


# Linear Algebra Basics

Mahdi Roozbahani  
Georgia Tech

# Some logistics

- Creating your project's team.
- Office hours are started.

# 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  $Ax = b$

$$A = \begin{bmatrix} 4 & -5 \\ -2 & 3 \end{bmatrix} \quad b = \begin{bmatrix} -13 \\ 9 \end{bmatrix}$$

- $A \in \mathbb{R}^{n \times d}$  denotes a matrix with  $n$  rows and  $d$  columns, where elements belong to real numbers.
- $x \in \mathbb{R}^d$  denotes a vector with  $d$  real entries. In this case,  $\mathbb{R}^d$  is a column vector ( $d$  rows 1 column), but  $\mathbb{R}^d$  can also be thought of as a matrix with 1 row and  $d$  columns in other situations.

# Linear Algebra Basics

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

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


$$H = H^T$$

$$\underline{(A + A^T)} = \underline{(A + A^T)}^T = A^T + (A^T)^T = \underline{A^T + A} \quad \checkmark$$

$$G = -G^T$$

$$\underline{(A - A^T)} = \underline{(A - A^T)}^T = - \underline{(A^T - A^{TT})} = A - A^T$$

# 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

$$x \in \mathbb{R}^d \quad \|x\| \in \mathbb{R}$$

- More formally, a norm is any function  $f: \mathbb{R}^d \rightarrow \mathbb{R}$  that satisfies

- For all  $x \in \mathbb{R}^d$ ,  $f(x) \geq 0$  (non-negativity)
- $f(x) = 0$  if and only if  $x = 0$  (definiteness)
- For  $x \in \mathbb{R}^d$ ,  $t \in \mathbb{R}$ ,  $f(tx) = |t|f(x)$  (homogeneity)
- For all  $x, y \in \mathbb{R}^d$ ,  $f(x + y) \leq f(x) + f(y)$  (triangle inequality)

$$x^2 = \text{graph of } y = x^2$$

- Common norms used in machine learning are

- $\ell_2$  norm

$\|x\|$

$$\|x\|_2 = \sqrt{\sum_{i=1}^d x_i^2}$$

$$X = [1, 2, 3]$$

$$\|x\| = \|x\|_2 = \sqrt{1^2 + 2^2 + 3^2}$$

$$\|x\|^2 = \|x\|_2^2 = 1^2 + 2^2 + 3^2$$



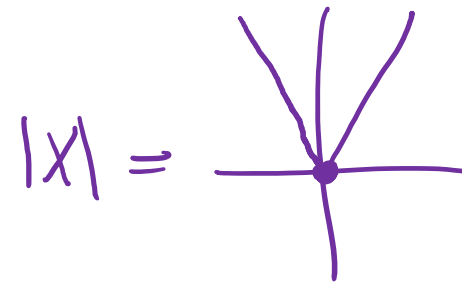
# Norms

$$x = [-1, -3, 2]$$

- $\ell_1$  norm

- $\|x\|_1 = \sum_{i=1}^d |x_i|$

$$\|x\|_1 = |-1| + |-3| + |2|$$



- $\ell_\infty$  norm

- $\|x\|_\infty = \max_i |x_i|$

$$\|x\|_\infty = \max(|-1|, |-3|, |2|) = 3$$

- All norms presented so far are examples of the family of  $\ell_p$  norms, which are parameterized by a real number  $p \geq 1$

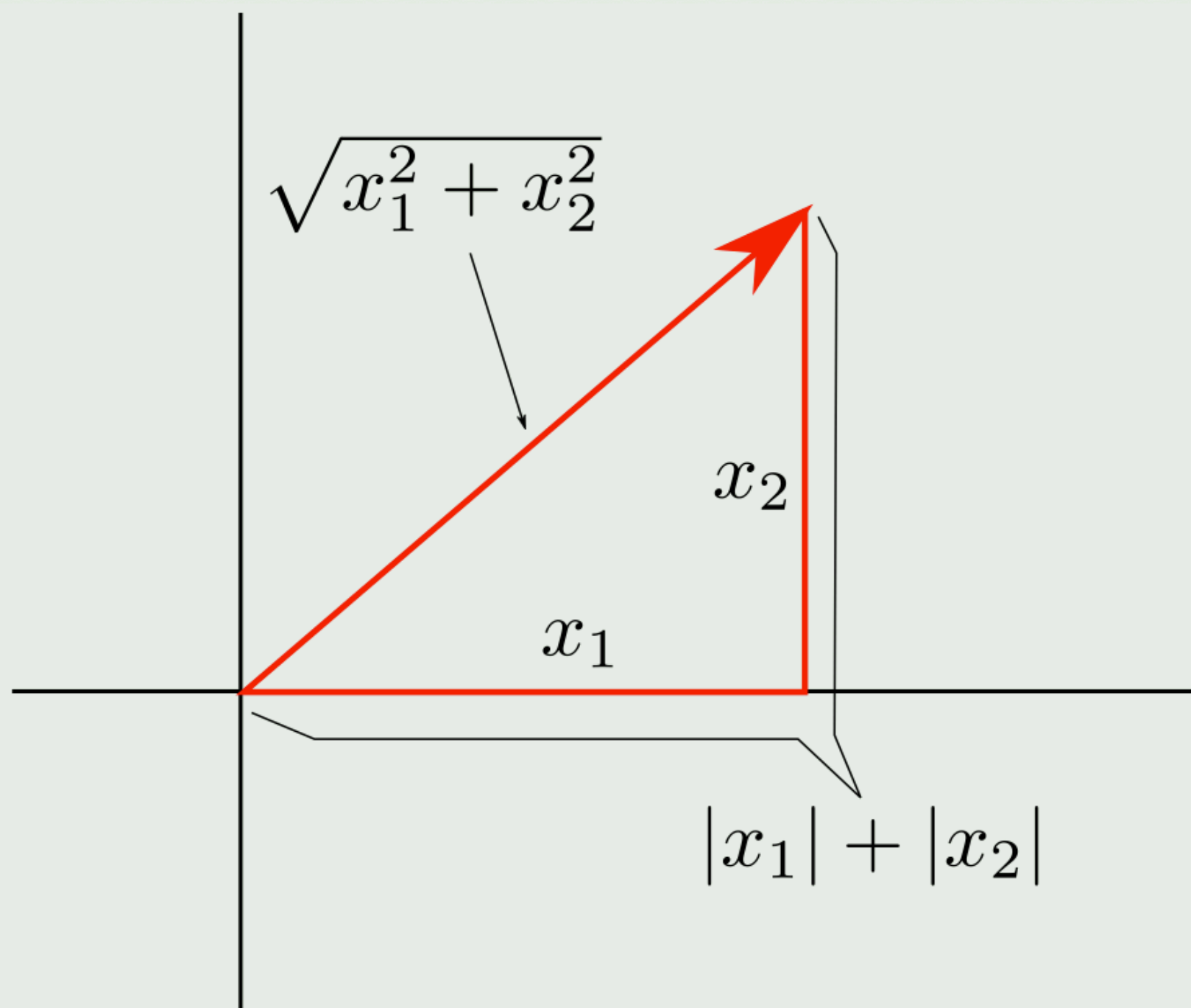
- $\|x\|_p = \left( \sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}$

- Norms can be defined for matrices, such as the Frobenius norm.

- $\|A\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^d A_{ij}^2} = \sqrt{\text{tr}(A^\top A)}$

# Vector Norm Examples

## Example $\ell_1$ -norm and $\ell_2$ -norm



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

# Special Matrices

$$X \cdot X^T = X^2 = \begin{bmatrix} 1^2 & 0 & 0 \\ 0 & 2^2 & 0 \\ 0 & 0 & 3^2 \end{bmatrix}$$

$$X^{-1} = \begin{bmatrix} \frac{1}{1} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}$$

- The identity matrix, denoted by  $I \in \mathbb{R}^{d \times d}$  is a square matrix with ones on the diagonal and zeros everywhere else
- A diagonal matrix is a matrix where all non-diagonal 'ELEMENTS' are 0. This is typically denoted as  $D = \text{diag}(d_1, d_2, \dots, d_d)$
- Two vectors  $x, y \in \mathbb{R}^d$  are orthogonal if  $x \cdot y = 0$ . A square matrix  $U \in \mathbb{R}^{d \times d}$  is **Orthonormal** if all its columns are orthogonal to each other and are normalized

Unitary = orthonormal

$$X = \begin{bmatrix} a & b & c \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- It follows from orthogonality and normality that

- $U^T U = I = U U^T$


- $\|Ux\|_2 = \|x\|_2$

Is the inverse of a unitary matrix equal to its transpose?

$$U U^{-1} = I$$

$$U U^T = I$$

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

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

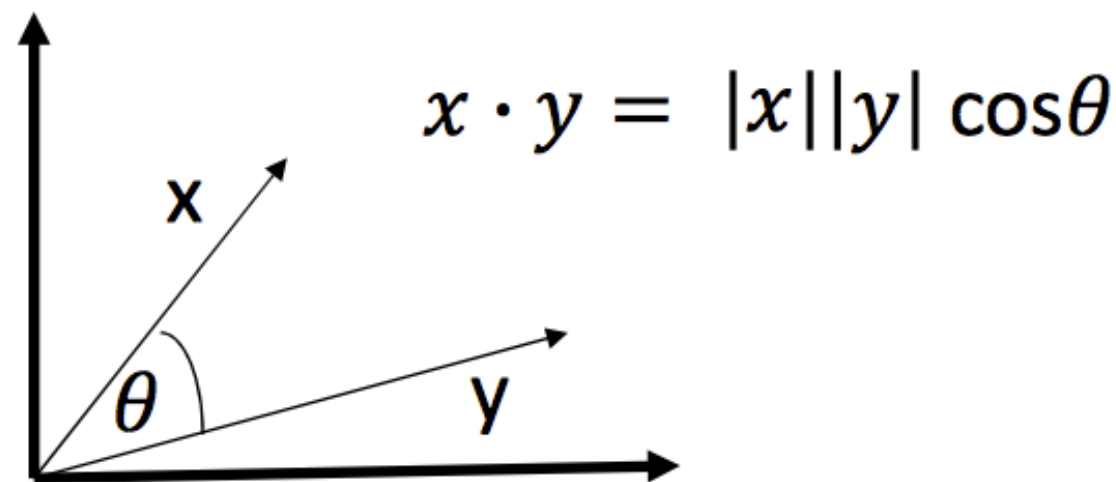
- Given two vectors  $x \in \mathbb{R}^d$ ,  $y \in \mathbb{R}^n$ , the term  $x^T y$  is called the **outer product** of the vectors:  $x \otimes y$

*Is Dot Product a linear operation?*

# Multiplications

$$x \otimes y = x^T y = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} [y_1 \quad y_2 \quad y_3] = \begin{bmatrix} x_1 y_1 & x_1 y_2 & x_1 y_3 \\ x_2 y_1 & x_2 y_2 & x_2 y_3 \\ x_3 y_1 & x_3 y_2 & x_3 y_3 \end{bmatrix}$$

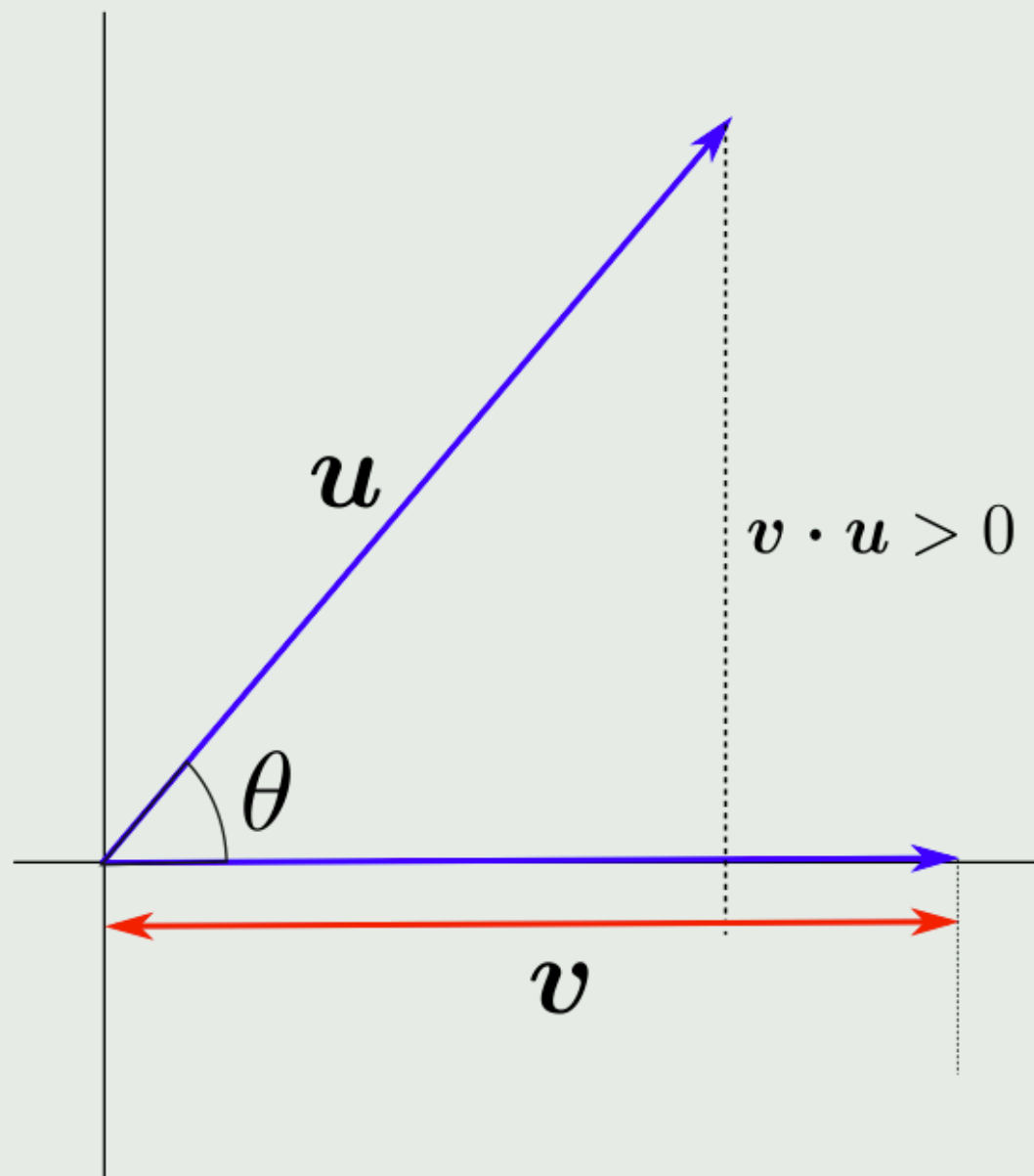
- The dot product also has a geometrical interpretation, for vectors in  $x, y \in \mathbb{R}^2$  with angle  $\theta$  between them



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

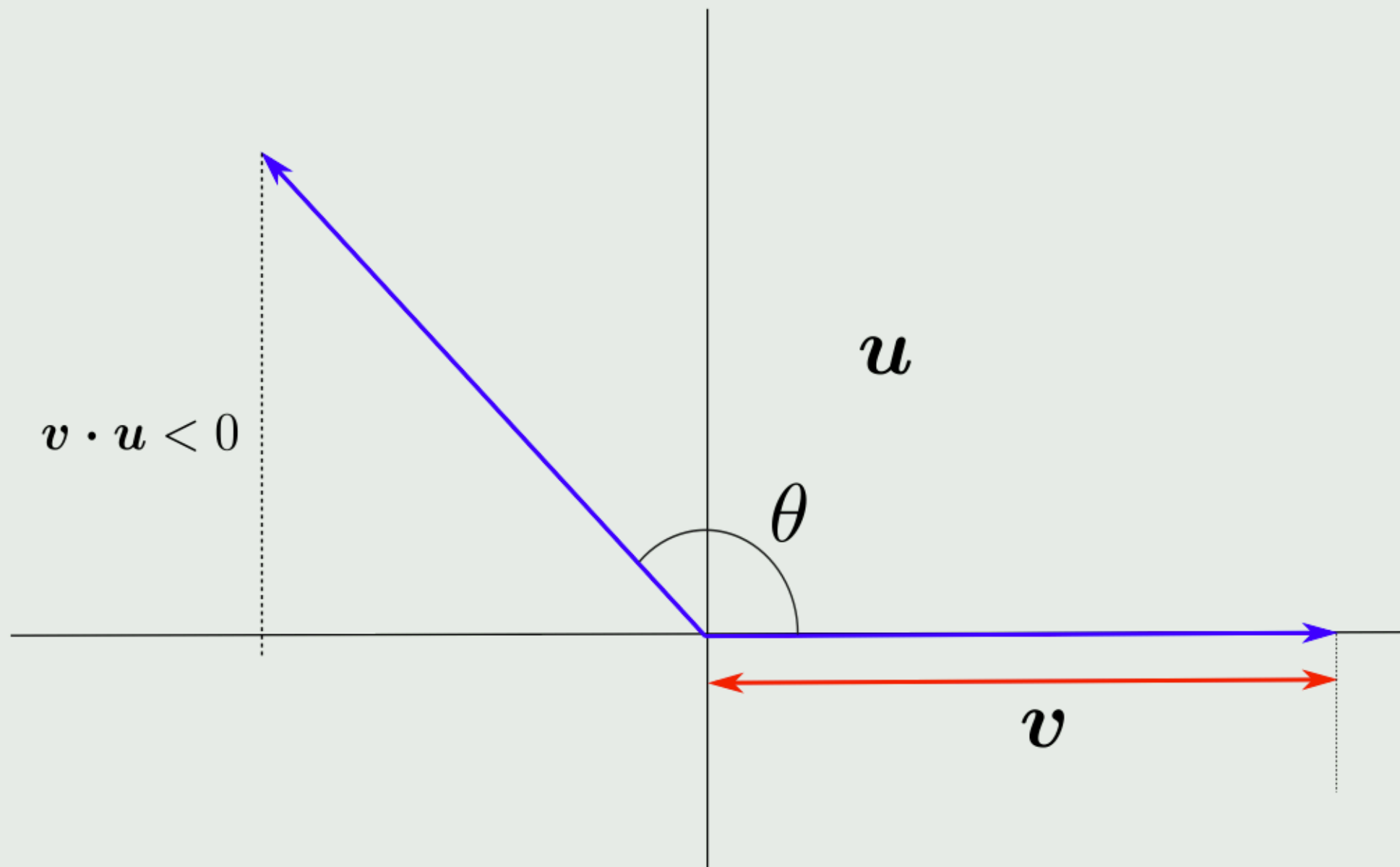
# Inner Product Properties

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



# Inner Product Properties

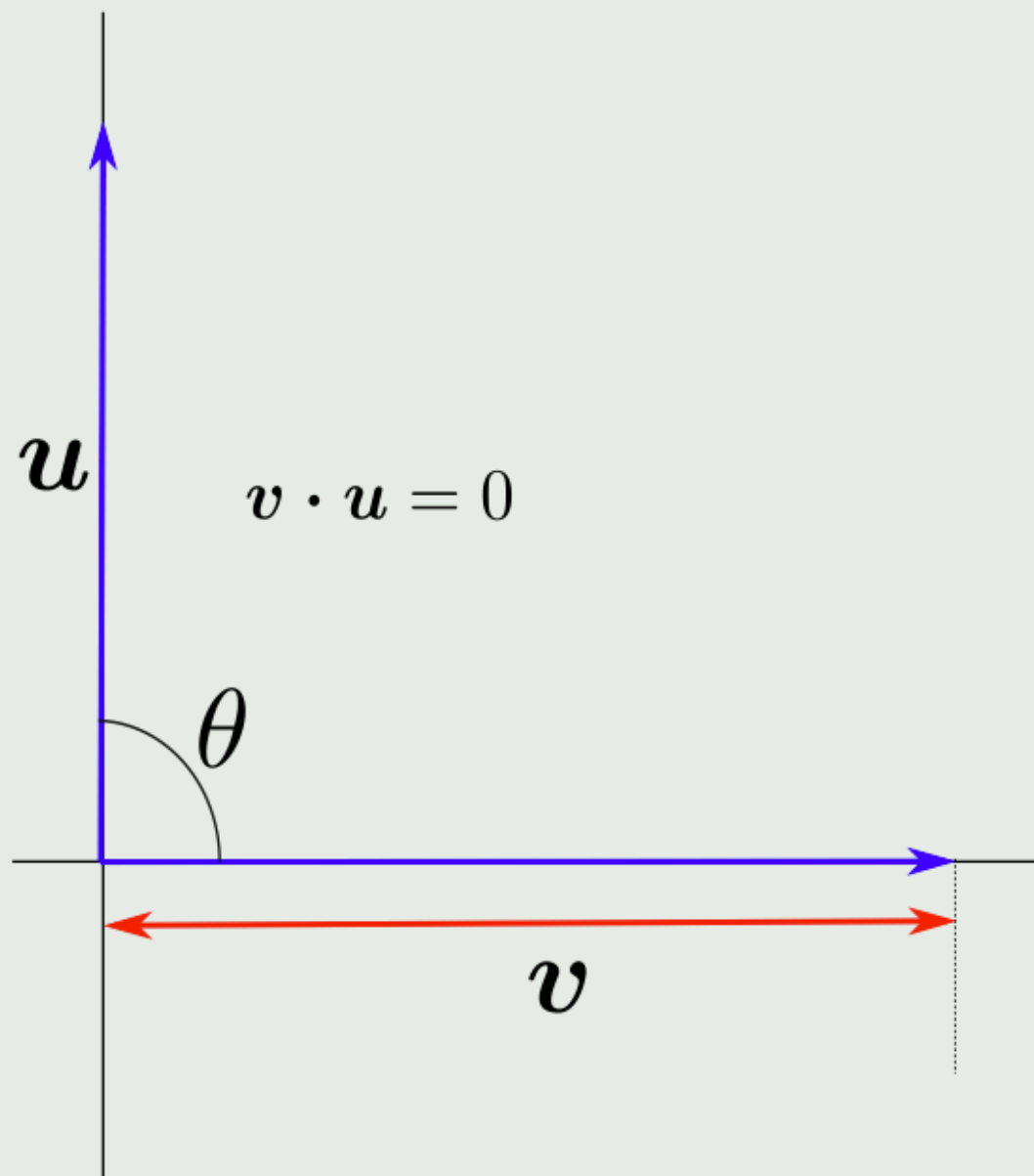
The inner product is a measure of correlation between two vectors, scaled by the norms of the vectors ?






# Inner Product Properties

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



If two variables are uncorrelated, they are orthogonal and if two variables are orthogonal, they are uncorrelated. (Can I really say that?)

# 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  $\{x_1, x_2, \dots, x_d\} \subset \mathbb{R}^d$  are said to be **(linearly) independent** if no vector can be represented as a linear combination of the remaining vectors. That is if  $\alpha a + \beta b = c$

$$x_d = \sum_{i=1}^{d-1} \alpha_i x_i$$

$\begin{bmatrix} a & b & c \end{bmatrix}_{3 \times 3}$

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

$$A_{30 \times 40} \quad B_{30 \times 10}$$

- The **column rank** of a matrix  $A \in \mathbb{R}^{n \times d}$  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.

It is a full rank if the rank is  $\min\{n, d\}$ . This is the maximum rank.

# Matrix Rank: Examples

What are the ranks for the following matrices? How about an identity matrix

$$\mathbf{A} = \begin{matrix} & \begin{matrix} a & b & c \end{matrix} \\ \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{bmatrix} \end{matrix} \quad \text{Rank} = 1$$

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 2 \\ 2 & 1 & 0 \\ 3 & 2 & 1 \end{bmatrix} \quad \text{Rank} = 3 = \text{full rank}$$

# Matrix Inverse

- The inverse of a square matrix  $A \in \mathbb{R}^{d \times d}$  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**


$$\underline{A_{n \times d}}$$

$$(A^T A)_{d \times d}$$

$$\text{Inv}(A^T A) = \frac{1}{A^T A}$$

$$P_{\text{inv}}(A_{n \times d}) = \frac{1}{A^T A} \times A^T$$

# 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  $A \in \mathbb{R}^{d \times d}$ , denoted as  $\mathbf{tr}(A)$ , is the sum of the diagonal elements in the matrix

$$\mathbf{tr}(A) = \sum_{i=1}^d A_{ii}$$

- The trace has the following properties
  - For  $A \in \mathbb{R}^{d \times d}$ ,  $\mathbf{tr}(A) = \mathbf{tr}A^\top$
  - For  $A, B \in \mathbb{R}^{d \times d}$ ,  $\mathbf{tr}(A + B) = \mathbf{tr}(A) + \mathbf{tr}(B)$
  - For  $A \in \mathbb{R}^{d \times d}$ ,  $t \in \mathbb{R}$ ,  $\mathbf{tr}(tA) = t \cdot \mathbf{tr}(A)$
  - For  $A, B, C$  such that  $ABC$  is a square matrix  $\mathbf{tr}(ABC) = \mathbf{tr}(BCA) = \mathbf{tr}(CAB)$
- 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

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

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

For a  $2 \times 2$  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  $|A^{-1}| = \frac{1}{|A|}$ .

# 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  $A \in \mathbb{R}^{d \times d}$  we say that  $\lambda \in \mathbb{C}$  is an eigenvalue of  $A$  and  $x \in \mathbb{C}^d$  is an eigenvector if

$A = \begin{bmatrix} & \\ & \end{bmatrix}_{2 \times 2}$

- Intuitively this means that upon multiplying the matrix  $A$  with a vector  $x$ , we get the same vector, but scaled by a parameter  $\lambda$
- 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  $\lambda$

# Computing Eigenvalues and Eigenvectors

- We can rewrite the original equation in the following manner

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

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

# Eigenvalue Example

Eigenvalues

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

$$\lambda_2 = 2$$

$$| \mathbf{A} - \underline{\underline{sI}} | = \left| \begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix} - \begin{bmatrix} s & 0 \\ 0 & s \end{bmatrix} \right| =$$

$$\left| \begin{bmatrix} 1-s & 2 \\ 3 & -4-s \end{bmatrix} \right| = 0$$

$$(1-s)(-4-s) - 6 = 0 \quad \begin{matrix} s_1 = -5 \\ s_2 = 2 \end{matrix}$$

Determine eigenvectors:  $\mathbf{Ax} = \lambda \mathbf{x}$

$$\begin{aligned} x_1 + 2x_2 &= \lambda x_1 \\ 3x_1 - 4x_2 &= \lambda x_2 \end{aligned} \Rightarrow \begin{cases} (1-\lambda)x_1 + 2x_2 = 0 \\ 3x_1 - (4+\lambda)x_2 = 0 \end{cases}$$

$$\begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = s \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Eigenvector for  $\lambda_1 = -5$

$$\begin{cases} 6x_1 + 2x_2 = 0 \\ 3x_1 + x_2 = 0 \end{cases} \Rightarrow \mathbf{x}_1 = \begin{bmatrix} -0.3162 \\ 0.9487 \end{bmatrix} \text{ or } \mathbf{x}_1 = \begin{bmatrix} 1 \\ -3 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix} \begin{bmatrix} 1 \\ -3 \end{bmatrix} = \begin{bmatrix} -5 \\ 6 \end{bmatrix} = \begin{bmatrix} -5 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ -3 \end{bmatrix}$$

Eigenvector for  $\lambda_1 = 2$

$$\begin{aligned} -x_1 + 2x_2 &= 0 \\ 3x_1 - 6x_2 &= 0 \end{aligned} \Rightarrow \mathbf{x}_2 = \begin{bmatrix} 0.8944 \\ 0.4472 \end{bmatrix} \text{ or } \mathbf{x}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

# Matrix Eigen Decomposition

- All the eigenvectors can be written together as  $AX = X\Lambda$  where the columns of  $X$  are the eigenvectors of  $A$ , and  $\Lambda$  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
  - $Tr(A) = \sum_{i=1}^d \lambda_i$
  - $|A| = \prod_{i=1}^d \lambda_i$
  - 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!

Can a matrix have the same eigenvalues?

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{Yes}$$

Are the eigenvectors of a matrix orthogonal against each other? NO; but if the matrix is symmetric, then Yes


If two vectors are linearly independent, does it mean they are orthogonal against each other? NO

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

$$\begin{matrix} a & b \\ \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \end{matrix}$$



# Outline

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



# Singular Value Decomposition

$\bar{X}_{n \times d}$       n: instances  
                     d: dimensions  
                     X is a centered matrix

*np.svd(X)*

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

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

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

$V_{d \times d} \rightarrow \text{unitary matrix} \rightarrow V \times V^T = I$

$$\begin{matrix}
 X = \begin{bmatrix} u_{1 \times 1} & \dots & \dots & \dots & u_{1 \times n} \\ \vdots & \ddots & \dots & \dots & \vdots \\ \vdots & \vdots & \ddots & \dots & \vdots \\ \vdots & \vdots & \dots & \ddots & \vdots \\ u_{1 \times 1} & \dots & \dots & \dots & u_{n \times n} \end{bmatrix} & \times & \begin{bmatrix} \Sigma_{1 \times 1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Sigma_{d \times d} \\ 0 & 0 & 0 \end{bmatrix} & \times & \begin{bmatrix} v_{1 \times 1} & \dots & \dots & \dots & v_{1 \times d} \\ \vdots & \ddots & \dots & \dots & \vdots \\ \vdots & \vdots & \ddots & \dots & \vdots \\ \vdots & \vdots & \dots & \ddots & \vdots \\ v_{d \times 1} & \dots & \dots & \dots & v_{d \times d} \end{bmatrix} \\
 U & & \Sigma & & V^T \\
 & & d < n & &
 \end{matrix}$$

Covariance matrix:

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

$$C_{d \times d} = \frac{\bar{X}^T \bar{X}}{n}$$

Eigendecomposition of  $\underline{\underline{C}}$

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

$$\left. \begin{array}{l} \bar{X} = U \Sigma V^T \\ C = \frac{\bar{X}^T \bar{X}}{n} \end{array} \right\} C = \frac{V \Sigma^T U^T U \Sigma V^T}{n} = \frac{V \Sigma^2 V^T}{n}$$

$$A = U \Sigma V^T$$

Handwritten annotations:  $\Sigma^2/n$  is circled, and arrows point from the circled  $\Sigma^2/n$  and the circled  $V^T$  in the equation above to the  $CV = V\Lambda$  equation below.

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

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

$$CV = V\Lambda$$

Remember:

$$\boxed{AX = X\Lambda}$$

Handwritten annotations: Vertical double lines connect  $\Lambda$  in the top equation to  $\Lambda$  in the boxed equation. The entire boxed equation is circled in purple.

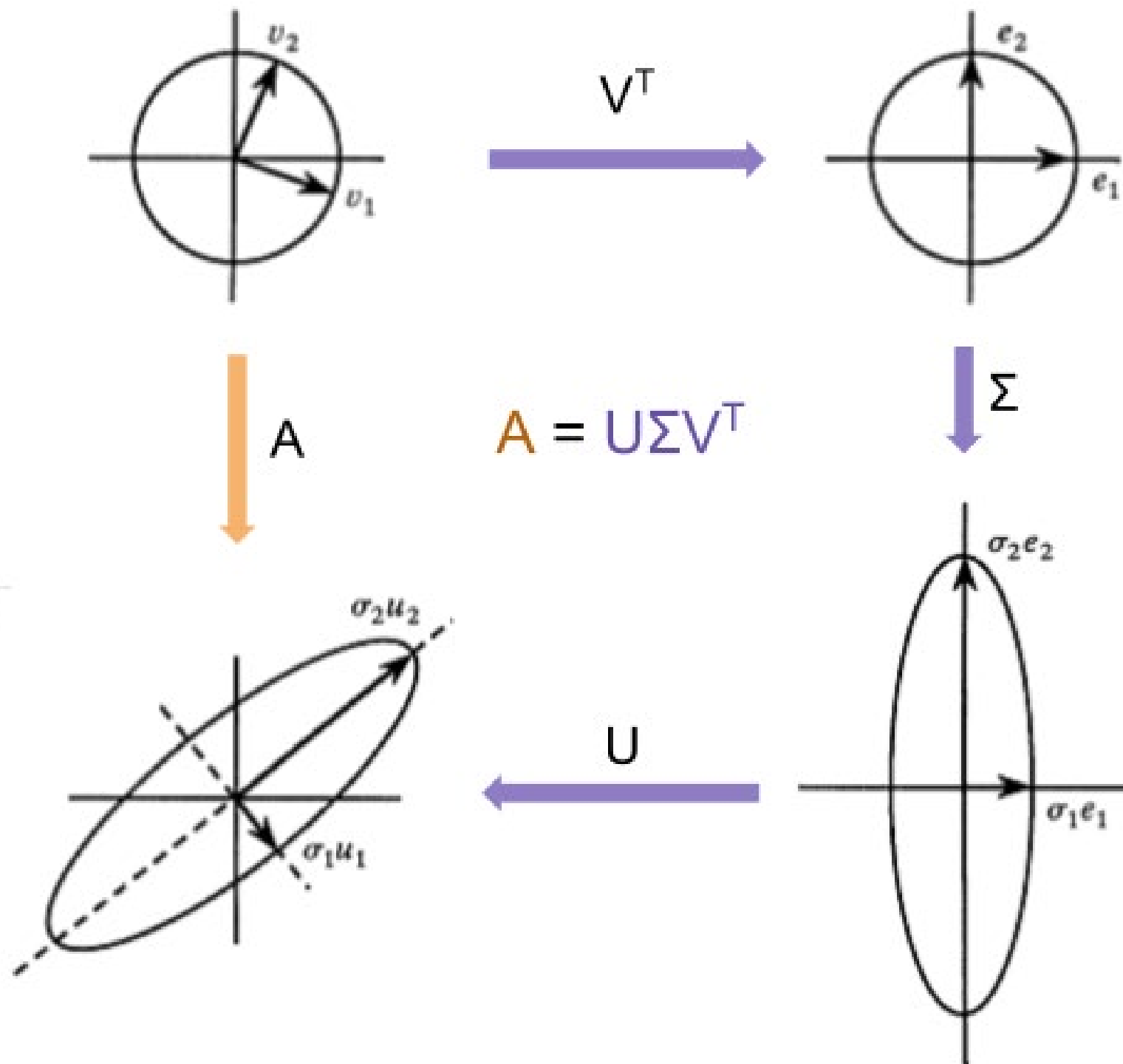
$\lambda_i = \frac{\Sigma_i^2}{n} \rightarrow$  The eigenvalues of covariance matrix

$\lambda_i$ : Eigenvalue of  $C$  or covariance matrix

$\Sigma_i$ : Singular value of  $X$  matrix

So, we can directly calculate eigenvalue of a covariance matrix by having the singular value of matrix  $X$  **directly**

# Geometric Meaning of SVD



# Summary

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