



Módulo VI - Aprendizaje NO supervisado

Clase 23: Single - Value Decomposition



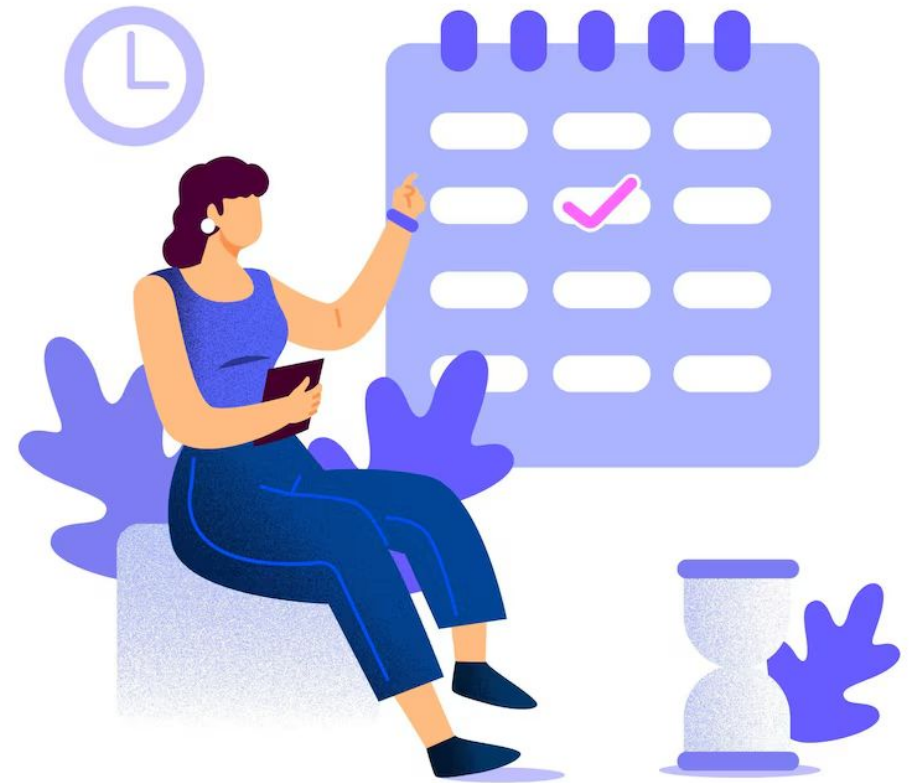


¿Ponemos a grabar el
taller?

¿QUÉ VAMOS A VER HOY?



- Single Value Decomposition





Repasemos

Eligiendo algoritmo

Tarea

Definir de forma
clara el objetivo

1

Información

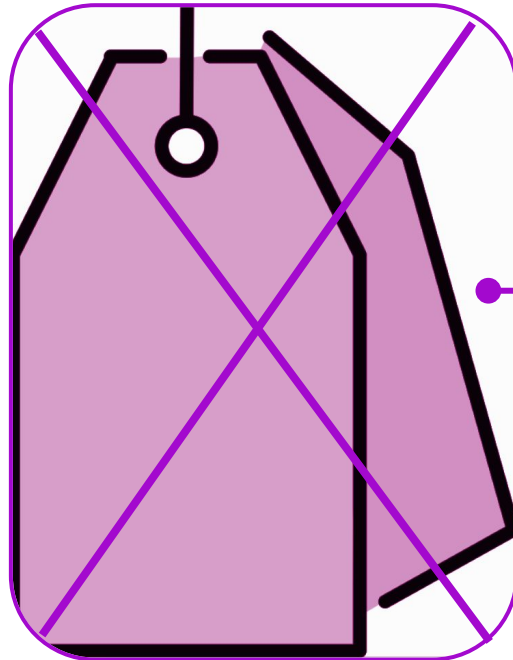
Con qué datos se cuenta
para lograr el objetivo

2

APRENDIZAJE
SUPERVISADO

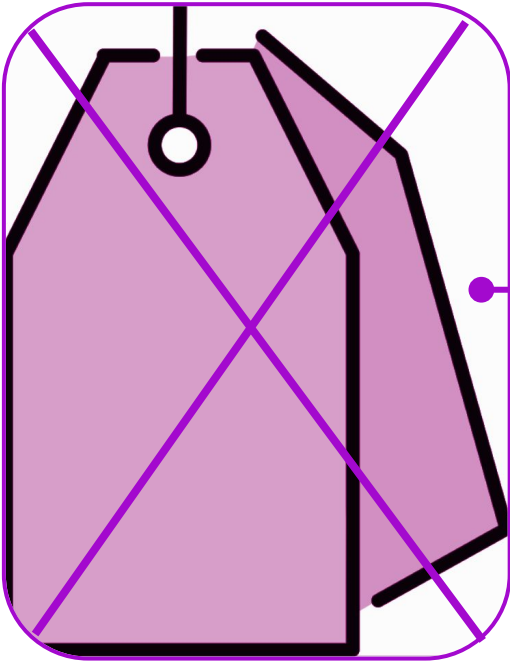
APRENDIZAJE
**NO
SUPERVISADO**

Aprendizaje No Supervisado



El algoritmo infiere patrones de un conjunto de datos que, a diferencia del aprendizaje supervisado, **no** están **etiquetados**. Puede utilizarse para descubrir la estructura subyacente de los datos

Aprendizaje No Supervisado



- Clustering
- Reducción de la Dimensionalidad



Reducción de la dimensionalidad

Reducción de la dimensionalidad

El objetivo consta de **reducir la cantidad de features** de un dataset, pero **reteniendo la mayor** cantidad de **información** posible.

Reducción de la dimensionalidad

Consiste en **reducir** la cantidad de **features** de un dataset, pero **reteniendo** la mayor cantidad de **información** posible.

- **Aplicaciones**
 - Reducir la complejidad del input en un modelo de regresión o clasificación
 - Visualización
 - Detectar features relevantes en datasets
- **Algoritmos**
 - **Principal Component Analysis**
 - Multidimensional Scaling
 - t-SNE: t-distributed Stochastic Neighbor Embedding
 - LDA: Linear Discriminant Analysis

Principal Component Analysis

Técnica utilizada para describir un conjunto de datos en términos de nuevas variables o componentes no correlacionados.

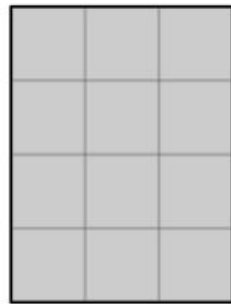
- El algoritmo encuentra nuevos componentes que describen los datos
- Los componentes se ordenan por la cantidad de varianza original que describen, reduciendo la dimensionalidad del conjunto de datos.
- Utiliza **Single Vector Decomposition (SVD)**



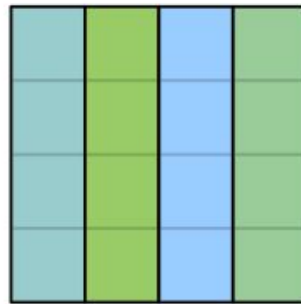
Single Value Decomposition

Single Vector Decomposition

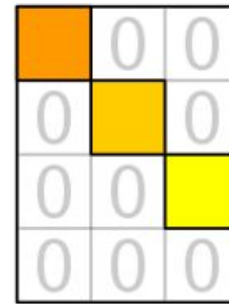
Es un **método de álgebra lineal** que nos permite representar cualquier matriz en términos de la multiplicación de otras 3 matrices.



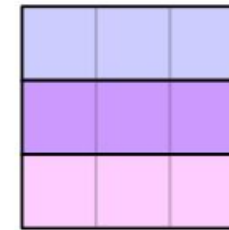
M
 $m \times n$



U
 $m \times m$



Σ
 $m \times n$



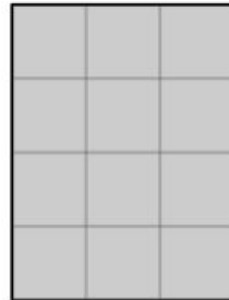
V*
 $n \times n$

$$\mathbf{M}_{m \times n} = \mathbf{U}_{m \times m} \mathbf{\Sigma}_{m \times n} \mathbf{V}^*_{n \times n}$$

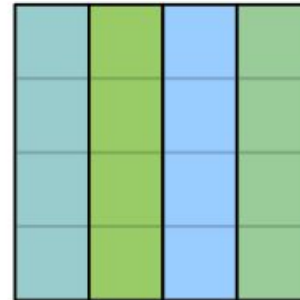
Single Vector Decomposition

Toda matriz M
se puede
escribir como :

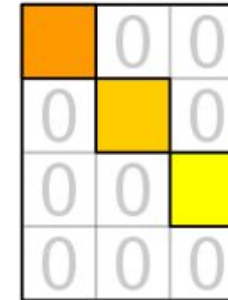
Matriz de Datos
(m instancias,
 n features)



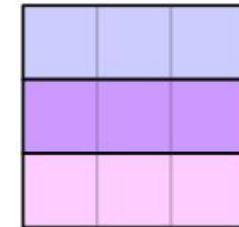
Matriz de
vectores
singulares por
izquierda



Matriz de los
valores
singulares



Matriz de vectores
singulares por
derecha



$$\begin{matrix} \mathbf{M} \\ m \times n \end{matrix} = \begin{matrix} \mathbf{U} \\ m \times m \end{matrix} \begin{matrix} \mathbf{\Sigma} \\ m \times n \end{matrix} \begin{matrix} \mathbf{V}^* \\ n \times n \end{matrix}$$

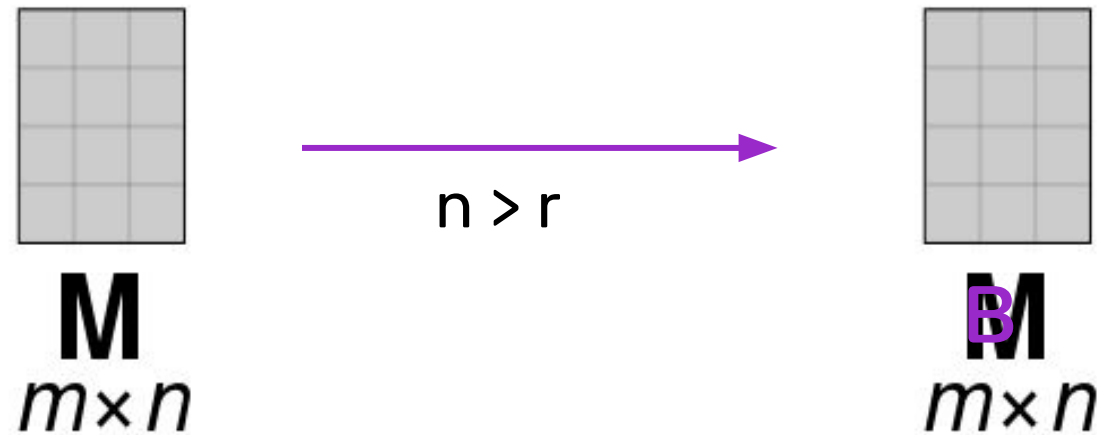
Matriz Unitaria

Matriz
Diagonal

Matriz Unitaria

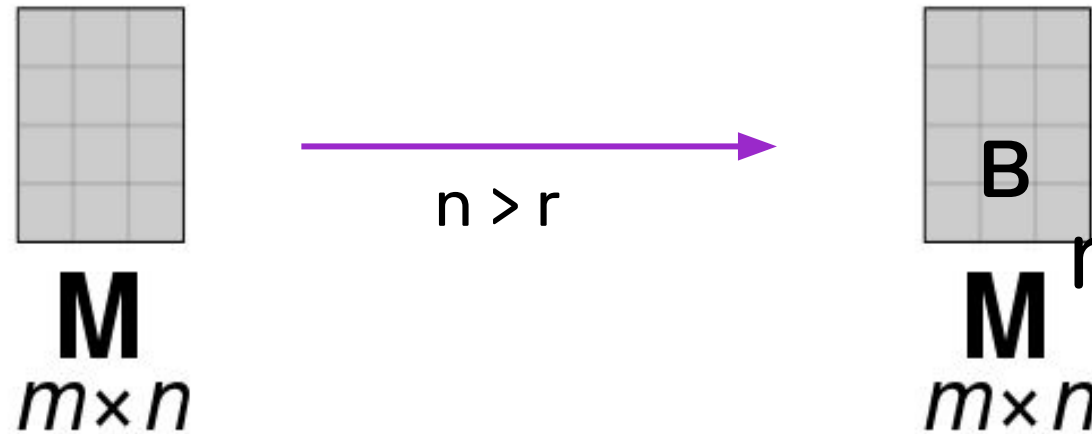
Single Vector Decomposition

Es parte central de varios algoritmos numéricos y es fundamental para reducir apropiadamente la matriz M : pasar de tener muchos features a tener menos, pero relevantes y explicativos.



Single Vector Decomposition

El objetivo es construir una nueva matriz B que reemplace a M , que tenga menos features.



Single Vector Decomposition

Para eso debemos tomar solo los valores principales (elementos en la diagonal de Sigma) de valor más grande y construir una matriz B que sea una **reducción** de X **relevante**.

| | | |
|--|--|--|
| | | |
| | | |
| | | |
| | | |

$$\Sigma_{m \times n}$$

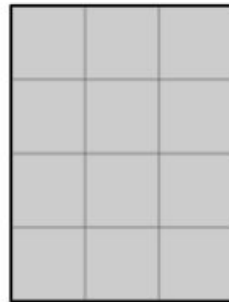
$$n > r$$

| | |
|--|--|
| | |
| | |
| | |
| | |

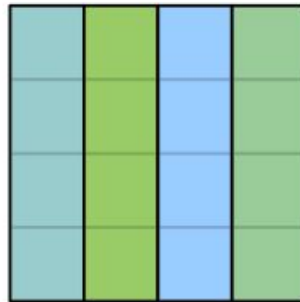
$$\Sigma_r_{m \times r}$$

Single Vector Decomposition

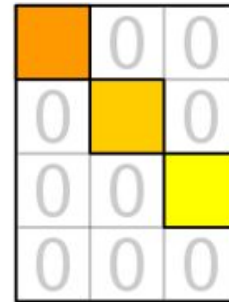
Matriz completa: es la M original, tiene toda la información.



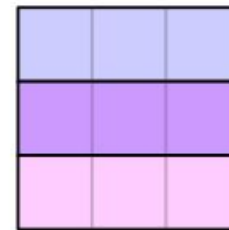
M
 $m \times n$



U
 $m \times m$



Σ
 $m \times n$



V*
 $n \times n$

Single Vector Decomposition

Matriz truncada: En esta matriz perdemos información. Pero al tomar un valor de r apropiado, la matriz truncada es muy parecida a M . Construimos una matriz B con menos features que M , pero relevante para el problema. Esta matriz es la que se utiliza para trabajar y representar los datos.

$$\begin{array}{c}
 \begin{array}{cccc}
 \begin{array}{|c|c|c|c|} \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \end{array} & = & \begin{array}{|c|c|c|c|} \hline \text{teal} & \text{green} & \text{blue} & \text{green} \\ \hline \text{teal} & \text{green} & \text{blue} & \text{green} \\ \hline \text{teal} & \text{green} & \text{blue} & \text{green} \\ \hline \end{array} & \begin{array}{|c|c|c|} \hline \text{orange} & 0 & 0 \\ \hline 0 & \text{yellow} & 0 \\ \hline 0 & 0 & \text{yellow} \\ \hline 0 & 0 & 0 \\ \hline \end{array} & \begin{array}{|c|c|c|} \hline \text{light blue} & \text{light blue} & \text{light blue} \\ \hline \text{purple} & \text{purple} & \text{purple} \\ \hline \text{pink} & \text{pink} & \text{pink} \\ \hline \end{array} \\
 \mathbf{M}_{m \times n} & = & \mathbf{U}_{m \times m} & \mathbf{\Sigma}_{m \times n} & \mathbf{V}^*_{n \times n}
 \end{array}
 \end{array}$$

$\downarrow \quad n > r$

$$\begin{array}{cccc}
 \begin{array}{|c|c|c|c|} \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \square & \square & \square & \square \\ \hline \end{array} & = & \begin{array}{|c|c|c|c|} \hline \text{teal} & \text{green} & \text{blue} & \text{green} \\ \hline \text{teal} & \text{green} & \text{blue} & \text{green} \\ \hline \text{teal} & \text{green} & \text{blue} & \text{green} \\ \hline \end{array} & \begin{array}{|c|c|} \hline \text{orange} & 0 \\ \hline 0 & \text{yellow} \\ \hline 0 & 0 \\ \hline 0 & 0 \\ \hline \end{array} & \begin{array}{|c|c|} \hline \text{light blue} & \text{light blue} \\ \hline \text{purple} & \text{purple} \\ \hline \end{array} \\
 \hat{\mathbf{M}}_{m \times n} & = & \mathbf{U}_{m \times m} & \mathbf{\Sigma}_r_{m \times r} & \mathbf{V}_r^*_{r \times n}
 \end{array}$$

Single Vector Decomposition

La matriz
original y la
truncada son
parecidas

$$\begin{matrix} \text{[Grid]} & \text{[Grid]} & \text{[Grid]} & \text{[Grid]} \\ \mathbf{M} & = & \mathbf{U} & \mathbf{\Sigma} & \mathbf{V}^* \\ m \times n & & m \times m & m \times n & n \times n \end{matrix}$$

$n > r$

$$\begin{matrix} \text{[Grid]} & \text{[Grid]} & \text{[Grid]} & \text{[Grid]} \\ \tilde{\mathbf{M}} & = & \mathbf{U} & \mathbf{\Sigma}_r & \mathbf{V}_r^* \\ m \times n & & m \times m & m \times r & r \times n \end{matrix}$$

$\mathbf{B}_{m \times r}$

Single Vector Decomposition

¿Cómo elegir el **valor de r**?

1) **Distancia entre M y \tilde{M}**

$$\|M - \tilde{M}\|_F = \sqrt{\sum_{ij} (M_{ij} - \tilde{M}_{ij})^2}$$

El método de Single Value Decomposition GARANTIZA que elegimos los mejores r vectores o combinaciones de features para minimizar esta norma.

Single Vector Decomposition

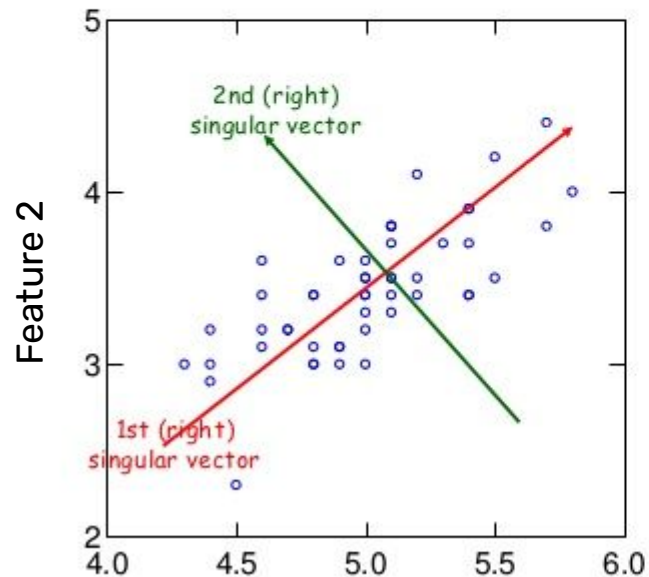
¿Cómo elegir el **valor de r** ?

2) **Criterio sobre el peso relativo** de los valores singulares seleccionados respecto a la suma de todos.

Esta técnica es más costosa ya que se deben calcular todos los valores singulares.

Single Vector Decomposition

Representación Gráfica



- Espacio original: 2 features (2 coordenadas). Esto define la posición de todas las instancias del dataset.
- SVD calcula nuevos vectores, el 1er y 2do vector singular. Utilizando estos valores como coordenadas, se puede definir la posición de cada punto.



Principal Component Analysis

Principal Component Analysis

PCA utiliza SVD. Por lo cual, aplicar PCA y SVD es muy parecido. Pero existe una diferencia:

PCA = Centrar datos + SVD truncado

Principal Component Analysis

**Componentes
Principales**



**Direcciones de
máxima varianza**

El primer componente principal se ubica en la dirección donde los datos presentan varianza máxima.

El segundo componente principal se ubica en la segunda dirección en términos de varianza. Y así sucesivamente.

PCA vs. SVD

Parametros y Caracteristicas

| PCA | SVD |
|-------------------------|---------------------------------|
| Número de componentes | Rango r |
| Componentes principales | Vectores singulares por derecha |
| Autovalores | Valores singulares |
| Maximiza Varianza | Minimiza Distancia |



Descanso

Nos vemos en 10 minutos

Sección Práctica

TRABAJAMOS EN SALAS

SVD

Trabajaremos con la
Notebook 22

En los grupos establecidos,
ejercitamos como se
implementa el SVD



45 minutos de actividad



Repasamos dudas

**Trabajamos con la
Notebook 22**

Revisamos los conceptos y el
código trabajados en la
notebook 22



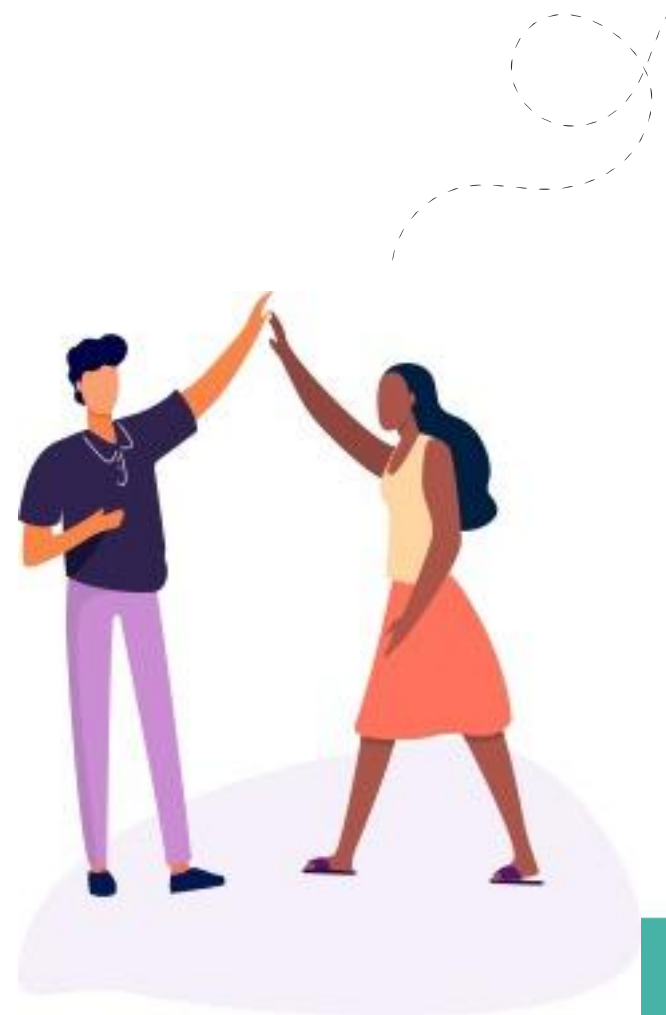
Preparando la cuarta pre-entrega

En la clase 33 deberán realizar la cuarta pre-entrega

La misma incluirá los desafíos vistos, en relación a los modelos de aprendizaje no supervisado.

- Deberán desarrollar un modelo de clustering, desde el análisis exploratorio hasta la mejora
- Elegir algún algoritmo y aplicarlo a su proyecto.
- Será importante que hagan una revisión y tomen decisiones, para aplicar lo visto con un criterio que deberán explicitar.

Presentarán lo trabajado entregando el link a su Github en el foro del aula virtual





¿Alguna consulta?

FUNDACIÓN
YPF

¡Muchas gracias!

