UNIVERSITY AT ALBANY
STATE UNIVERSITY OF NEW YORK

CSI 436/536 (Fall 2024)

Machine Learning

Lecture 2: Review of Linear Algebra

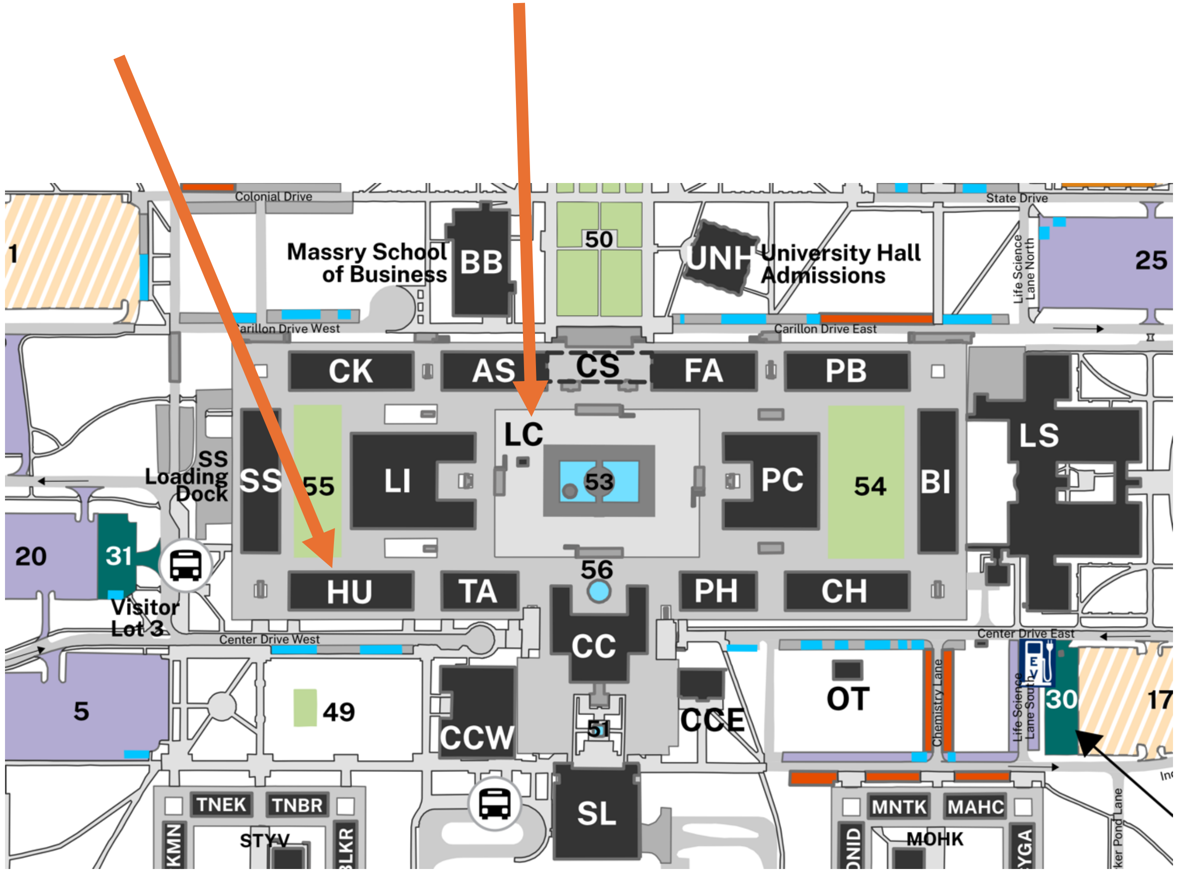
Chong Liu

Assistant Professor of Computer Science

Aug 29, 2024

Announcement

- Office hours:
  - Instructor: Tue 1:30-2:30pm @ HU 25
  - TA: Wed 1:30-2:30pm @ HU 25
  - Starting next week
- Enroll in Gradescope!
  - Entry code: **EV6862**
  - All homework via Gradescope
- No participation score today
  - Starting next week



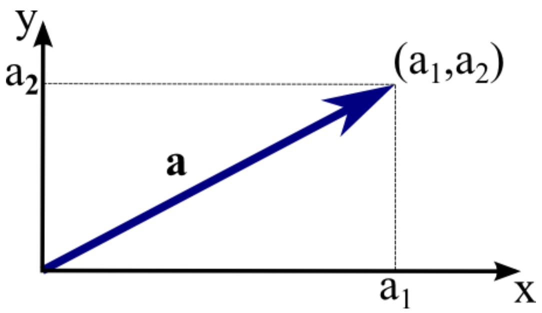
- Project list will be released next Tuesday

Today’s agenda

- Key objects:
  - Vector, matrix
- Operations:
  - Matrix-vector multiplication, matrix-matrix multiplication
- Properties vectors:
  - Norm (one vector), distance and angle (two vectors), linear (in)dependence, orthogonality (a “bag” of vectors)
- Properties of a matrix:
  - Rank, trace, determinant, symmetric, invertible
- Eigenvalues and eigenvectors

Vector and matrix

- Geometric meaning of a vector:
  - An arrow pointing from 0
  - A point in a coordinate system



$$\boldsymbol{a} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \in \mathbb{R}^3$$

- Matrix is a "bag" of vectors.
  - n-column vectors or m-row vectors.

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad a_{ij} \in \mathbb{R}\,.$$

Norms are “metrics”. A few useful properties:

Generally, a vector norm is a mapping $R^n \to R$, with the properties

- $||x|| \geq 0$, for all x
- $||x|| = 0$, if and only if $x = 0$
- $||\alpha x|| = |\alpha| ||x||$, $\alpha \in \mathbb{R}$
- $||x + y|| \leq ||x|| + ||y||$, for all x and y

l_p-norm is the most used vector norm

- Definition:

$$\|\mathbf{x}\|_p := \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$

- Different norms:
  - When $p = 1$, l_1-norm, Taxicab norm, Manhattan norm

$$\|\boldsymbol{x}\|_1 := \sum_{i=1}^{n} |x_i|$$

  - When $p = 2$, l_2-norm, Euclidean norm, quardratic norm, square norm
    - In literature, $\|x\|$ usually denotes Euclidean norm

$$\|\boldsymbol{x}\|_2 := \sqrt{x_1^2 + \cdots + x_n^2}$$

  - When $p \to \infty$, l_∞-norm

$$\|\mathbf{x}\|_\infty := \max_i |x_i|$$

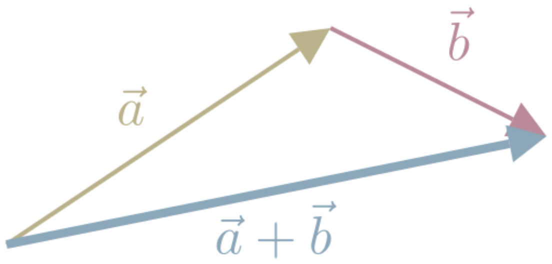
In-class exercise

- Find l_1-norm, l_2-norm, l_∞-norm of vector $x = [1,2,3,4,-5]$.

- Answer: $15, \sqrt{55}, 5$.

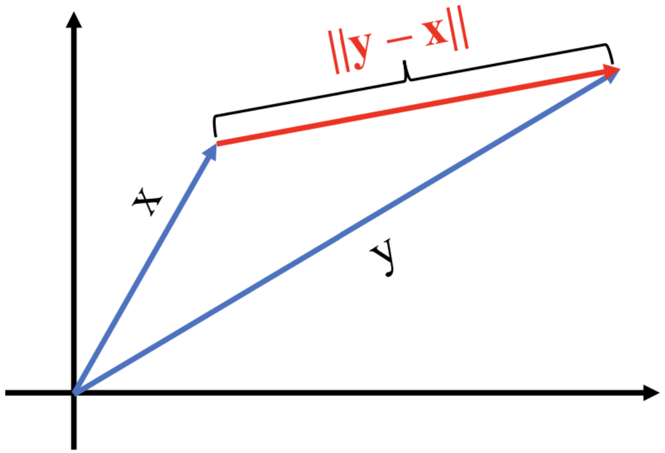
Properties of two vectors

- What can you do with them?
  - Add
    - $z = x + y$
    - $[5,6,-2] = [1,3,5] + [4,3,-7]$
  - Subtract
    - $g = x - y$
    - $[-3,0,12] = [1,3,5] - [4,3,-7]$
  - Weighted combination / linear combination
    - $h = x + 2y$
    - $[9,10,-9] = [1,3,5] + 2 * [4,3,-7]$

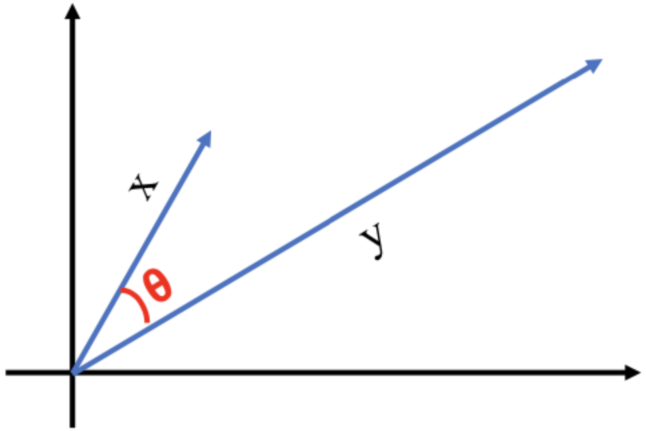


Relationship (similarity) of two vectors

- Direction



- Angle
  - Dot product / inner product
    - $\langle x, y \rangle = x^T y = \sum_{i=1}^{n} x_i y_i$
  - $\theta = cos^{-1}\left(\frac{\langle x,y \rangle}{\|x\|\|y\|}\right)$



Two vectors are **orthogonal** (perpendicular to each other) iff their dot-product is 0.

Three interpretations of matrix-vector Multiplication

- Interpretation 1: "Projecting x to m-directions"
  - Treat matrix A is as a "bag" of row-vectors

  - A is a m by n matrix
  - x is a n–dimensional vector
  - $Ax = \begin{bmatrix} 6 & 2 & 4 \\ -1 & 4 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 24 \\ -9 \end{bmatrix}$

  - Projecting x from 3 dimensions to 2 dimensions.

Three interpretations of Matrix-Vector Multiplication

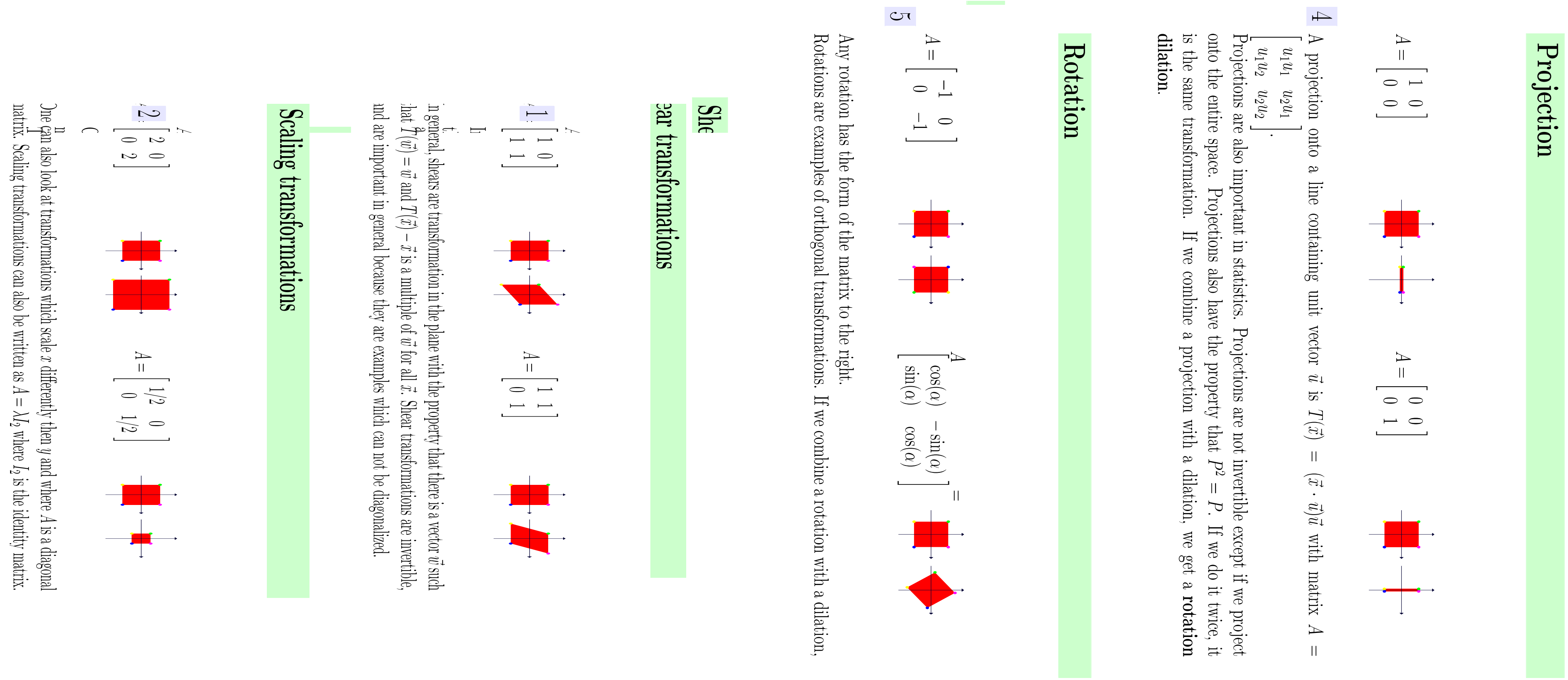
- Interpretation 2: "Weighted linear combination of column vectors"
  - Treat matrix A is as a "bag" of column-vectors

  - A is a m by n matrix
  - x is a n–dimensional vector
  - $Ax = \begin{bmatrix} 6 & 2 & 4 \\ -1 & 4 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 24 \\ -9 \end{bmatrix}$

  - The weight of column 1 is 4
  - The weight of column 2 is -1
  - …

Three interpretations of matrix-vector Multiplication

- Interpretation 3: “A linear transformation of input vector x”
  - Treat matrix A is as an “operator” or a “function that takes a vector input and output another vector” $A : \mathbb{R}^n \to \mathbb{R}^m$



Projection

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

4 A projection onto a line containing unit vector $\vec{u}$ is $T(\vec{x}) = (\vec{x} \cdot \vec{u})\vec{u}$ with matrix $A = \begin{bmatrix} u_1u_1 & u_2u_1 \\ u_1u_2 & u_2u_2 \end{bmatrix}$.
Projections are also important in statistics. Projections are not invertible except if we project onto the entire space. Projections also have the property that $P^2 = P$. If we do it twice, it is the same transformation. If we combine a projection with a dilation, we get a **rotation dilation**.

Rotation

$A = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$ $A = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix}$

5 Any rotation has the form of the matrix to the right.
Rotations are examples of orthogonal transformations. If we combine a rotation with a dilation,

She

ear transformations

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

n general, shears are transformation in the plane with the property that there is a vector $\vec{w}$ such hat $T(\vec{w}) = \vec{w}$ and $T(\vec{x}) - \vec{x}$ is a multiple of $\vec{w}$ for all $\vec{x}$. Shear transformations are invertible, nd are important in general because they are examples which can not be diagonalized.

Scaling transformations

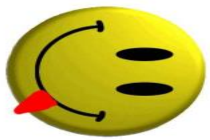
2 $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ $A = \begin{bmatrix} 1/2 & 0 \\ 0 & 1/2 \end{bmatrix}$

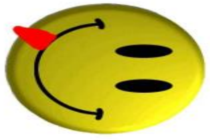
One can also look at transformations which scale x differently then y and where A is a diagonal matrix. Scaling transformations can also be written as $A = \lambda I_2$ where I_2 is the identity matrix.

In-class exercise: map each pixel to a new location

b) The **smiley face** visible to the right is transformed with various linear transformations represented by matrices $A-F$. Find out which matrix does which transformation:

$$A=\begin{bmatrix}1 & -1\\ 1 & 1\end{bmatrix},\quad B=\begin{bmatrix}1 & 2\\ 0 & 1\end{bmatrix},\quad C=\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix},$$
$$D=\begin{bmatrix}1 & -1\\ 0 & -1\end{bmatrix},\quad E=\begin{bmatrix}-1 & 0\\ 0 & 1\end{bmatrix},\quad F=\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}/2$$



A-F	image	A-F	image	A-F	image
					
					

Matrix-Matrix multiplication

- Let $A \in \mathbb{R}^{m \times p}$ and $B \in \mathbb{R}^{p \times n}$. Then,
  $C = AB = (c_{ij}) \in \mathbb{R}^{m \times n}$ is defined as follows:

$$c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj}, \text{ for all } i = 1, \cdots, m, j = 1, \cdots, n.$$

- Key things to remember
  - Dimension check!

- Properties of a scalar-scalar multiplications (which ones are still valid for matrix-matrix multiplication?)
  - Commutative law: AB=BA?
  - Associative law: (AB)C=A(BC)?
  - Distributive law: A(B+C)=AB+BC?

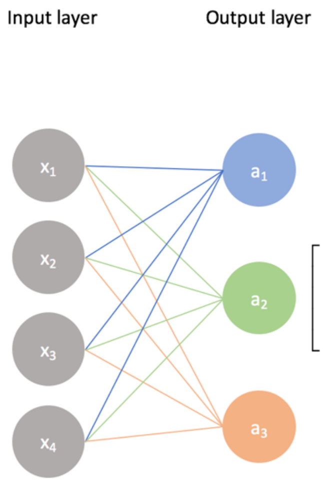
Examples of matrix-matrix multiplication

- Inner product and outer product of two vectors

$$\mathbf{u} \otimes \mathbf{v} = \mathbf{u}\mathbf{v}^{\mathsf{T}} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} = \begin{bmatrix} u_1v_1 & u_1v_2 & u_1v_3 \\ u_2v_1 & u_2v_2 & u_2v_3 \\ u_3v_1 & u_3v_2 & u_3v_3 \\ u_4v_1 & u_4v_2 & u_4v_3 \end{bmatrix}$$

- Page rank (mathematics behind Google Search)
    - https://pi.math.cornell.edu/~mec/Winter2009/RalucaRemus/Lecture3/lecture3.html

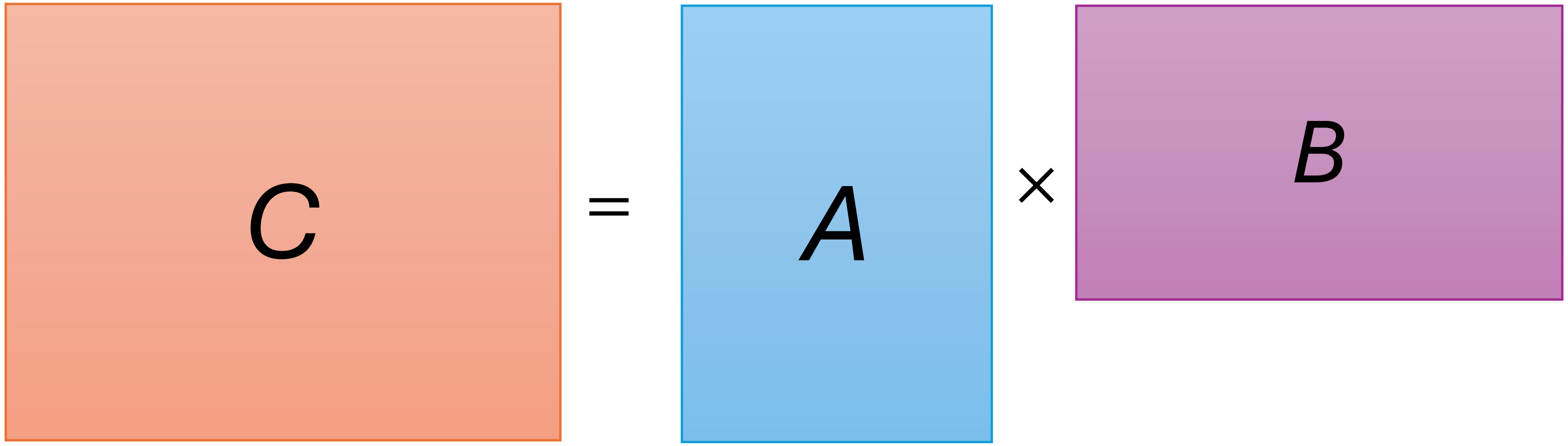
- Neural networks



A simple neural network

$$\begin{bmatrix} w_1 & w_2 & w_3 & w_4 \\ w_1 & w_2 & w_3 & w_4 \\ w_1 & w_2 & w_3 & w_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} b \\ b \\ b \end{bmatrix} = \begin{bmatrix} w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \\ w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \\ w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \end{bmatrix} \xrightarrow[\text{activation}]{} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

Computational Complexity of matrix Multiplication?



- How many dot product needed? (A is m by n and B is n by p)

Fun fact: complexity of matrix multiplication is still an open problem

- 2 by 2 matrix multiplication
  - Naïve algorithm takes 8 multiplication
  - **Strassen** showed that one can get away with 7
- Divide and conquer gives $O(n^{\log_2 7}) \approx O(n^{2.807})$
  - Improves over $O(n^3)$ for reasonable sized matrices
- Actually used in practice!

Timeline of matrix multiplication exponent

Year	Bound on omega	Authors
1969	2.8074	Strassen[1]
1978	2.796	Pan[11]
1979	2.780	Bini, Capovani [it], Romani[12]
1981	2.522	Schönhage[13]
1981	2.517	Romani[14]
1981	2.496	Coppersmith, Winograd[15]
1986	2.479	Strassen[16]
1990	2.3755	Coppersmith, Winograd[17]
2010	2.3737	Stothers[18]
2013	2.3729	Williams[19][20]
2014	2.3728639	Le Gall[21]
2020	2.3728596	Alman, Williams[6][22]
2022	2.371866	Duan, Wu, Zhou[3]
2023	2.371552	Williams, Xu, Xu, and Zhou[2]

Best lower bound is still $\Omega(n^2 \log n)$

Properties of a bag of vectors: linear independence

Important to consider for machine learning algorithm design

- Given a set of vectors $\{v_1, v_2, \cdots, v_n\} \in \mathbb{R}^m$, with $m \geq n$, consider the set of **linear combinations** $y = \sum_{j=1}^{n} \alpha_j v_j$ for arbitrary coefficients α_j's.
- The vectors $\{v_1, v_2, \cdots, v_n\}$ are **linearly independent**, if $\sum_{j=1}^{n} \alpha_j v_j = 0$, if and only if $\alpha_j = 0$ for all $j = 1, \cdots, n$.
- Implication: if a set of vectors are linearly dependent, then one of them can be written as a linear combination of the others

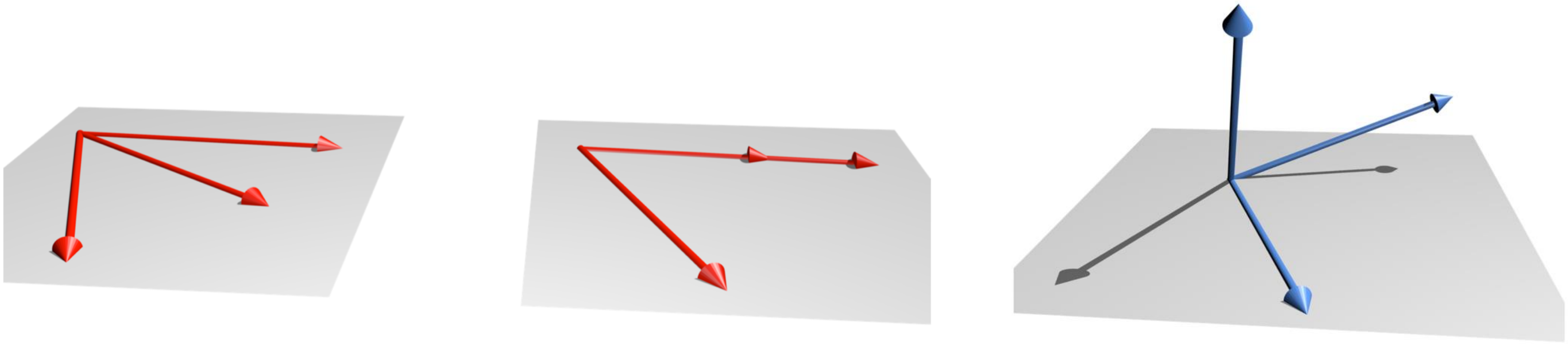
In-class exercise: linear independence

Are these vectors linear dependent?

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, v_2 = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}, v_3 = \begin{pmatrix} 3 \\ 1 \\ 4 \end{pmatrix}$$

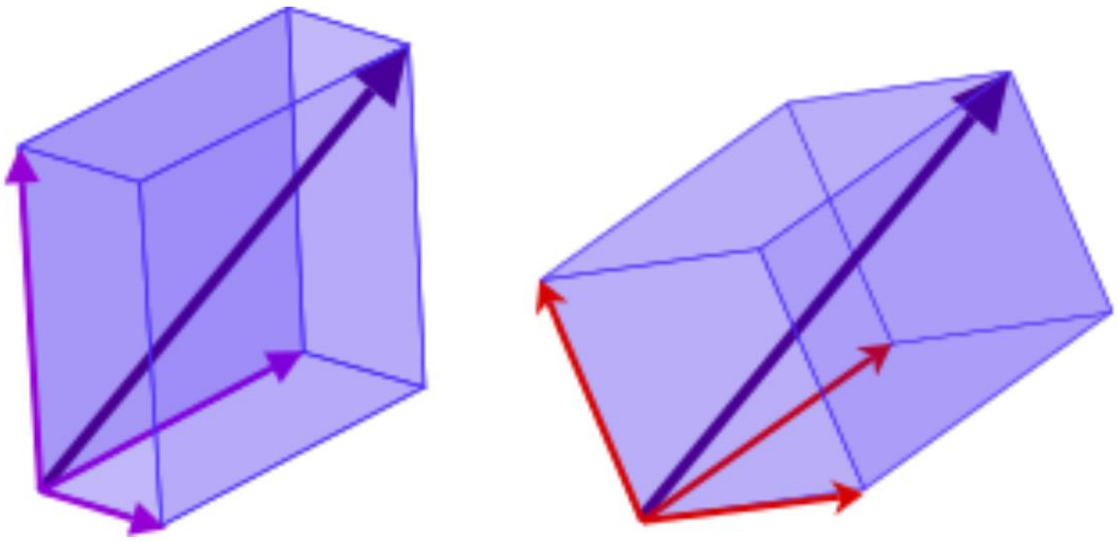
Yes, because that $2v_1 + v_2 - v_3 = 0$. Or equivalently, $v_3 = 2v_1 + v_2$.

Discussion: are these vectors linearly independent?



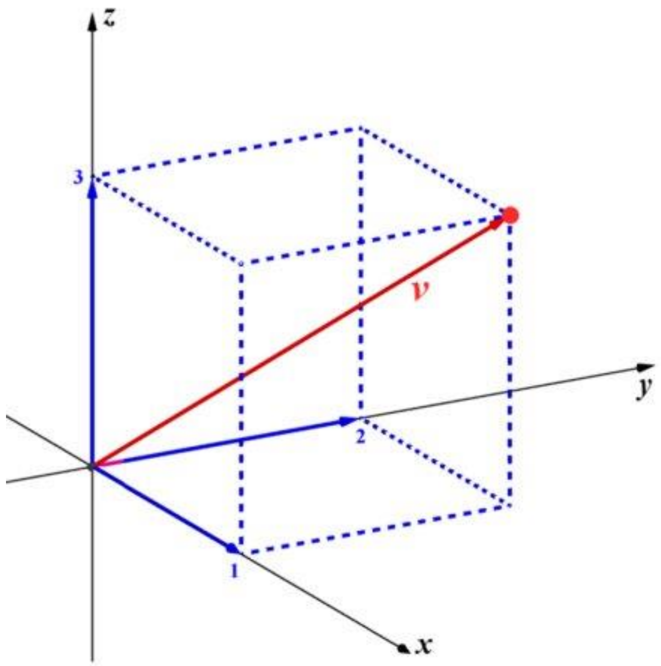
When they are linearly independent, we call this "bag" of vectors a basis. A basis of size m *spans* an m-dimensional vector space.

- A set of m linearly independent vectors of $\mathbb{R}^m$ is called a **basis** in $\mathbb{R}^m$: any vector in $\mathbb{R}^m$ can be expressed as a linear combination of the basis vectors.



Properties of basis

- Vectors in a basis are mutually orthogonal
  - Dot product of any two of them is 0.



$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Properties of a matrix

- General matrix
  - Rank: max number of independent column vectors / row vectors
  - Transpose: switch rows and columns

$$A \in \mathbb{R}^{m \times n} \qquad A^T \in \mathbb{R}^{n \times m}$$

- Square matrix
  - Trace: Sum of diagonal elements
  - Determinant:

$$tr\left(\begin{bmatrix} 5 & 3 & 5 \\ 4 & -1 & 2 \\ -3 & 8 & 7 \end{bmatrix}\right) = 5 - 1 + 7 = 11.$$

$$\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

Invertible matrix

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

Orthogonal matrix

$$A^{-1} = A^T$$

Symmetric matrix

$$A^T = A$$

Eigenvalues and eigenvectors of a (square) matrix

Let A be a $n \times n$ matrix. The vector $v \neq 0$ that satisfies

$$Av = \lambda v$$

for some scalar λ is called the eigenvector of A and λ is the eigenvalue corresponding to the eigenvector v.

1. A is symmetric, then $\lambda \in \mathbb{R}$.
2. A is symmetric and positive semi-definite, then $\lambda \geq 0$
3. A is symmetric and positive definite, then $\lambda > 0$

Positive (semi)-definite matrix

Very important property for optimization, kernel methods

- A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is positive semi-definite, if and only if $x^T A x \geq 0$, for any $x \in \mathbb{R}^n$.
  - All eigenvalues of A are non-negative.
  - $X^T A X$ for any $X \in \mathbb{R}^{n \times m}$ is positive semi-definite.
- A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is positive definite, if and only if $x^T A x > 0$, for any $0 \neq x \in \mathbb{R}^n$.
  - All eigenvalues of A are positive.
  - All diagonal entries of A are positive.