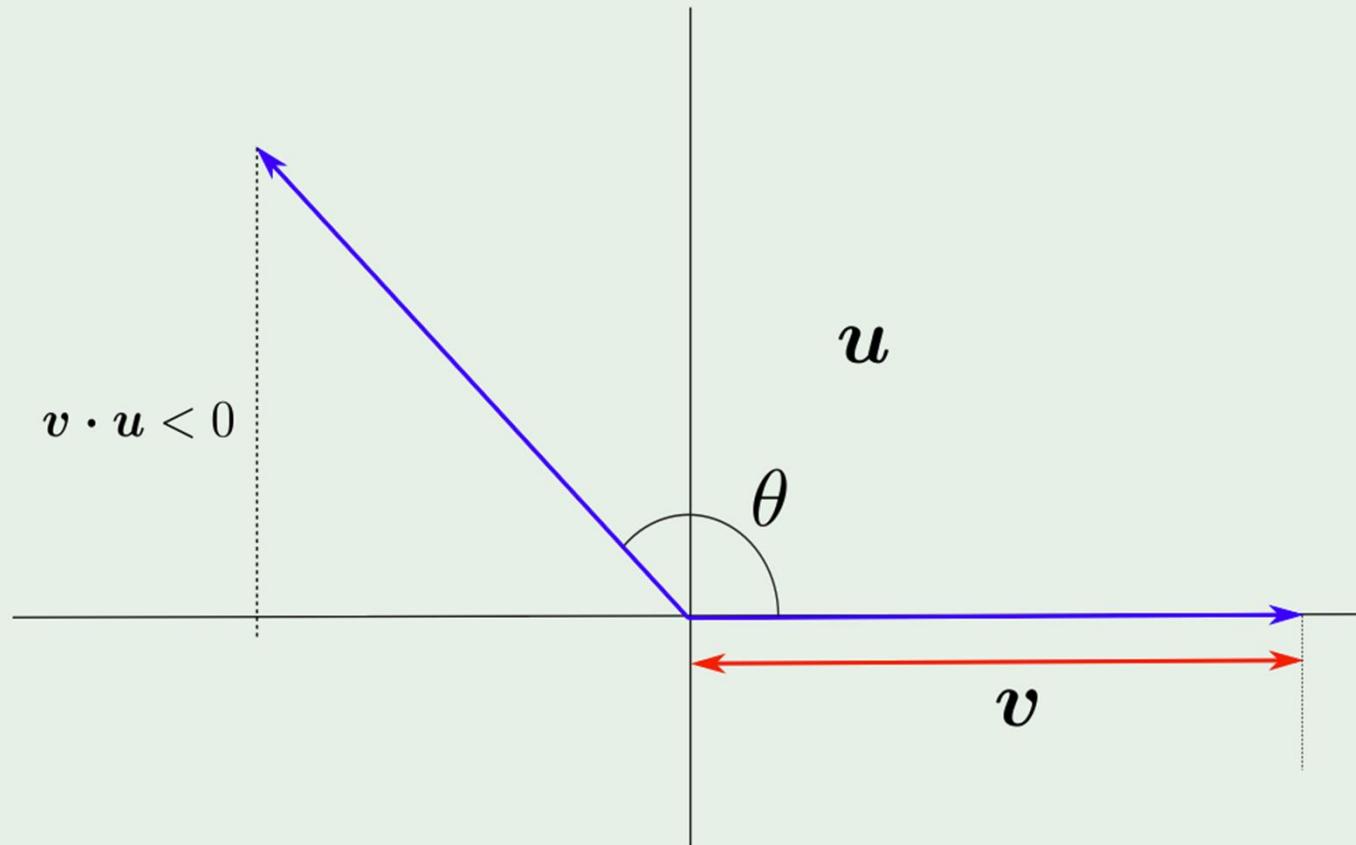
Inner Product Properties

The inner product is a measure of correlation between two vectors, scaled by the norms of the vectors



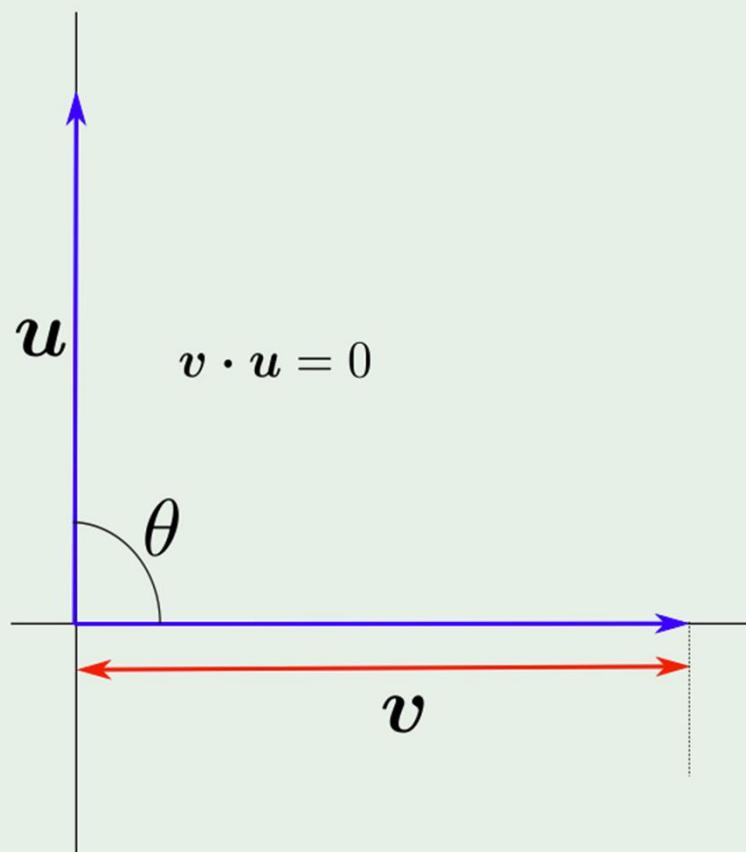
Inner Product Properties

The inner product is a measure of correlation between two vectors, scaled by the norms of the vectors



Inner Product Properties

The inner product is a measure of correlation between two vectors, scaled by the norms of the vectors



Outline

- Linear Algebra Basics
- Norms
- Multiplications
- Matrix Inversion 
- Trace and Determinant
- Eigen Values and Eigen Vectors
- Matrix Decomposition

Linear Independence and Matrix Rank

- A set of vectors $\{x_1, x_2, \dots, x_n\} \subset \mathbb{R}^m$ are said to be (linearly independent) if no vector can be represented as a linear combination of the remaining vectors. That is if

$$x_n = \sum_{i=1}^{n-1} \alpha_i x_i$$

for some scalar values $\alpha_1, \alpha_2, \dots \in \mathbb{R}$ then we say that the vectors are linearly **dependent**; otherwise the vectors are linearly independent

- The **column rank** of a matrix $A \in \mathbb{R}^{m \times n}$ is the size of the largest subset of columns of A that constitute a linearly independent set. **Row rank** of a matrix is defined similarly for rows of a matrix.

Range and Null Space

- The span of a set of vectors $\{x_1, x_2, \dots, x_n\}$ is the set of all vectors that can be expressed as a linear combination of the set $\{v: v = \sum_{i=1}^n \alpha_i x_i, \alpha_i \in \mathbb{R}\}$
- If $\{x_1, x_2, \dots, x_n\} \in \mathbb{R}^n$ is a set of linearly independent set of vectors, then $\text{span}(\{x_1, x_2, \dots, x_n\}) = \mathbb{R}^n$
- The range of a matrix $A \in \mathbb{R}^{m \times n}$, denoted as $\mathcal{R}(A)$, is the span of the columns of A
- The nullspace of a matrix $A \in \mathbb{R}^{m \times n}$, denoted $\mathcal{N}(A)$, is the set of all vectors that equal 0 when multiplied by A
 - $\mathcal{N}(A) = \{x \in \mathbb{R}^n : Ax = 0\}$

Row and Column Space

- The row space and column space are the linear subspaces generated by row and column vectors of a matrix
- Linear subspace, is a vector space that is a subset of some other higher dimension vector space
- For a matrix $A \in \mathbb{R}^{m \times n}$
 - $\text{Col space}(A) = \text{span}(\text{columns of } A)$
 - $\text{Rank}(A) = \dim(\text{row space}(A)) = \dim(\text{col space}(A))$

Matrix Rank: Examples

What are the ranks for the following matrices?

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

Rank 2 (circled 2x3 submatrix)

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

Rank 2 (circled 2x2 submatrix)

$$P_1 = \begin{bmatrix} * & * & * \\ 1 & 2 & 3 \\ 0 & 0 & 0 \end{bmatrix}$$

Rank 1 (circled 1x3 submatrix)

$$P_2 = \begin{bmatrix} * & * & * \\ 1 & 2 & 3 \\ 0 & 0 & 0 \end{bmatrix}$$

Rank 1 (circled 1x3 submatrix)

Matrix Inverse

- The inverse of a square matrix $A \in \mathbb{R}^{n \times n}$ is denoted A^{-1} and is the unique matrix such that $A^{-1}A = I = AA^{-1}$
- For some square matrices A^{-1} may not exist, and we say that A is **singular or non-invertible**. In order for A to have an inverse, A must be **full rank**.
- For non-square matrices the inverse, denoted by A^+ , is given by $A^+ = (A^T A)^{-1} A^T$ called the **pseudo inverse**

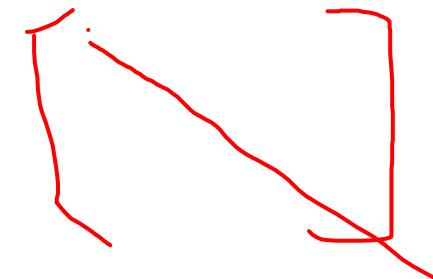
Outline

- Linear Algebra Basics
- Norms
- Multiplications
- Matrix Inversion
- Trace and Determinant 
- Eigen Values and Eigen Vectors
- Singular Value Decomposition

Matrix Trace

- The trace of a matrix $A \in \mathbb{R}^{n \times n}$, denoted as $\text{tr}(A)$, is the sum of the diagonal elements in the matrix

$$\text{tr}(A) = \sum_{i=1}^n A_{ii}$$



- The trace has the following properties
 - For $A \in \mathbb{R}^{n \times n}$, $\text{tr}(A) = \text{tr}A^\top$
 - For $A, B \in \mathbb{R}^{n \times n}$, $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$
 - For $A \in \mathbb{R}^{n \times n}$, $t \in \mathbb{R}$, $\text{tr}(tA) = t \cdot \text{tr}(A)$
 - For A, B, C such that ABC is a square matrix $\text{tr}(ABC) = \text{tr}(BCA) = \text{tr}(CAB)$
- The trace of a matrix helps us easily compute norms and eigenvalues of matrices as we will see later

Matrix Determinant

Definition (Determinant)

The determinant of a square matrix A , denoted by $|A|$, is defined as

$$\det(A) = \sum_{j=1}^n (-1)^{i+j} a_{ij} M_{ij}$$

where M_{ij} is determinant of matrix A without the row i and column j .

For a 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$

$$|A| = ad - bc$$

Properties of Matrix Determinant

Basic Properties

- $|A| = |A^T|$
- $|AB| = |A||B|$
- $|A| = 0$ if and only if A is not invertible
- If A is invertible, then $|A^{-1}| = \frac{1}{|A|}$.

$$|A| = \begin{vmatrix} A & A^{-1} \\ A & A \end{vmatrix} = \begin{vmatrix} A & A^{-1} \\ A & A \end{vmatrix} = \frac{1}{|A|}$$

Outline

- Linear Algebra Basics
- Norms
- Multiplications
- Matrix Inversion
- Trace and Determinant
- Eigen Values and Eigen Vectors ←
- Singular Value Decomposition

Eigenvalues and Eigenvectors

- Given a square matrix $A \in \mathbb{R}^{n \times n}$ we say that $\lambda \in \mathbb{C}$ is an eigenvalue of A and $x \in \mathbb{C}^n$ is an eigenvector if

$$Ax = \lambda x, \quad x \neq 0$$

- Intuitively this means that upon multiplying the matrix A with a vector x , we get the same vector, but scaled by a parameter λ
- Geometrically, we are transforming the matrix A from its original orthonormal basis/co-ordinates to a new set of orthonormal basis x with magnitude as λ

Computing Eigenvalues and Eigenvectors

- We can rewrite the original equation in the following manner

$$\begin{aligned} Ax &= \lambda x, \quad x \neq 0 \\ \Rightarrow (\lambda I - A)x &= 0, \quad x \neq 0 \end{aligned}$$

- This is only possible if $(\lambda I - A)$ is singular, that is $|(\lambda I - A)| = 0$.
- Thus, eigenvalues and eigenvectors can be computed.
 - Compute the determinant of $A - \lambda I$.
 - This results in a polynomial of degree n .
 - Find the roots of the polynomial by equating it to zero.
 - The n roots are the n eigenvalues of A . They make $A - \lambda I$ singular.
 - For each eigenvalue λ , solve $(A - \lambda I)x$ to find an eigenvector x

Eigenvalue Example

Eigenvalues

$$A = \begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix} \quad \lambda_1 = -5 \quad \lambda_2 = 2$$

Determine eigenvectors: $Ax = \lambda x$

$$x_1 + 2x_2 = \lambda x_1$$

$$3x_1 - 4x_2 = \lambda x_2$$

$$(1 - \lambda)x_1 + 2x_2 = 0$$

$$3x_1 - (4 + \lambda)x_2 = 0$$

Eigenvector for $\lambda_1 = -5$

$$6x_1 + 2x_2 = 0$$

$$3x_1 + x_2 = 0$$

$$\Rightarrow \mathbf{x}_1 = \begin{bmatrix} -0.3162 \\ 0.9487 \end{bmatrix} \text{ or } \mathbf{x}_1 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$$

Eigenvector for $\lambda_1 = 2$

$$-x_1 + 2x_2 = 0$$

$$3x_1 - 6x_2 = 0$$

$$\Rightarrow \mathbf{x}_2 = \begin{bmatrix} 0.8944 \\ 0.4472 \end{bmatrix} \text{ or } \mathbf{x}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Slide credit: Shubham Kumbhar

Matrix Eigen Decomposition

- All the eigenvectors can be written together as $AX = X\Lambda$ where the diagonals of X are the eigenvectors of A , and Λ is a diagonal matrix whose elements are eigenvalues of A
- If the eigenvectors of A are invertible, then $A = X\Lambda X^{-1}$
- There are several properties of eigenvalues and eigenvectors
 - $\text{Tr}(A) = \sum_{i=1}^n \lambda_i$
 - $|A| = \prod_{i=1}^n \lambda_i$
 - Rank of A is the number of non-zero eigenvalues of A
 - If A is non-singular then $1/\lambda_i$ are the eigenvalues of A^{-1}
 - The eigenvalues of a diagonal matrix are the diagonal elements of the matrix itself!

Properties of Eigendecomposition

- For a symmetric matrix A , it can be shown that eigenvalues are real and the eigenvectors are orthonormal. Thus it can be represented as $U\Lambda U^T$
- Considering quadratic form of A ,
 - $x^T A x = x^T U \Lambda U^T x = y^T \Lambda y = \sum_{i=1}^n \lambda_i y_i^2$ (where $y = U^T x$)
- Since y_i^2 is always positive the sign of the expression always depends on λ_i . If $\lambda_i > 0$ then the matrix A is positive definite, if $\lambda_i \geq 0$ then the matrix A is positive semidefinite
- For a multivariate Gaussian, the variances of x and y do not fully describe the distribution. The eigenvectors of this covariance matrix capture the directions of highest variance and eigenvalues the variance

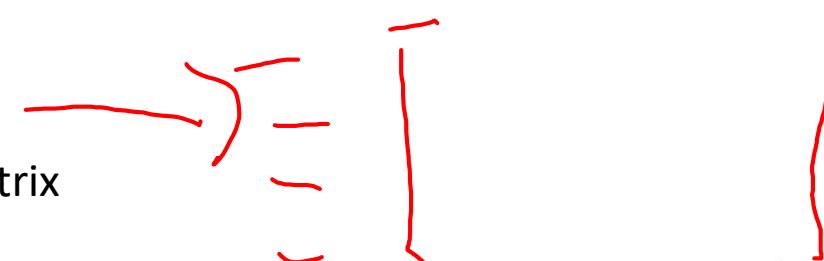
Outline

- Linear Algebra Basics
- Norms
- Multiplications
- Matrix Inversion
- Trace and Determinant
- Eigen Values and Eigen Vectors
- Singular Value Decomposition ←

Singular Value Decomposition

$X_{n \times d}$

n: instances
d: dimensions
 X is a centered matrix



$\underline{X = U\Sigma V^T}$

$\Sigma_{n \times d} \rightarrow \text{diagonal matrix}$

$U_{n \times n} \rightarrow \text{unitary matrix} \rightarrow U \times U^T = I$

A 2x3 matrix with red annotations. The first row is labeled {1, 2} and the second row is labeled {5, 6}. The first column is labeled 3, the second column is labeled 4, and the third column is labeled 7. Red curly braces group the rows and columns respectively.

$$\begin{bmatrix} 1 & 2 & 3 \\ 5 & 6 & 7 \end{bmatrix}$$

A 3x3 diagonal matrix with red annotations. The diagonal elements are labeled 6, 2, and 0. Red curly braces group the rows and columns respectively.

$$\begin{bmatrix} 6 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

A detailed diagram of the Singular Value Decomposition formula $X = U\Sigma V^T$. The formula is shown as $X = \begin{bmatrix} u_{1 \times 1} & \cdots & u_{1 \times n} \end{bmatrix} \times \begin{bmatrix} \Sigma_{1 \times 1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma_{p \times d} \end{bmatrix} \times \begin{bmatrix} v_{1 \times 1} & \cdots & v_{1 \times d} \\ \vdots & \ddots & \vdots \\ v_{d \times 1} & \cdots & v_{d \times d} \end{bmatrix}$. Red annotations include curly braces grouping the columns of U and the rows of V^T , and a large oval enclosing the entire product Σ . Red numbers 1, 2, and 3 are written near the U matrix, and red numbers 4, 5, and 6 are written near the V^T matrix.

$$X = \begin{bmatrix} u_{1 \times 1} & \cdots & u_{1 \times n} \\ \vdots & \ddots & \vdots \\ u_{n \times 1} & \cdots & u_{n \times n} \end{bmatrix} \times \begin{bmatrix} \Sigma_{1 \times 1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma_{p \times d} \end{bmatrix} \times \begin{bmatrix} v_{1 \times 1} & \cdots & v_{1 \times d} \\ \vdots & \ddots & \vdots \\ v_{d \times 1} & \cdots & v_{d \times d} \end{bmatrix}$$

$$\begin{aligned}
 \text{Var}(X) &= E((X - E(X))^2) \\
 &= E[(X - u)^2] = E(X^2) - E(X)^2 \\
 &= E(X^2) - 2E(X)u + u^2 = E(X^2) - u^2 \\
 \text{Cov}(X, Y) &= E[(X - E(X))(Y - E(Y))] \\
 &= E[XY - E(X)Y - XE(Y) + \\
 &\quad E(X)E(Y)] \\
 X &\sim \text{Unif}[a_0, a_n] \\
 \text{Cov}(X) &
 \end{aligned}$$

Covariance matrix:

$$X^T = \begin{bmatrix} a_{00} & \dots & a_{0d} \\ a_{10} & \ddots & \vdots \\ \vdots & \ddots & a_{dd} \end{bmatrix} \quad A_{d \times d} = \frac{X^T X}{n}$$

$$\cancel{X^T X} = \begin{bmatrix} \sum (x_{00}^2 + x_{10}^2) & \dots \\ \dots & \ddots \end{bmatrix}$$

$X = U \Sigma V^T$

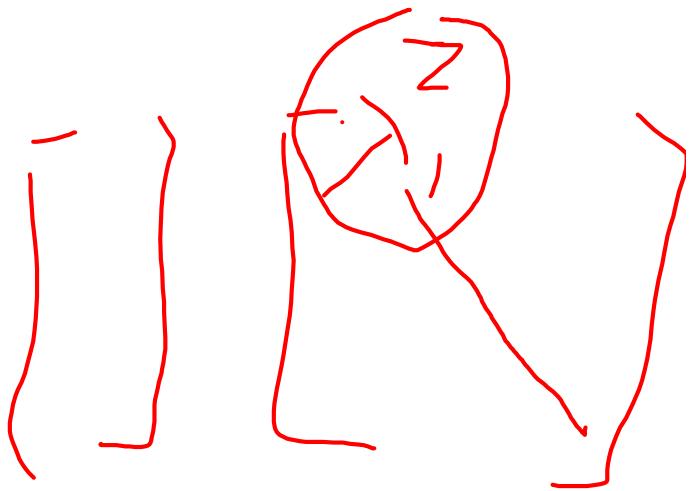
$$C = \frac{X^T X}{n}$$

$C = \frac{V \Sigma^T U^T U \Sigma V^T}{n} = \frac{V \Sigma^2 V^T}{n}$

$\boxed{\Sigma}$

$$A = \frac{V\Sigma^2V^T}{n} = V\frac{\Sigma^2}{n}V^T$$

$$AV = V\frac{\Sigma^2}{n}V^TV = V\frac{\Sigma^2}{n}$$



According to Eigen-decomposition definition $\rightarrow AV = \Lambda V$

$$\lambda_i = \frac{\sigma_i^2}{n} \rightarrow \text{The eigenvalues of covariance matrix}$$

V is the right singular vectors (Principal directions)

Geometric Meaning of SVD

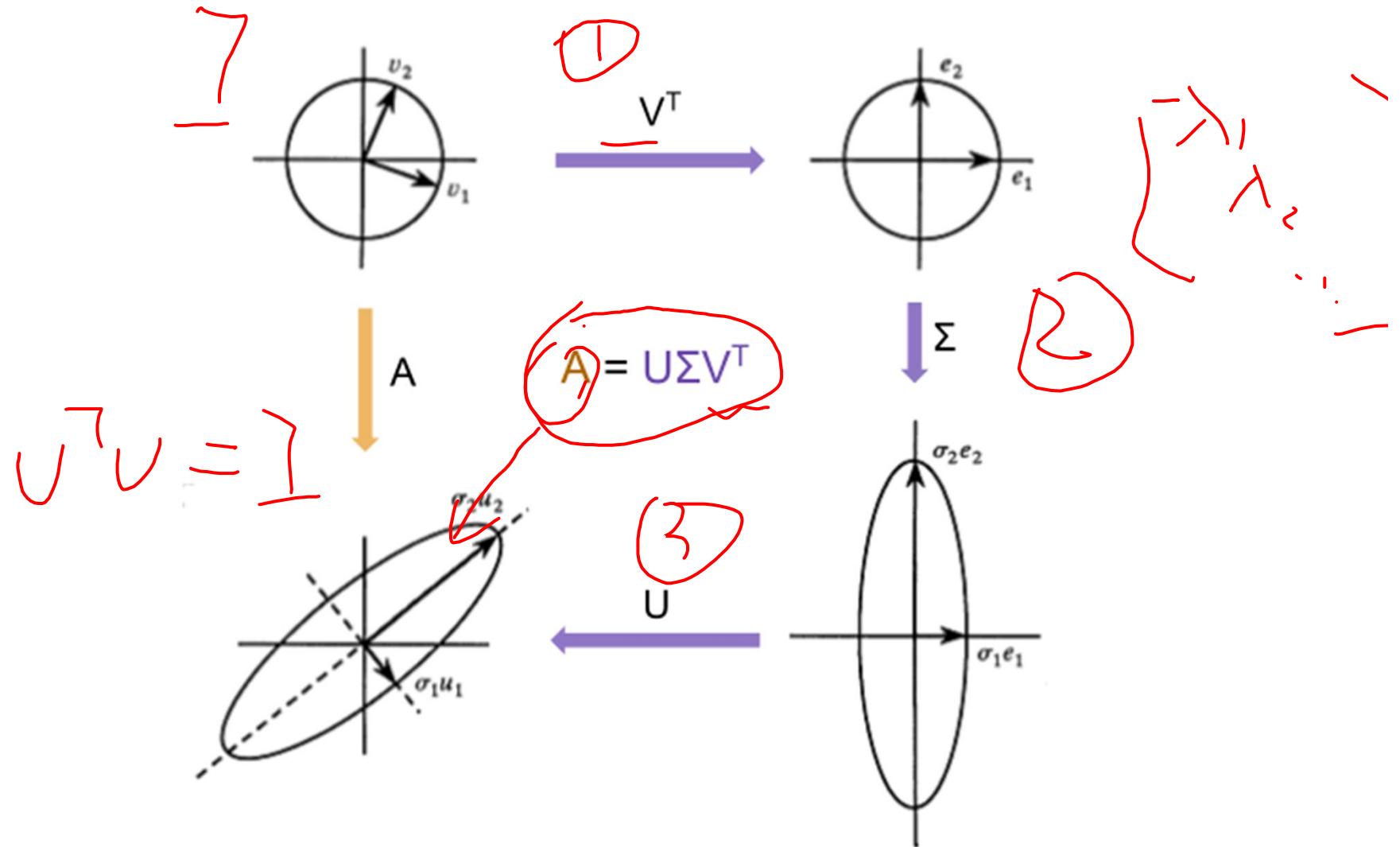


Image Credit: Kevin Binz

SVD Example

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

\mathbf{M} \mathbf{U} Σ \mathbf{V}^T

From [Strang]

Matrix Calculus

- For a vector $x, b \in \mathbb{R}^n$, let $f(x) = b^\top x$, then $\nabla_x b^\top x$ is equal to b
 - $\frac{\partial f(x)}{\partial x_k} = \frac{\partial}{\partial x_k} \sum_{i=1}^n b_i x_i = b_k$
- Now for a quadratic function, $f(x) = x^\top A x$, with $A \in \mathbb{S}^n$,
 $\frac{\partial f(x)}{\partial x_k} = 2Ax$
 - $\frac{\partial f(x)}{\partial x_k} = \frac{\partial}{\partial x_k} \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j$
 - $= \sum_{i \neq k} A_{ik} x_i + \sum_{j \neq k} A_{kj} x_j + 2A_{kk} x_k$
 - $= 2 \sum_{i=1}^n A_{ki} x_i$
- Let $f(X) = X^{-1}$, then $\partial(X^{-1}) = -X^{-1}(\partial X)X^{-1}$

Summary

- Linear Algebra Basics
- Norms
- Multiplications
- Matrix Inversion
- Trace and Determinant
- Eigen Values and Eigen Vectors
- Singular Value Decomposition