



# Procesamiento de Lenguaje Natural

## Topic Modelling

Mauricio Toledo-Acosta  
[mauricio.toledo@unison.mx](mailto:mauricio.toledo@unison.mx)

Departamento de Matemáticas  
Universidad de Sonora



## Section 1

# Introducción



## Section 2

# Ejemplos ilustrativos



# La matriz term-document

- La Revolución Francesa fue un período de grandes **cambios** políticos y sociales en **Europa**.
- La paella es un plato tradicional de España que lleva **arroz**, mariscos y verduras.
- El Imperio Romano dominó gran parte de **Europa** durante siglos, expandiéndose por toda **Europa**.
- El sushi es una comida japonesa hecha con **arroz** y **pescado** crudo, acompañado de algas.

Texto	Europa	cambios	arroz	pescado
$d_1$	1	1	0	0
$d_2$	0	0	1	0
$d_3$	2	0	0	0
$d_4$	0	0	1	1



Navigation icons: back, forward, search, and other controls. 5/20



- El gato come ratones y juega con el perro. El perro duerme al lado y come.
- El gato come pescado.
- El perro ladra fuerte y come.
- El código tiene un error.
- El programa ejecuta código.

- gato
- come
- ratones
- juega
- perro
- duerme
- lado
- pescado
- ladra
- fuerte
- código
- error
- programa
- ejecuta



## Section 3

# LSA



- **Definición:** LSA (Latent Semantic Analysis) es una técnica de procesamiento de lenguaje natural.
- **Contexto:** Usado en Topic Modelling para descubrir temas en textos.
- **Objetivo:** Identificar temas ocultos en un conjunto de documentos.





- **Matriz Término-Documento:** Representación numérica de textos.
- **SVD:** Reducción de dimensionalidad para capturar relaciones semánticas.
- **Espacio semántico latente:** Representación compacta de palabras y documentos.

## Idea



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

- ◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ↺ 🔍 ↻ 10/20



# Matrices de SVD

- **Matriz  $U$ :**
  - Vectores singulares izquierdos (relacionados con las palabras).
  - Dimensiones:  $m \times r$  ( $m$  = palabras,  $r$  = rango de  $A$ ).
- **Matriz  $\Sigma$ :**
  - Matriz diagonal con valores singulares  $(\sigma_1, \sigma_2, \dots, \sigma_r)$ .
  - Dimensiones:  $r \times r$ .
- **Matriz  $V^T$ :**
  - Vectores singulares derechos (relacionados con los documentos).
  - Dimensiones:  $r \times n$  ( $n$  = documentos).



# Singular Values and Singular Vectors

- **Singular Value:** A non-negative real number  $\sigma$  is a singular value for  $\mathbf{M}$  if there exist unit-length vectors  $\mathbf{u} \in \mathbb{R}^m$  and  $\mathbf{v} \in \mathbb{R}^n$  such that:

$$\mathbf{M}\mathbf{v} = \sigma\mathbf{u},$$

$$\mathbf{M}^*\mathbf{u} = \sigma\mathbf{v}.$$

- **Singular Vectors:**
  - $\mathbf{u}$ : Left-singular vector for  $\sigma$ .
  - $\mathbf{v}$ : Right-singular vector for  $\sigma$ .
- **Singular Value Decomposition (SVD):**

$$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$$

- $\mathbf{\Sigma}$ : Diagonal matrix of singular values.
- $\mathbf{U}$ : Columns are left-singular vectors.
- $\mathbf{V}$ : Columns are right-singular vectors.



# Espacio semántico latente

Procesamiento  
de Lenguaje  
Natural

Introducción

Ejemplos  
ilustrativos

LSA

...



# Proceso de LSA

## Procesamiento de Lenguaje Natural

### Introducción

### Ejemplos ilustrativos

### LSA

- **Preprocesamiento:** Tokenización, eliminación de stopwords, etc.
- **Matriz Término-Documento:** Creación y ponderación (TF-IDF).
- **SVD:** Aplicación y reducción de dimensionalidad.
- **Interpretación:** Identificación de temas latentes.



# Ventajas de LSA

## Procesamiento de Lenguaje Natural

### Introducción

### Ejemplos ilustrativos

### LSA

- Captura relaciones semánticas entre palabras.
- Reduce el ruido en grandes conjuntos de datos.
- Simple y fácil de implementar.



# Limitaciones de LSA

## Procesamiento de Lenguaje Natural

### Introducción

### Ejemplos ilustrativos

### LSA

- Dificultad para interpretar temas explícitamente.
- No captura relaciones no lineales entre palabras.
- Depende del preprocesamiento y parámetros.





- Recuperación de información.
- Clasificación de textos.
- Análisis de sentimientos.
- Recomendación de contenido.



# Comparación con otras técnicas

Procesamiento  
de Lenguaje  
Natural

Introducción

Ejemplos  
ilustrativos

LSA

- **LSA:** Basado en álgebra lineal (SVD).
- **LDA:** Basado en probabilidad (distribuciones de temas).
- **Uso:** LSA para simplicidad, LDA para interpretabilidad.



- **Dataset:** Textos de noticias.
- **Proceso:** Aplicación de LSA y visualización de temas.
- **Resultados:** Grupos de palabras y documentos relacionados.



# Conclusión

Procesamiento  
de Lenguaje  
Natural

Introducción

Ejemplos  
ilustrativos

LSA

- LSA es una herramienta poderosa para Topic Modelling.
- Captura relaciones semánticas en textos.
- Invitación a explorar más sobre el tema.