COMP 1433: Introduction to Data Analytics &
COMP 1003: Statistical Tools and Applications

Lecture 4 – Linear Algebra Basics

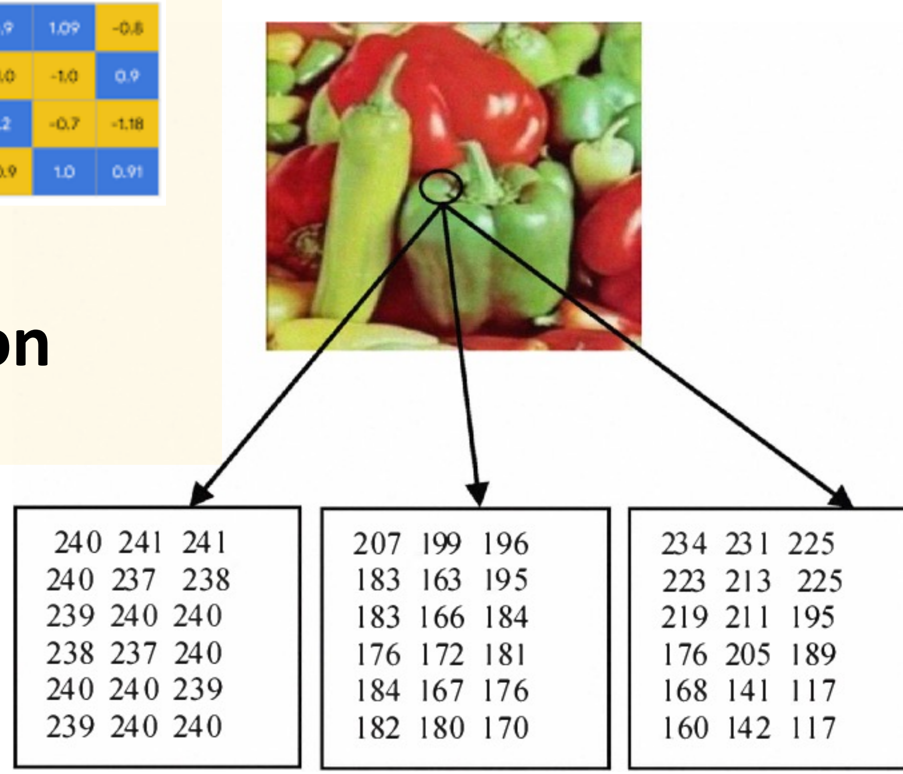
Dr. Jing Li
Department of Computing
The Hong Kong Polytechnic University

6&7 Feb 2023

Why Learn Linear Algebra

Product Recommendation

Computer Vision



More Examples?

Roadmap

- Vectors and Operations
  - Concepts
  - **Operations**: Addition, Scalar Multiplication, Dot Product, etc.
- Norm and Distance of Vectors
  - Definition
  - **Application**: Clustering
- Matrices
  - Concepts
  - **Operations**: Addition, Transpose, Multiplication, etc.

Roadmap

- **Vectors and Operations**
  - Concepts
  - **Operations**: Addition, Scalar Multiplication, Dot Product, etc.
- Norm and Distance of Vectors
  - Definition
  - **Application**: Clustering
- Matrices
  - Concepts
  - **Operations**: Addition, Transpose, Multiplication, etc.

Roadmap

- **Vectors and Operations**
  - **Concepts**
  - **Operations**: Addition, Scalar Multiplication, Dot Product, etc.
- Norm and Distance of Vectors
  - Definition
  - **Application**: Clustering
- Matrices
  - Concepts
  - **Operations**: Addition, Transpose, Multiplication, etc.

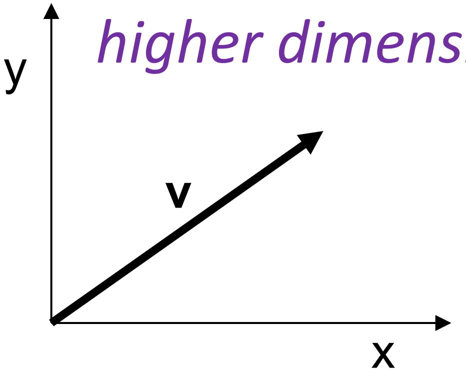
What is a *Vector*?

- A vector is an ordered list of *numbers*, such as
  - $(-1 \quad 0 \quad 3.6 \quad 7.2)$ or $\begin{pmatrix} -1 \\ 0 \\ 3.6 \\ 7.2 \end{pmatrix}$

Elements or *entries*,
e.g., the 3rd entry is 3.6

- Seen as a *directed line segment* in n-dimensions.
  - Count of *entries*: *dimension*.
  - Vector above has dimension 4
  - Vectors of dimension n: *n-vector*.
  - Numbers are called *scalars*.
  - Denoted as symbols, such as a, b, c, ...

Algebra in a higher dimension.



Example: Word Count Vector

A short sentence.

Word count vectors are used **in** computer based **document** analysis.

Dictionary

Dictionary	Word Count Vector
word	1
in	1
number	0
house	0
the	0
document	1

Word Count Vector

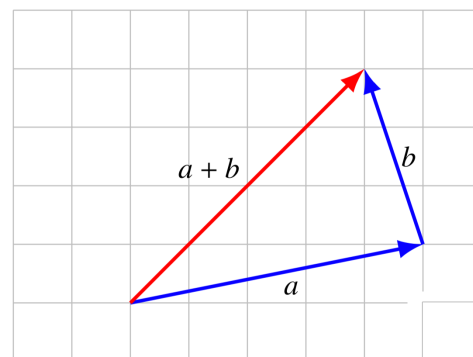
$$\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

Roadmap

- **Vectors and Operations**
  - Concepts
  - **Operations: Addition, Scalar Multiplication, Dot Product, etc.**
- Norm and Distance of Vectors
  - Definition
  - **Application**: Clustering
- Matrices
  - Concepts
  - **Operations**: Addition, Transpose, Multiplication, etc.

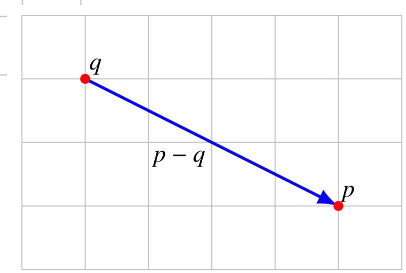
Vector Addition

- n-vectors a and b can be added, the sum is $a + b$
- Add corresponding entries to get the sum
  - e.g., $\begin{pmatrix} 0 \\ 7 \\ 3 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 9 \\ 3 \end{pmatrix}$



Head-to-tail methods

- Subtraction is similar.



- Properties:
  - ***Communicative***. $a + b = b + a$
  - ***Associative***. $(a + b) + c = a + (b + c)$
  - $a + 0 = 0 + a$
  - $a - a = 0$

0 is a zero vector with all entries as 0

Example: Word Count Vector Addition

A sentences. Yet another sentence.

Word count vectors are used **in** computer based **document** analysis.

Each entry of **the** **word** count vector is **the** **number** of times **the** associated dictionary **word** appears **in** **the** **document**.

Dictionary

word
in
number
house
the
document

$$\begin{pmatrix}1\\1\\0\\0\\0\\1\end{pmatrix}+\begin{pmatrix}2\\1\\1\\0\\4\\1\end{pmatrix}=\begin{pmatrix}3\\2\\1\\0\\4\\2\end{pmatrix}$$

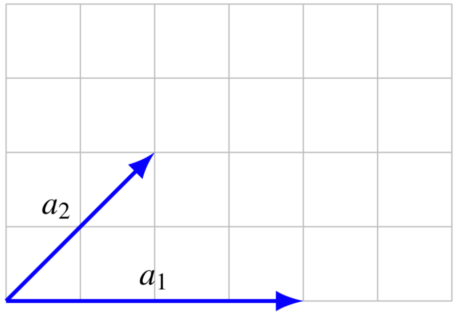
Word Count Vector Addition

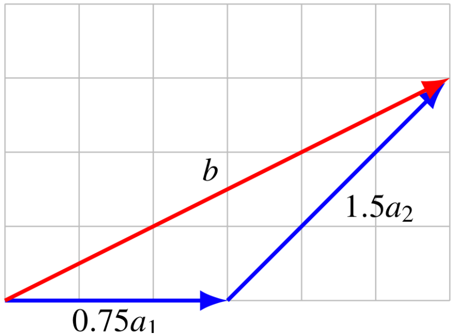
Scalar-Vector Multiplication

- Scalar β and n-vector a can be multiplied
  - $\beta a = (\beta a_1, \beta a_2, \ldots, \beta a_n)$
  - E.g., $(-2)\begin{pmatrix}1\\9\\6\end{pmatrix} = \begin{pmatrix}-2\\-18\\-12\end{pmatrix}$
- ***Associative***. $(\beta\gamma)a = \beta(\gamma a)$
- ***Left Distributive***. $(\beta + \gamma)a = \beta a + \gamma a$.
- ***Right Distributive***. $\beta(a + b) = \beta a + \beta b$

Linear Combination

- For vectors $a_1, a_2, \ldots, a_m$ and scalars $\beta_1, \beta_2, \ldots, \beta_m$
- We define linear combination of the vectors as:
  - $\beta_1 a_1 + \beta_2 a_2 + \cdots + \beta_m a_m$
- $\beta_1, \beta_2, \ldots, \beta_m$ are *coefficients*
- A simple identity, for any n-vector b,
  - $b = b_1 e_1 + b_2 e_2 + \cdots + b_n e_n$
  - e_i is a *unit vector* with 1 at the i-th entry and others 0





$$b = 0.75a_1 + 1.5a_2$$

Example: Scalar-Vector Multiplication

Two sentences, where their weights vary.

Word count vectors are used **in** computer based **document** analysis. *Weight 0.75*

Each entry of **the** **word** count vector is **the** **number** of times **the** associated dictionary **word** appears **in** **the** **document**. *Weight 0.25*

word
in
number
house
the
document

$$0.75\begin{pmatrix}1\\1\\0\\0\\0\\1\end{pmatrix}+0.25\begin{pmatrix}2\\1\\1\\0\\4\\1\end{pmatrix}=\begin{pmatrix}1.25\\1\\0.25\\0\\1\\1\end{pmatrix}$$

Weighted Addition

Example: Weights on Words

Two sentences, where their weights vary.

Word count vectors are used **in** computer based **document** analysis.

Each entry of **the** **word** count vector is **the** **number** of times **the** associated dictionary **word** appears **in** **the** **document**.

word 0.1
in 0.8
number 0.2
house 0.1
the 0.9
document 0.1

$$(0.1 \quad 0.8 \quad 0.2 \quad 0.1 \quad 0.9 \quad 0.1)\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = 1$$

Example: Weights on Words

Two sentences, where their weights vary.

Word count vectors are used **in** computer based **document** analysis.

Each entry of **the** **word** count vector is **the** **number** of times **the** associated dictionary **word** appears **in** **the** **document**.

word 0.1
in 0.8
number 0.2
house 0.1
the 0.9
document 0.1

$$\begin{pmatrix} 0.1 & 0.8 & 0.2 & 0.1 & 0.9 & 0.1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 1 \\ 0 \\ 4 \\ 1 \end{pmatrix} = 5$$

Inner Product

- ***Inner Product*** (or dot product) of n-vectors a and b:
    - $a^T b = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n$
- Example:
    - $\begin{pmatrix} -1 \\ 2 \\ 2 \end{pmatrix}^T \begin{pmatrix} 1 \\ 0 \\ -3 \end{pmatrix} = -1 * 1 + 2 * 0 + 2 * (-3) = -7$
- **PROPERTIES**
    - $a^T b = b^T a$
    - $(\gamma a)^T b = \gamma(a^T b)$
    - $(a + b)^T c = a^T c + b^T c$
    - $(a + b)^T (c + d) = a^T c + b^T c + a^T d + b^T d$

Example: Inner Product

- *Pick out the i-th entry*: $e_i^T a = a_i$
- *Sum of entries*: $1^T a = a_1 + a_2 + \cdots + a_n$
- Sum of squares of entries:
  - $a^T a = a_1^2 + a_2^2 + \cdots + a_n^2$
- More examples.
  - w is a weight vector, f is feature vector, $w^T f$ is weighted score.
  - p is vector of prices, q is vector of quantities, $p^T q$ is total cost.

Roadmap

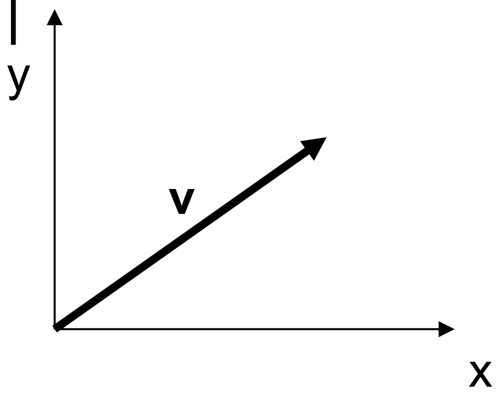
- Vectors and Operations
  - Concepts
  - Operations: Addition, Scalar Multiplication, Dot Product, etc.
- **Norm and Distance of Vectors**
  - Definition
  - **Application**: Clustering
- Matrices
  - Concepts
  - **Operations**: Addition, Transpose, Multiplication, etc.

Roadmap

- Vectors and Operations
  - Concepts
  - Operations: Addition, Scalar Multiplication, Dot Product, etc.
- **Norm and Distance of Vectors**
  - **Definition**
  - **Application**: Clustering
- Matrices
  - Concepts
  - **Operations**: Addition, Transpose, Multiplication, etc.

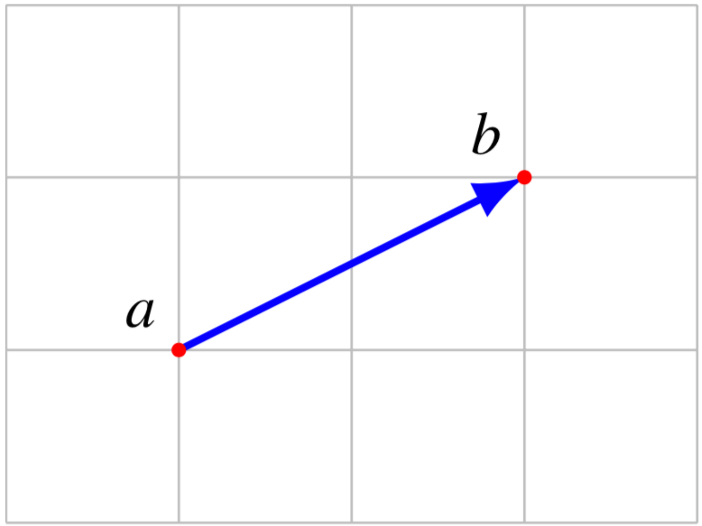
What is *Norm*?

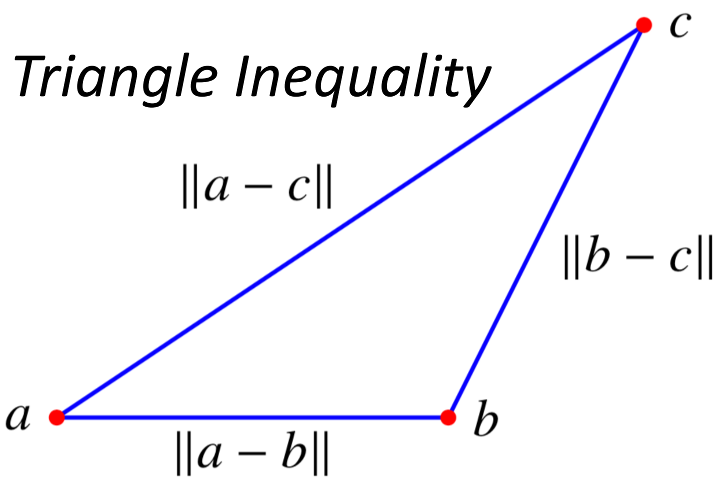
- The Euclidean norm (or norm) of an n-vector x is:
  - $||x|| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{x^T x}$
- Used to measure the *length* of a vector.
- **PROPERTIES.** For any n-vectors x, y and scalar β:
  - *Homogeneity*. $||\beta x|| = |\beta| ||x||$
  - *Triangle Inequality*. $||x + y|| \leq ||x|| + ||y||$
  - *Non-negativity*. $||x|| \geq 0$
  - *Definiteness*. $||x|| = 0$ only if $x = 0$



What is *Distance*?

- The Euclidean *distance* (or distance) of two n-vectors x and y is:
  - $\|x - y\| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}$
- *Length* of the subtraction of the two vectors.





Example: Document Distance

- 5 Wikipedia articles:
  - *Veterans Day, Memorial Day, Academy Awards, Golden Globe Awards, Super Bowl*
- Word count vectors with 4,423 words in dictionary.

Pairwise Distance

	Veterans Day	Memorial Day	Academy Awards	Golden Globe Awards	Super Bowl
Veterans Day	0	0.095	0.130	0.153	0.170
Memorial Day	0.095	0	0.122	0.147	0.164
Academy A.	0.130	0.122	0	0.108	0.164
Golden Globe A.	0.153	0.147	0.108	0	0.181
Super Bowl	0.170	0.164	0.164	0.181	0

Cauchy-Schwarz Inequality

- For two n-vectors a and b, $|a^T b| \leq \left|\left|a\right|\right| \cdot ||b||$
- Or $(\sum_{i=1}^{n} a_i b_i)^2 \leq (\sum_{i=1}^{n} a_i^2)(\sum_{i=1}^{n} b_i^2)$
- Now we can further look at *triangle inequality*.
  - $$\begin{aligned} \left|\left|a+b\right|\right|^2 &= \left|\left|a\right|\right|^2 + 2a^T b + \left|\left|b\right|\right|^2 \\ &\leq \left|\left|a\right|\right|^2 + 2\left|\left|a\right|\right| \cdot \left|\left|b\right|\right| + \left|\left|b\right|\right|^2 \\ &\leq \left(\left|\left|a\right|\right| + \left|\left|b\right|\right|\right)^2 \end{aligned}$$

What is an *angle*?

- Angle θ between two non-zero vectors a and b
  - $cos\theta = \frac{a^T b}{||a||\cdot||b||}$ where $0 \leq \theta \leq \pi$
- Several cases of θ:
  - $\theta = \frac{\pi}{2} = 90°$: a and b are orthogonal, i.e., $a \perp b$
  - $\theta = 0$: a and b are aligned. Here $a^T b = ||a|| \cdot ||b||$
  - $\theta = \pi = 180°$: a and b are anti-aligned. $a^T b = -||a|| \cdot ||b||$
  - $\theta \in (0, \frac{\pi}{2})$: a and b make an acute angle. $a^T b > 0$.
  - $\theta \in (\frac{\pi}{2}, \pi)$: a and b make an obtuse angle. $a^T b < 0$.

Example: Document Dissimilarity

- 5 Wikipedia articles:
  - *Veterans Day, Memorial Day, Academy Awards, Golden Globe Awards, Super Bowl*
- Word count vectors with 4,423 words in dictionary.

Pairwise Angles in Degrees

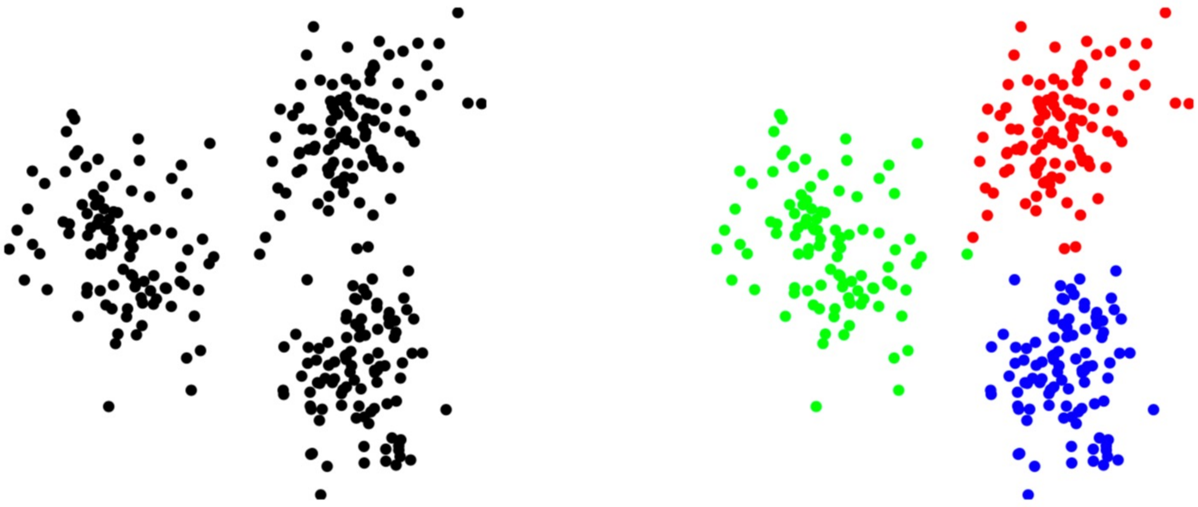
	Veterans Day	Memorial Day	Academy Awards	Golden Globe Awards	Super Bowl
Veterans Day	0	60.6	85.7	87.0	87.7
Memorial Day	60.6	0	85.6	87.5	87.5
Academy A.	85.7	85.6	0	58.7	85.7
Golden Globe A.	87.0	87.5	58.7	0	86.0
Super Bowl	87.7	87.5	86.1	86.0	0

Roadmap

- Vectors and Operations
  - Concepts
  - Operations: Addition, Scalar Multiplication, Dot Product, etc.
- **Norm and Distance of Vectors**
  - Definition
  - **Application: Clustering**
- Matrices
  - Concepts
  - **Operations**: Addition, Transpose, Multiplication, etc.

Clustering

- Given N n-vectors, $x_1, x_2, \ldots, x_N$
- Partition (*cluster*) them into k clusters
- Our goal is to let vectors in the same cluster to be close to each other.

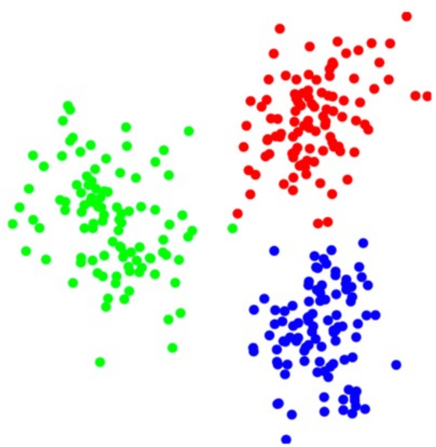


Clustering Objective

- Given N n-vectors, $x_1, x_2, \ldots, x_N$
- Partition (*cluster*) them into k clusters: $G_1, G_2, \ldots, G_k$
- **Group assignment**: c_i is the index of the group assigned to vector x_i, i.e., $x_i \in G_{c_i}$
- **Group representatives**:
  - n-vectors $z_1, z_2, \ldots, z_k$
- Clustering objective is:
  - $J^{cluster} = \frac{1}{N}\sum_{i=1}^{N} \left|\left|x_i - z_{c_i}\right|\right|^2$
  - Smaller, the better!

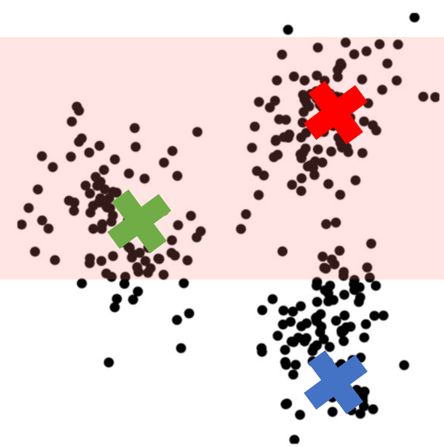
Clustering Objective

- Given N n-vectors, $x_1, x_2, \dots, x_N$
- Partition (*cluster*) them into k clusters: $G_1, G_2, \dots, G_k$
- **Group assignment**: c_i is the index of the group assigned to vector x_i, i.e., $x_i \in G_{c_i}$
- **Group representatives**:
  - n-vectors $z_1, z_2, \dots, z_k$
- Clustering objective is:
  - $J^{cluster} = \frac{1}{N} \sum_{i=1}^{N} \left|\left|x_i - z_{c_i}\right|\right|^2$
  - z_{c_i} should be the mean of G_{c_i}



Clustering Objective

- Given N n-vectors, $x_1, x_2, \ldots, x_N$
- Partition (*cluster*) them into k clusters: $G_1, G_2, \ldots, G_k$
- **Group assignment**: c_i is the index of the group assigned to vector x_i, i.e., $x_i \in g_{c_i}$
- **Group representatives**:
  - n-vectors $z_1, z_2, \ldots, z_k$
- Clustering objective is:
  - $J^{cluster} = \frac{1}{N}\sum_{i=1}^{N} \left|\left|x_i - z_{c_i}\right|\right|^2$
  - Align x_i to be in the same group as the closet representative.



K-means Clustering Algorithm

- Alternatively updating the group assignment, then the representatives.
- $J^{cluster}$ goes down in each step.
- No guarantee to minimize $J^{cluster}$

---

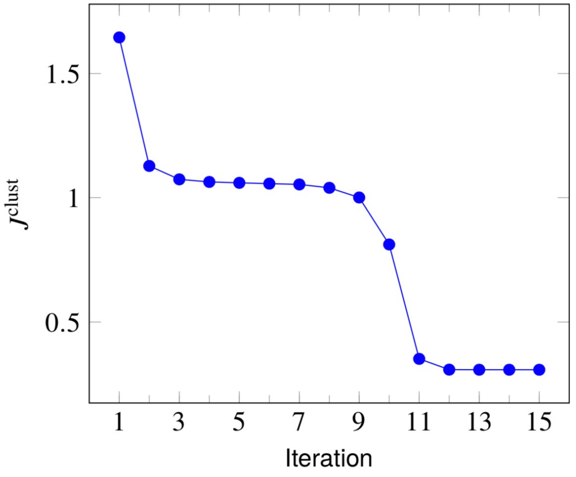
given $x_1, \ldots, x_N \in \mathbf{R}^n$ and $z_1, \ldots, z_k \in \mathbf{R}^n$

repeat

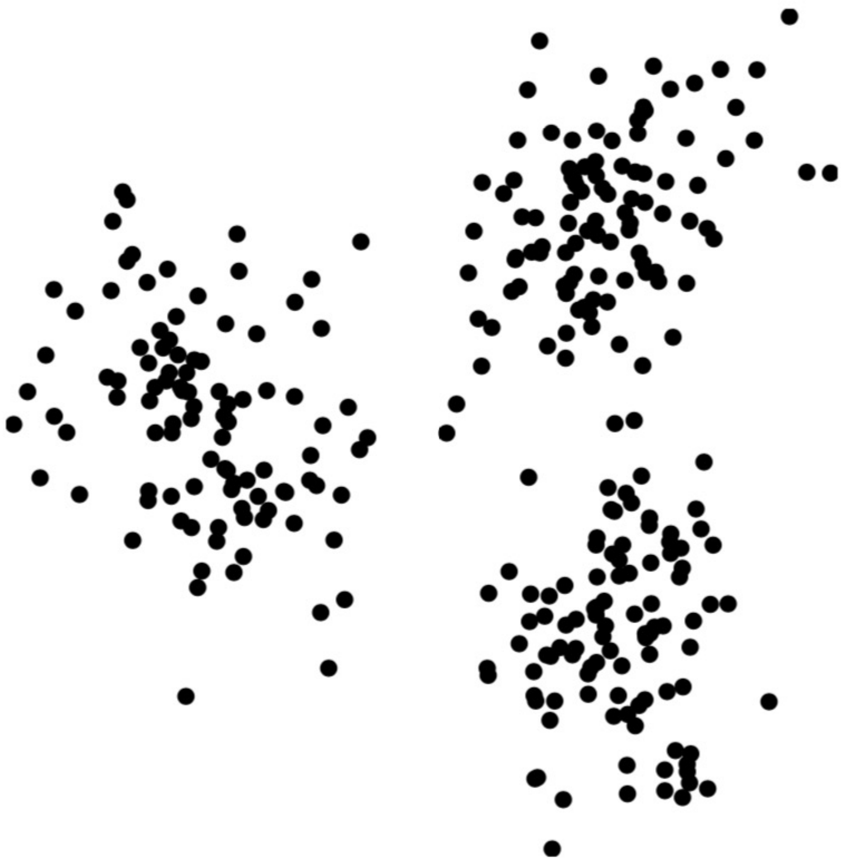
Update partition: assign i to G_j, $j = \operatorname{argmin}_{j'} \|x_i - z_{j'}\|^2$

Update centroids: $z_j = \frac{1}{|G_j|} \sum_{i \in G_j} x_i$

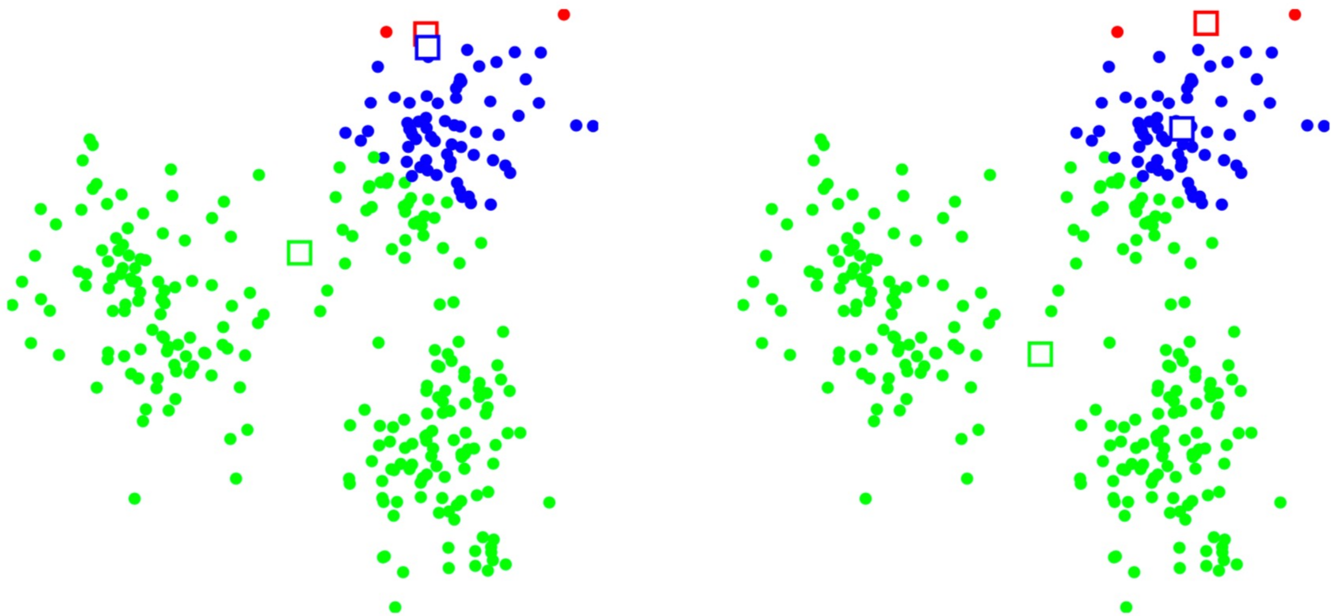
until $z_1, \ldots, z_k$ stop changing



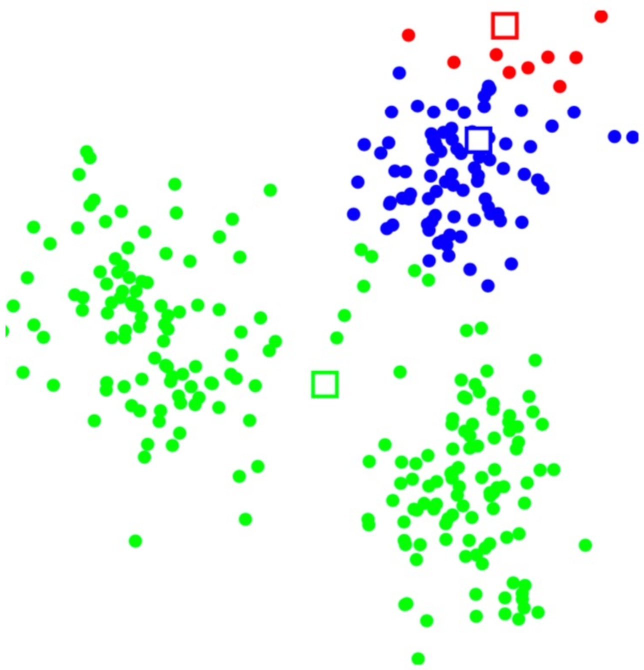
Running K-means Clustering (at beginning)

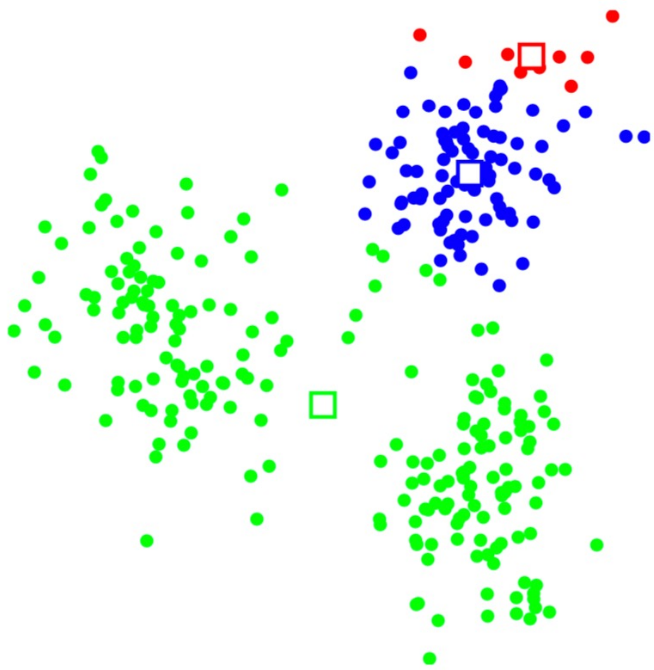


Running K-means Clustering (Iteration 1)

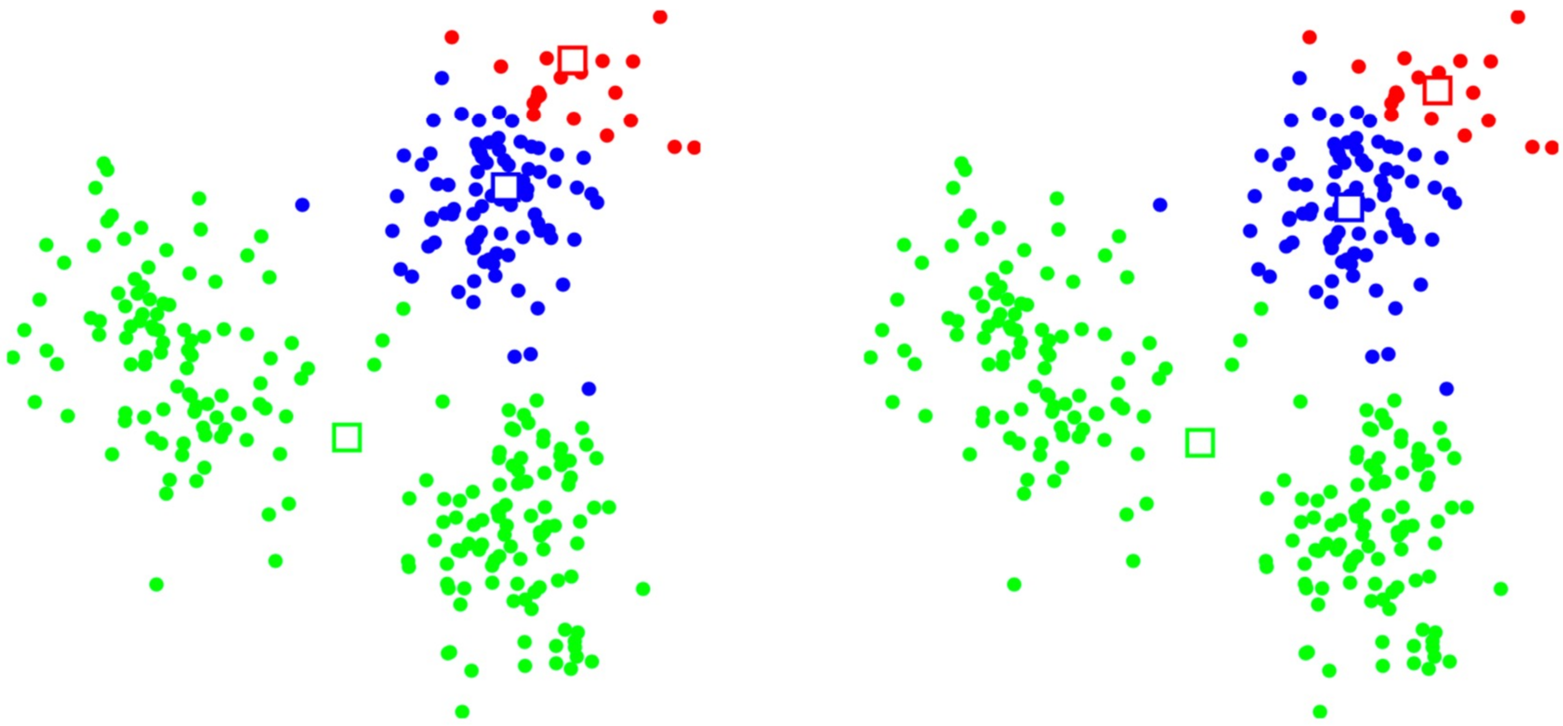


Running K-means Clustering (Iteration 2)

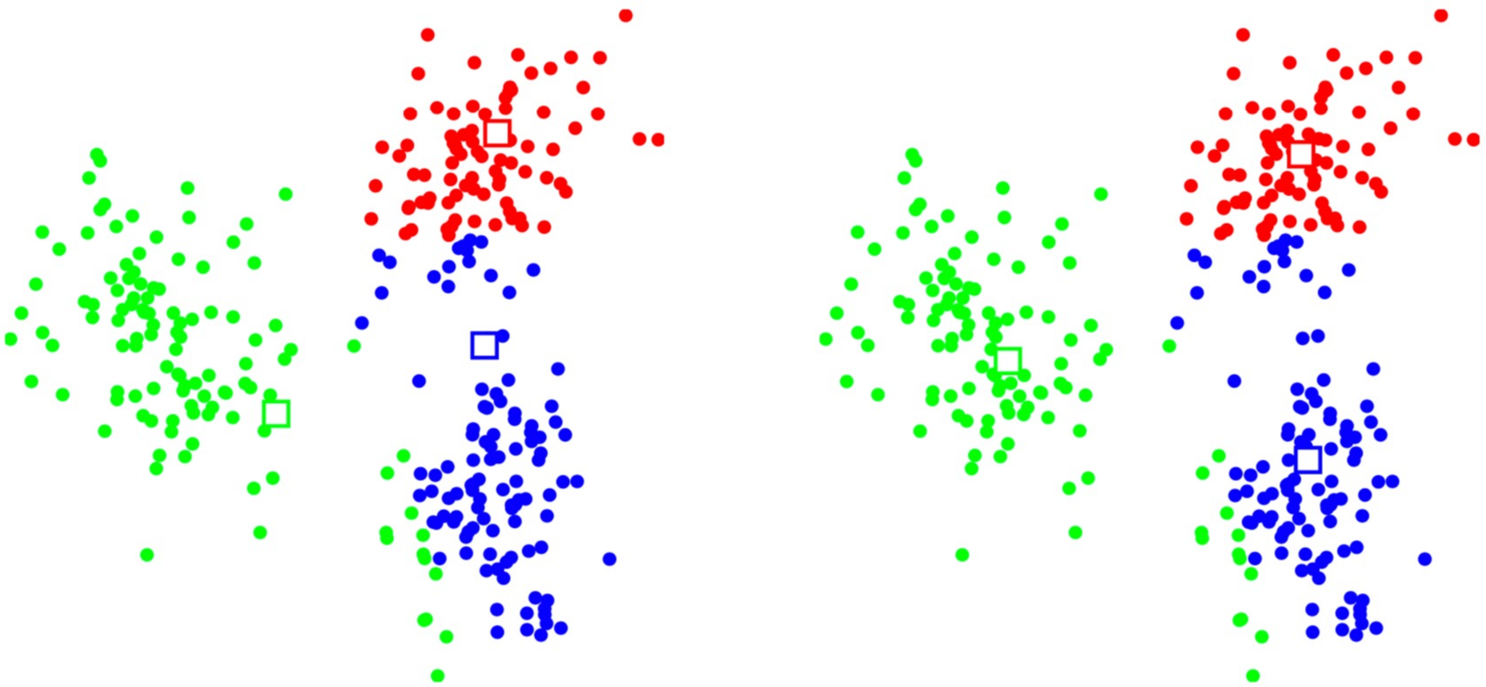




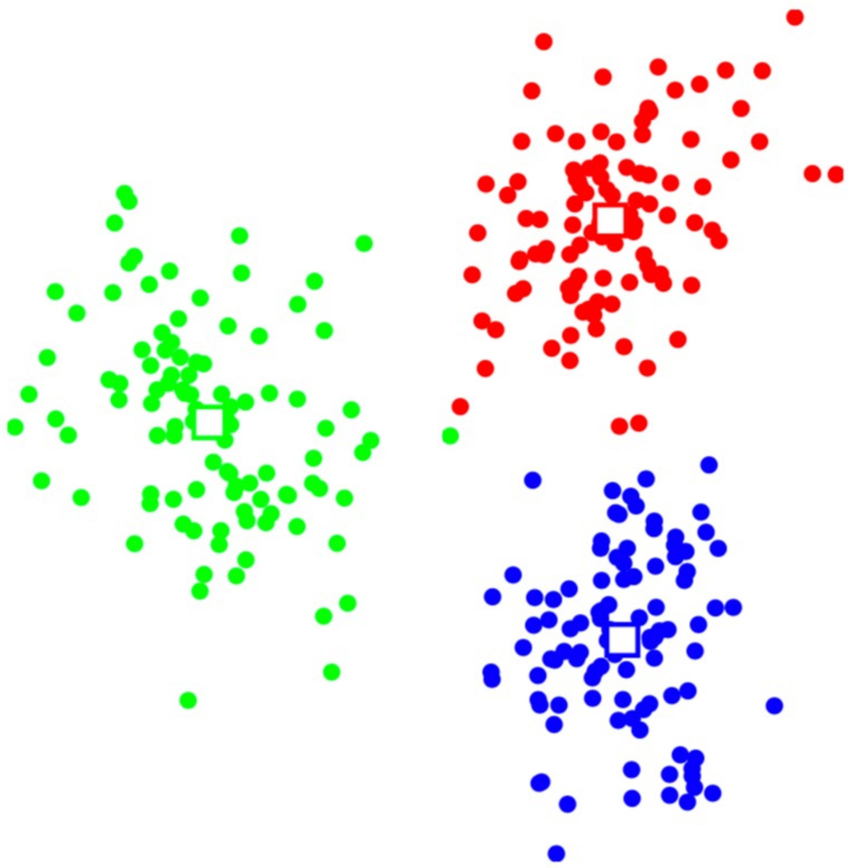
Running K-means Clustering (Iteration 3)



Running K-means Clustering (Iteration 10)



Running K-means Clustering (At last)



Example: Topic Discovery

- $N = 500$ Wikipedia articles
- Dictionary size $n = 4423$
- Run K-means algorithm with $k = 9$.
- **Results:**
  - Top words in the cluster representatives, mean of word vectors in the cluster.
  - Titles of articles closest to the representatives.

39

Example: Topic Discovery (C1-3)

Cluster 1		Cluster 2		Cluster 3	
Word	Coef.	Word	Coef.	Word	Coef.
fight	0.038	holiday	0.012	united	0.004
win	0.022	celebrate	0.009	family	0.003
event	0.019	festival	0.007	party	0.003
champion	0.015	celebration	0.007	president	0.003
fighter	0.015	calendar	0.006	government	0.003

Top 5 words in the cluster representatives --- mean of word vectors in the cluster in the normalized form).

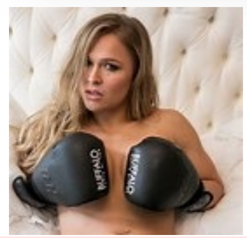
Titles of articles closest to the representatives.

1. “Floyd Mayweather, Jr”, “Kimbo Slice”, “Ronda Rousey”, “José Aldo”, “Joe Frazier”, “Wladimir Klitschko”, “Saul Álvarez”, “Gennady Golovkin”, “Nate Diaz”, . . .
2. “Halloween”, “Guy Fawkes Night” “Diwali”, “Hanukkah”, “Groundhog Day”, “Rosh Hashanah”, “Yom Kippur”, “Seventh-day Adventist Church”, “Remembrance Day”, . . .
3. “Mahatma Gandhi”, “Sigmund Freud”, “Carly Fiorina”, “Frederick Douglass”, “Marco Rubio”, “Christopher Columbus”, “Fidel Castro”, “Jim Webb”, . . .

Example: Topic Discovery (C1-3)

Cluster 1		Cluster 2		Cluster 3	
Word	Coef.	Word	Coef.	Word	Coef.
fight	0.038	holiday	0.012	united	0.004
win	0.022	celebrate	0.009	family	0.003
event	0.019	festival	0.007	party	0.003
champion	0.015	celebration	0.007	president	0.003
fighter	0.015	calendar	0.006	government	0.003



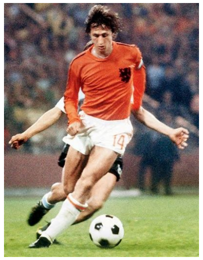
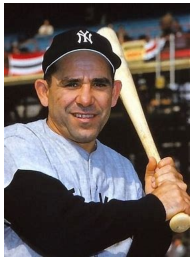



1. “Floyd Mayweather, Jr”, “Kimbo Slice”, “Ronda Rousey”, “José Aldo”, “Joe Frazier”, “Wladimir Klitschko”, “Saul Álvarez”, “Gennady Golovkin”, “Nate Diaz”, . . .
2. “Halloween”, “Guy Fawkes Night” “Diwali”, “Hanukkah”, “Groundhog Day”, “Rosh Hashanah”, “Yom Kippur”, “Seventh-day Adventist Church”, “Remembrance Day”, . . .
3. “Mahatma Gandhi”, “Sigmund Freud”, “Carly Fiorina”, “Frederick Douglass”, “Marco Rubio”, “Christopher Columbus”, “Fidel Castro”, “Jim Webb”, . . .

Example: Topic Discovery (C4-6)

Cluster 4		Cluster 5		Cluster 6	
Word	Coef.	Word	Coef.	Word	Coef.
album	0.031	game	0.023	series	0.029
release	0.016	season	0.020	season	0.027
song	0.015	team	0.018	episode	0.013
music	0.014	win	0.017	character	0.011
single	0.011	player	0.014	film	0.008

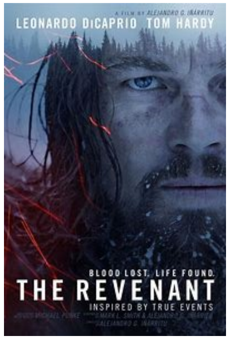



1. "David Bowie", "Kanye West" "Celine Dion", "Kesha", "Ariana Grande", "Adele", "Gwen Stefani", "Anti (album)", "Dolly Parton", "Sia Furler", . . .
2. "Kobe Bryant", "Lamar Odom", "Johan Cruyff", "Yogi Berra", "José Mourinho", "Halo 5: Guardians", "Tom Brady", "Eli Manning", "Stephen Curry", "Carolina Panthers", . . .
3. "The X-Files", "Game of Thrones", "House of Cards (U.S. TV series)", "Daredevil (TV series)", "Supergirl (U.S. TV series)", "American Horror Story", . . .

Example: Topic Discovery (C7-9)

Cluster 7		Cluster 8		Cluster 9	
Word	Coef.	Word	Coef.	Word	Coef.
match	0.065	film	0.036	film	0.061
win	0.018	star	0.014	million	0.019
championship	0.016	role	0.014	release	0.013
team	0.015	play	0.010	star	0.010
event	0.015	series	0.009	character	0.006



1. “Wrestlemania 32”, “Payback (2016)”, “Survivor Series (2015)”, “Royal Rumble (2016)”, “Night of Champions (2015)”, “Fastlane (2016)”, “Extreme Rules (2016)”, . . .
2. “Ben Affleck”, “Johnny Depp”, “Maureen O’Hara”, “Kate Beckinsale”, “Leonardo DiCaprio”, “Keanu Reeves”, “Charlie Sheen”, “Kate Winslet”, “Carrie Fisher”, . . .
3. “Star Wars: The Force Awakens”, “Star Wars Episode I: The Phantom Menace”, “The Martian (film)”, “The Revenant (2015 film)”, “The Hateful Eight”, . . .

Roadmap

- Vectors and Operations
  - Concepts
  - Operations: Addition, Scalar Multiplication, Dot Product, etc.
- Norm and Distance of Vectors
  - Definition
  - Application: Clustering
- **Matrices**
  - Concepts
  - **Operations**: Addition, Transpose, Multiplication, etc.

Roadmap

- Vectors and Operations
  - Concepts
  - Operations: Addition, Scalar Multiplication, Dot Product, etc.
- Norm and Distance of Vectors
  - Definition
  - Application: Clustering
- **Matrices**
  - **Concepts**
  - **Operations**: Addition, Transpose, Multiplication, etc.

What is a *Matrix*?

- A *matrix* is a rectangular array of numbers:
  - $\begin{pmatrix} 0 & 1 & -2.3 & 0.1 \\ 1.3 & 4 & -0.1 & 0 \\ 4.1 & -1 & 0 & 1.7 \end{pmatrix}$
- Its *size* is given by the *(row dimension)×(column dimension)*, say 3×4 for the above example.
- *Elements* are also called *entries*.
- $B_{i,j}$ is the entry at the i-th row and j-th column.
- Two matrices are the *equal* (=) if they have *the same size* and *all corresponding entries are equal*.

What is a *Matrix*?

- A *matrix* is a rectangular array of numbers:
  - $\begin{pmatrix} 0 & 1 & -2.3 & 0.1 \\ 1.3 & 4 & -0.1 & 0 \\ 4.1 & -1 & 0 & 1.7 \end{pmatrix}$
- Its *size* is given by the *(row dimension)×(column dimension)*, say 3×4 for the above example.
  - *Tall* if $m > n$
  - *Wide* if $m < n$
  - *Square* if $m = n$

Matrix and Vectors

- We consider a $n \times 1$ matrix to be *n-vector* (or *column vector*).
- We consider a *1×1* matrix to be a *number*.
- A *$1 \times n$* matrix is defined as a *row vector*.
  - E.g., $\begin{pmatrix} 1.2 & -0.3 & 1.4 & 2.6 \end{pmatrix}$
  - It should be distinguished from the column vector, e.g.,
  - $\begin{pmatrix} 1.2 \\ -0.3 \\ 1.4 \\ 2.6 \end{pmatrix}$

Columns and Rows of a Matrix

- Suppose A is an $m \times n$ matrix with entries $A_{i,j}$
- Its j-th column is (an m-vector): $\begin{pmatrix} A_{1,j} \\ A_{2,j} \\ \dots \\ A_{m,j} \end{pmatrix}$
- Its i-th row is (an n-row-vector): $(A_{i,1}\ A_{i,2}\ \ \dots A_{i,n})$
- *Slice* of matrix: $A_{p:q,r:s}$ is a $(q-p+1)\times(s-r+1)$ matrix:
    - $A_{p:q,r:s} = \begin{pmatrix} A_{p,r} & \cdots & A_{p,s} \\ \vdots & \ddots & \vdots \\ A_{q,r} & \cdots & A_{q,s} \end{pmatrix}$

Block Matrices

- We can form *block matrices* (for simplicity) whose entries are matrices, such as
  - $A = \begin{pmatrix} B & C \\ D & E \end{pmatrix}$, where B, C, D, E are matrices (called submatrices or blocks of A)
  - Matrices in the the same block row must have the same height (i.e., row dimension)
  - Matrices in the the same block column must have the same width (i.e., column dimension)
  - **Example**. $B = (0\ 2\ 3), \mathrm{C} = (-1), \mathrm{D} = \begin{pmatrix} 2 & 2 & 1 \\ 1 & 3 & 5 \end{pmatrix}, \mathrm{E} = \begin{pmatrix} 4 \\ 4 \end{pmatrix}$
  - $A = \begin{pmatrix} B & C \\ D & E \end{pmatrix} = \begin{pmatrix} 0 & 2 & 3 & -1 \\ 2 & 2 & 1 & 4 \\ 1 & 3 & 5 & 4 \end{pmatrix}$

Column and Row Representation

- Suppose A is an $m \times n$ matrix with entries $A_{i,j}$
- Can express as block matrix with its (m-vector) columns $a_1, a_2, \ldots, a_n$.
  - $A = (a_1 \; a_2 \; \ldots a_n)$
- Can also express as block matrix with its (n-row-vector) rows $b_1, b_2, \ldots, b_m$.
  - $A = \begin{pmatrix} b_1 \\ b_2 \\ \ldots \\ b_m \end{pmatrix}$

Example: Word Count Matrix

- We are examining are n sentences.
- The dictionary size is m words.
- How can you represent the count of each word in different sentences?
  - We define a $m \times n$ matrix A
  - $A_{i,j}$ denotes the count of the i-th word in dictionary occurring in the j-th sentence.
  - **QUESTION**. What do the rows and columns mean?

word
in
number
house
the
document

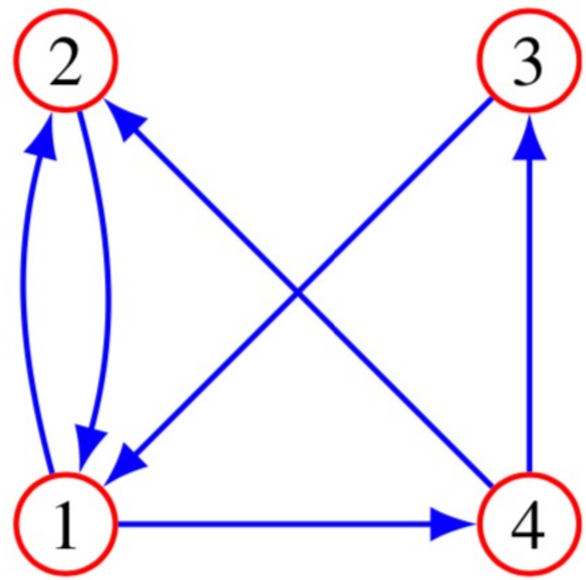
Dictionary

Sentence 1 **Sentence 2**

$$\begin{pmatrix} 1 & 2 \\ 1 & 1 \\ 0 & 1 \\ 0 & 0 \\ 0 & 4 \\ 1 & 1 \end{pmatrix}$$

Exercise

- Four nodes 1,2,3,4 and their connections in arrows.
- How to use matrix to represent the connections among the four nodes?



Roadmap

- Vectors and Operations
  - Concepts
  - Operations: Addition, Scalar Multiplication, Dot Product, etc.
- Norm and Distance of Vectors
  - Definition
  - Application: Clustering
- **Matrices**
  - Concepts
  - **Operations: Addition, Transpose, Multiplication, etc.**

Transpose of Matrices

- The *transpose* of an $m \times n$ matrix A is denoted as A^T, where $(A^T)_{i,j} = A_{j,i}$, for all possible i, j.
- For example, $\begin{pmatrix} 0 & 4 \\ 7 & 0 \\ 3 & 1 \end{pmatrix}^T = \begin{pmatrix} 0 & 7 & 3 \\ 4 & 0 & 1 \end{pmatrix}$
- Transpose converts columns to row vectors (and vice versa)
- $(A^T)^T = A$

Addition, Subtraction, and Scalar Multiplication of Matrices

- We can *add* or *subtract* matrices with the same size:
  - $(A + B)_{i,j} = A_{i,j} + B_{i,j}$ for all i, j
  - $(A - B)_{i,j} = A_{i,j} - B_{i,j}$ for all i, j
- For *scalar multiplication*:
  - $(\alpha A)_{i,j} = \alpha A_{i,j}$
- **PROPERTIES**.
  - $A + B = B + A$
  - $\alpha(A + B) = \alpha A + \alpha B$
  - $(A + B)^T = A^T + B^T$

Matrix–Vector Product

- Matrix-Vector Product of $m \times n$ matrix A and n-vector x, denoted as $y = Ax$, with
    - $y_i = A_{i,1}x_1 + A_{i,2}x_2 + \cdots + A_{i,n}x_n$
    - For example,
    - $\begin{pmatrix} 0 & 2 & -1 \\ -2 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 3 \\ -4 \end{pmatrix}$

Example: Vector-Matrix Product

Two sentences, where their weights vary.

Word count vectors are used **in** computer based **document** analysis.

Each entry of **the** **word** count vector is **the** **number** of times **the** associated dictionary **word** appears **in** **the** **document**.

word 0.1; 1
in 0.8;0
number 0.2;0
house 0.1;0
the 0.9;0
document 0.1;1

$$\begin{pmatrix} 0.1 & 0.8 & 0.2 & 0.1 & 0.9 & 0.1 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

Example: Vector-Matrix Product

Two sentences, where their weights vary.

Word count vectors are used **in** computer based **document** analysis.

Each entry of **the** **word** count vector is **the** **number** of times **the** associated dictionary **word** appears **in** **the** **document**.

word 0.1; 1
in 0.8;0
number 0.2;0
house 0.1;0
the 0.9;0
document 0.1;1

$$\begin{pmatrix} 0.1 & 0.8 & 0.2 & 0.1 & 0.9 & 0.1 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 1 \\ 0 \\ 4 \\ 1 \end{pmatrix} = \begin{pmatrix} 5 \\ 3 \end{pmatrix}$$

Example: Matrix Multiplication

Two sentences, where their weights vary.

Word count vectors are used **in** computer based **document** analysis.

Each entry of **the** **word** count vector is **the** **number** of times **the** associated dictionary **word** appears **in** **the** **document**.

word 0.1; 1
in 0.8;0
number 0.2;0
house 0.1;0
the 0.9;0
document 0.1;1

$$\begin{pmatrix} 0.1 & 0.8 & 0.2 & 0.1 & 0.9 & 0.1 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 1 & 1 \\ 0 & 1 \\ 0 & 0 \\ 0 & 4 \\ 1 & 1 \end{pmatrix} \overset{?}{=} \begin{pmatrix} 1 & 5 \\ 2 & 3 \end{pmatrix}$$

Matrix Multiplication

- We can multiply $m \times p$ matrix A and $p \times n$ matrix B:
  - $C = AB$ where $C_{i,j} = \sum_{k=1}^{p} A_{i,k} B_{k,j}$ for any i, j
  - Move along the i-th row of A and the j-th column of B
  - **Example**.
  - $\begin{pmatrix} -1.5 & 3 & 2 \\ 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} -1 & -1 \\ 0 & -2 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 3.5 & -4.5 \\ -1 & 1 \end{pmatrix}$

Block Matrix Multiplication

- Block matrices can be multiplied in the same way:
  - $\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} E & F \\ G & H \end{pmatrix} = \begin{pmatrix} AE + BG & AF + BH \\ CE + DG & CF + DH \end{pmatrix}$
- For example,
  - $AB = A(b_1 \quad b_2 \quad \cdots \quad b_n) = (Ab_1 \quad Ab_2 \quad \cdots \quad Ab_n)$
  - So AB is the *batch multiply* of A times columns of B
- Another example,
  - $A = \begin{pmatrix} a_1^T \\ a_2^T \\ \cdots \\ a_m^T \end{pmatrix} B = (b_1 \quad b_2 \quad \cdots \quad b_n)$
  - $(AB)_{i,j} = a_i^T b_j$

Properties of Matrix Multiplication

- *Associative*: $(AB)C = A(BC)$
- *Left Distributive*: $A(B + C) = AB + AC$
- *Right Distributive:* $(B + C)A = BA + CA$
- ~~*Communicative:*~~ $AB = BA$ *does not hold generally*
- $(AB)^T = B^T A^T$
- $AI = A$
- $IA = A$

I is named as an *identify matrix* $\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$, where

- I is a square matrix
- $I_{i,i} = 1$ for all $1 \le i \le n$
- $I_{i,j} = 0$ for $i \ne j$ and$1 \le i, j \le n$

A slide to takeaway

- What are *scalars* and *vectors*?
- How to do *addition*, *scalar multiplication*, and *dot production* for vectors?
- How to determine the *norm of a vector* and *distance (dissimilarity) of two vectors*?
- How to cluster data vectors?
- What are *matrices*?
- How to do *addition*, *scalar multiplication*, *transpose*, and *multiplication* for matrices?