

Latent Semantic Analysis

Introduction to SVD and LSA

Latent Semantic Analysis

“Latent Semantic Analysis is a technique of analysing relationships between a set of documents and the terms they contain by producing a set of concepts related to the documents and terms.”

- Wikipedia

Latent Semantic Analysis

Music

Food

News

Technology

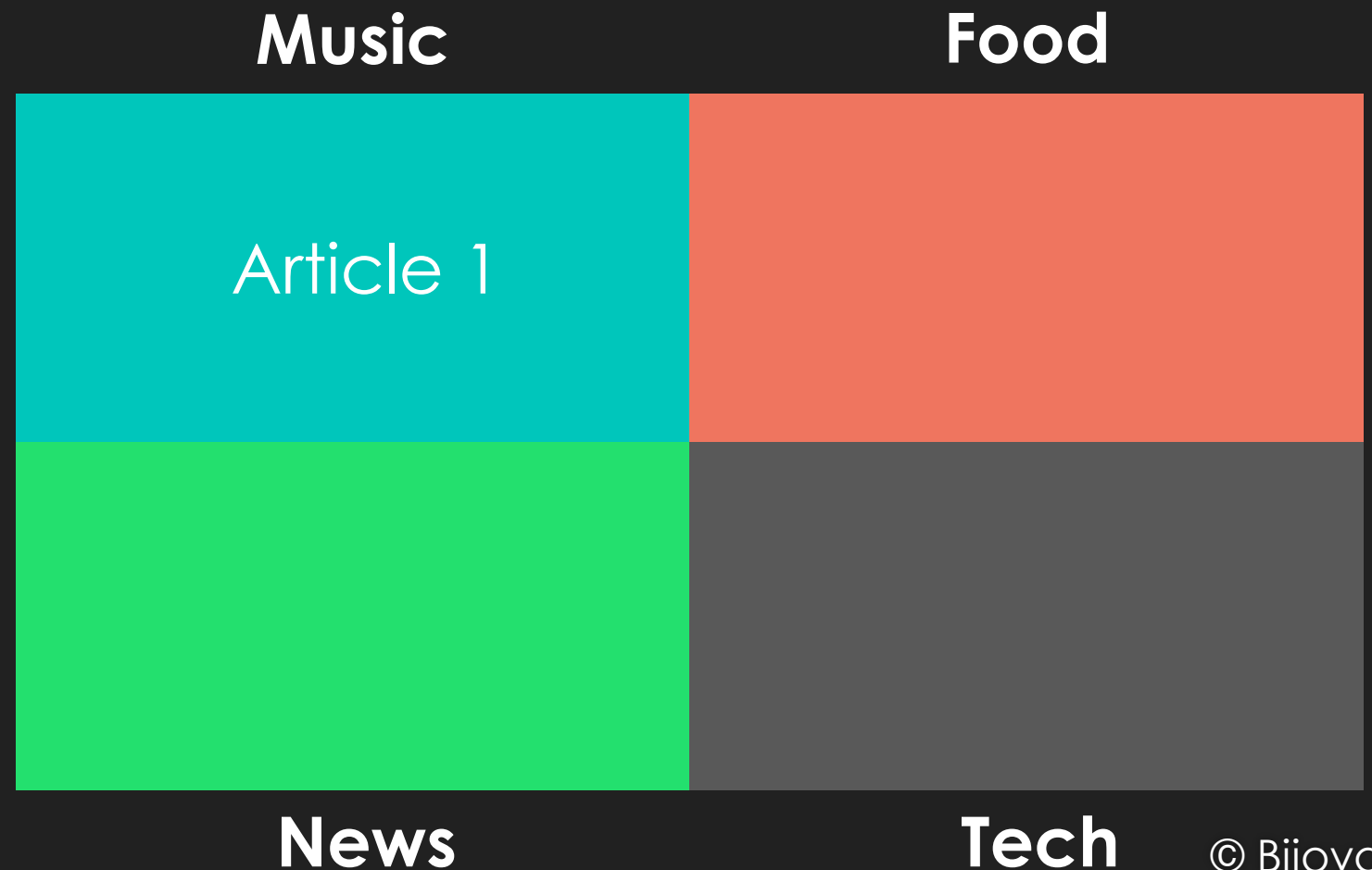
Latent Semantic Analysis

1. Article 1
2. Article 2
3. Article 3
4. Article 4
5. Article 5
6. Article 6
-

	Music	Food
News		
Tech		

Latent Semantic Analysis

- 1.
2. Article 2
3. Article 3
4. Article 4
5. Article 5
6. Article 6
-



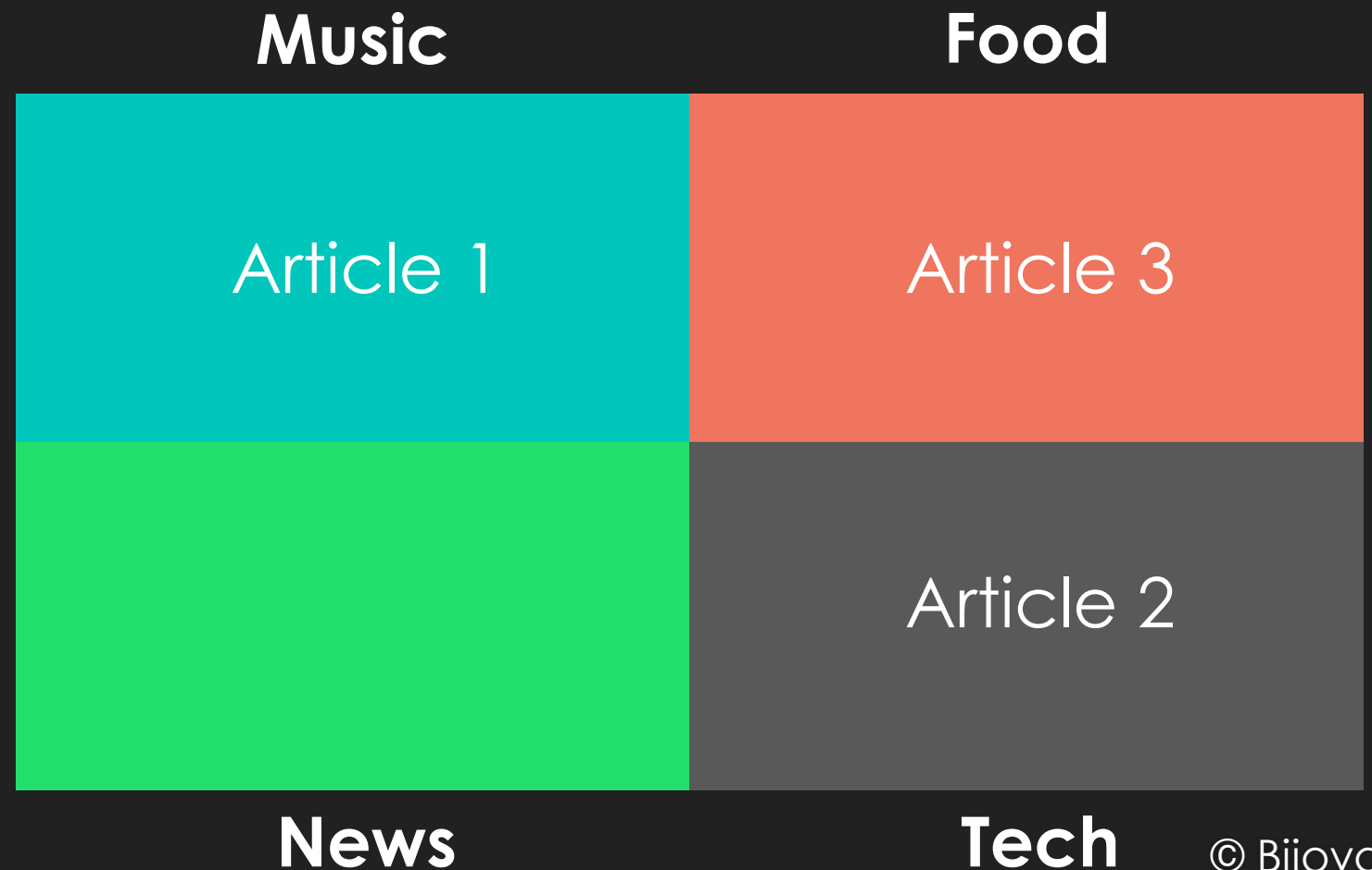
Latent Semantic Analysis

- 1.
- 2.
3. Article 3
4. Article 4
5. Article 5
6. Article 6
-



Latent Semantic Analysis

- 1.
- 2.
- 3.
4. Article 4
5. Article 5
6. Article 6
-



Latent Semantic Analysis

	Music	Food
1.	Article 1	Article 3
2.		
3.		
4.		
5. Article 5	Article 4	Article 2
6. Article 6		
.....		
	News	Tech

Latent Semantic Analysis

	Music	Food
1.	Article 1	Article 3 Article 5
2.		
3.	Article 4	Article 2
4.		
5.	Article 6	
6.		
.....		
	News	Tech

Latent Semantic Analysis

	Music	Food
1.	Article 1	Article 3 Article 5
2.		
3.	Article 4 Article 6	Article 2
4.		
5.		
6.		
.....		
	News	Tech

Latent Semantic Analysis

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
-

Music	Food
Article 1	Article 3 Article 5
Article 4 Article 6 Article 5	Article 2 Article 1
News	Tech

Latent Semantic Analysis

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
-

Music	Food
Article 1 – 85 %	Article 3 – 100% Article 5 – 73%
Article 4 – 100% Article 6 – 100% Article 5 – 27%	Article 2 – 100% Article 1 – 15%
News	Tech

Bag Of Words Model

Words/Documents	going	to	today	i	am	it	is	rain	not	outside
1	1	1	1	0	0	1	1	1	0	0
2	1	0	1	1	1	0	0	0	1	1
3	1	1	0	1	1	0	0	0	0	0

M x N matrix

M = Number of Rows/Documents

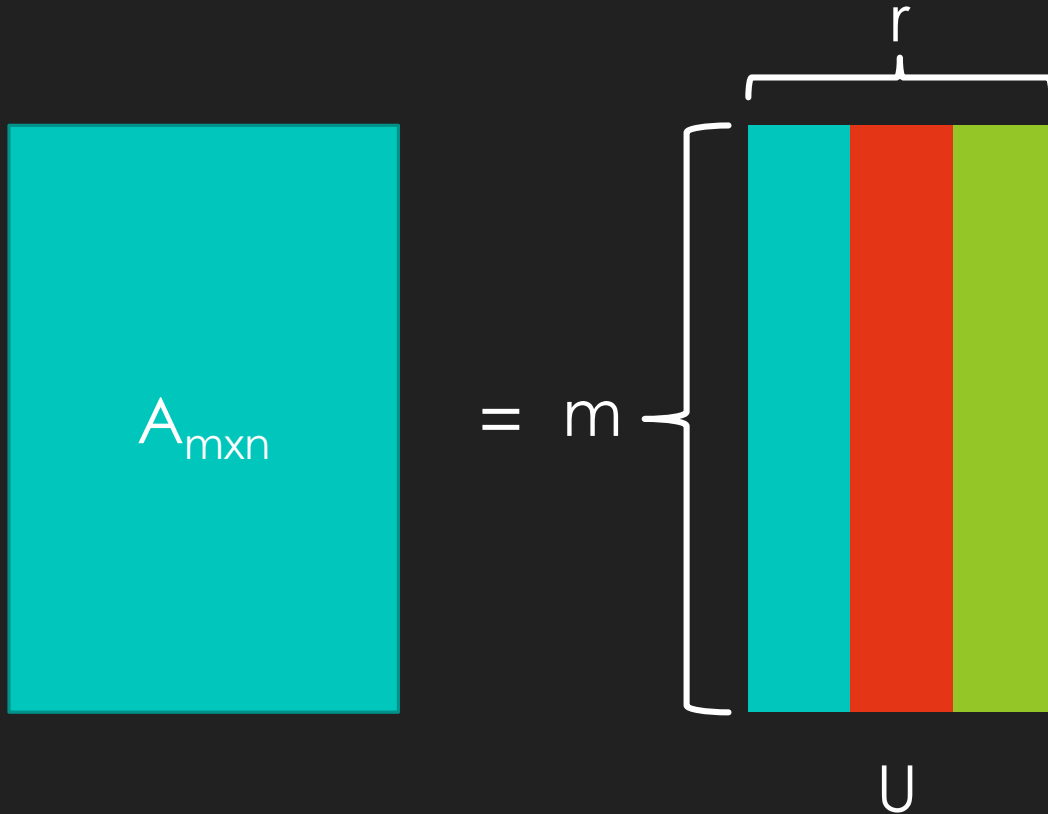
N = Number of columns/words

SVD - Definition

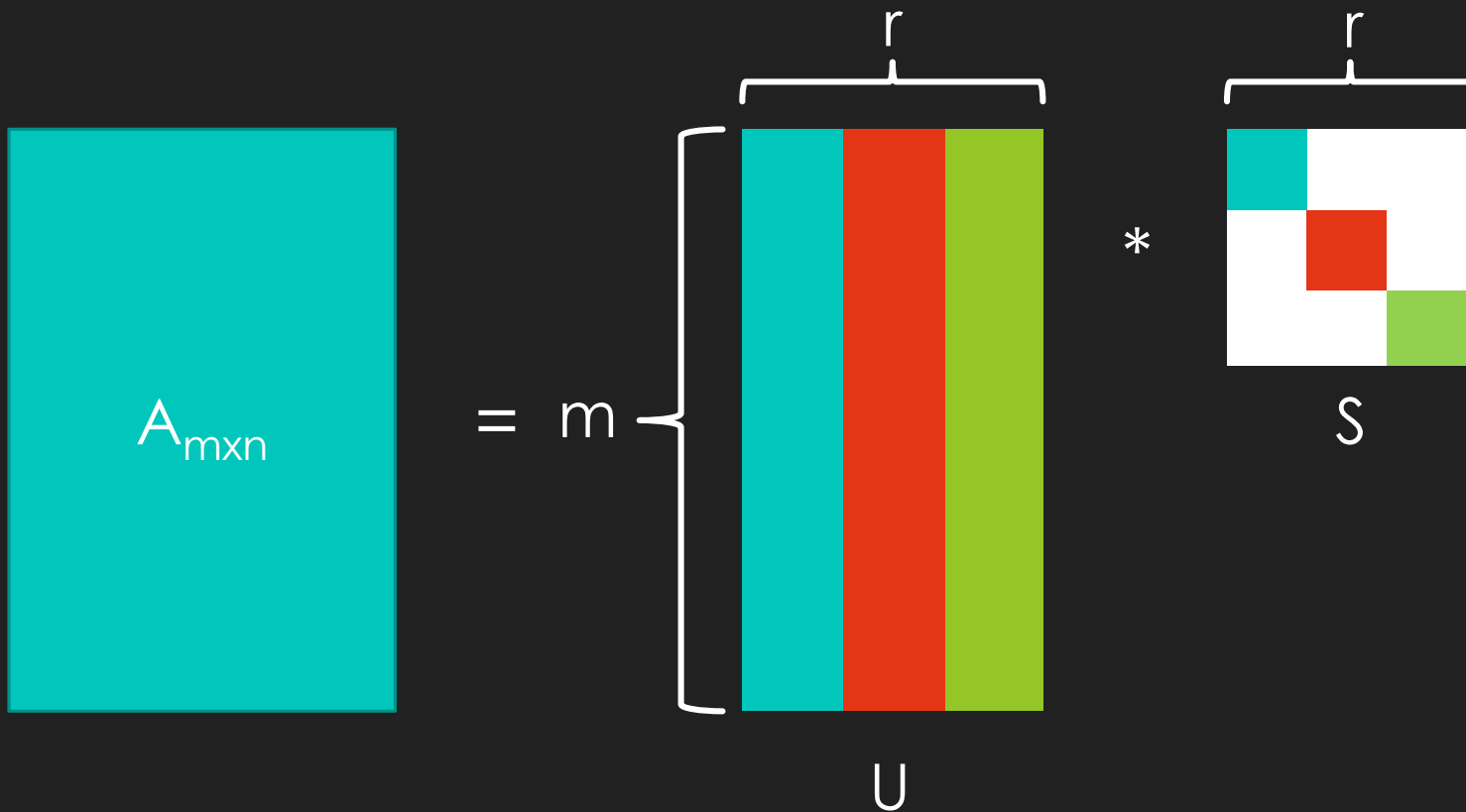
$$A_{[m \times n]} = U_{[m \times r]} * S_{[r \times r]} * (V_{[n \times r]})^T$$

- A : Input Data Matrix
 - m x n matrix (m = number of documents, n = number of words/features)
- U : Left Singular matrix
 - m x r matrix (m = number of documents, r = number of concepts)
- S : Rank Matrix
 - r x r matrix (r = rank of A)
- V : Right Singular Matrix
 - n x r matrix (n = number of words/features, r = number of concepts)

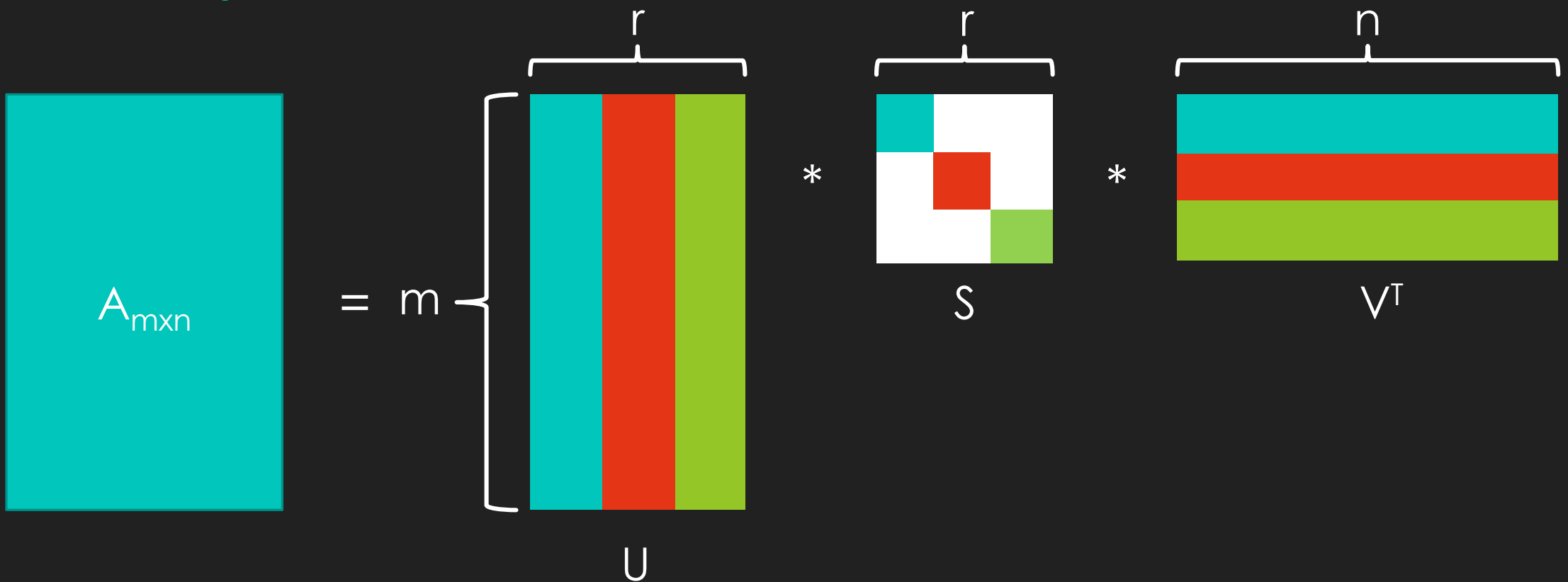
SVD – Visually Explained



SVD – Visually Explained



SVD – Visually Explained



Latent Semantic Analysis – Applications

- Article Bucketing in Websites
- Finding relations between articles/words
- Page Indexing in Search Engines