

ACÀMICA

¡Bienvenidas/os a Data Science!



Agenda

¿Cómo anduvieron?

Repaso: Aprendizaje No Supervisado

Explicación: SVD

Break

Hands-On

Cierre



¿Cómo anduvieron?



Invitado

ESTEBAN TOSO

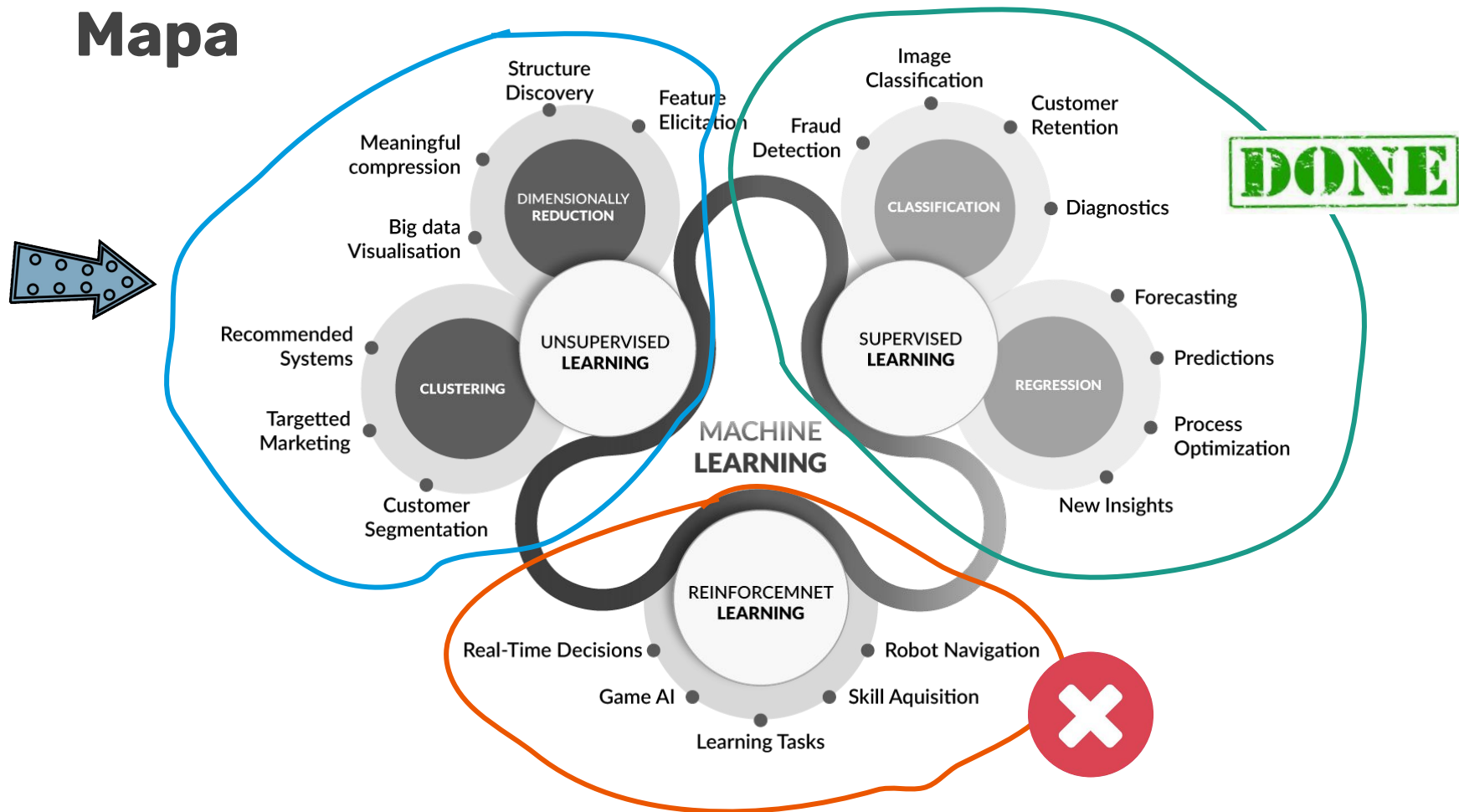
Data Scientist



Repaso: Aprendizaje no supervisado



Mapa



Solo datos

Llamamos **Aprendizaje No Supervisado** a los métodos para trabajar con datos (instancias) que no tienen asociados una etiqueta (una clase o un valor).

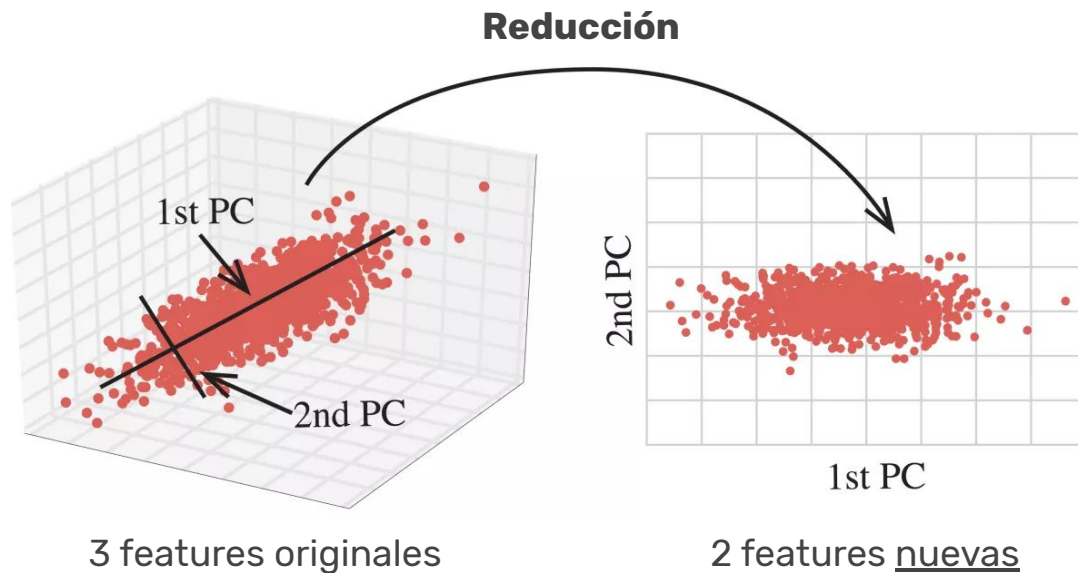
Los objetivos principales en Aprendizaje No Supervisado son:

- **Clustering**
- **Reducción de dimensionalidad**

- Clustering
- Reducción de dimensionalidad

Aprendizaje No Supervisado • Reducción de la dimensionalidad

Buscamos reducir la cantidad de features de un dataset, pero reteniendo la mayor cantidad de “información” posible.



Aprendizaje No Supervisado • Reducción de la dimensionalidad

¿Para qué sirve?

Reducir la cantidad de features en un dataset puede servir para:

- Reducir el input en un modelo de regresión o clasificación
- Compresión de archivos
- Visualización
- Detectar features relevantes en datasets
- Muchísimas mas cosas

Aprendizaje No Supervisado • Reducción de la dimensionalidad

¿Para qué sirve?

Reducir la cantidad de features en un dataset puede servir para:

- Reducir el input en un modelo de regresión o clasificación
- Compresión de archivos
- Visualización
- Detectar features relevantes en datasets
- Muchísimas mas cosas

¿Cómo se hace?

Algunos de los métodos de reducción de dimensionalidad son:

- PCA: Principal Component Analysis (usa SVD)
- MDS: Multidimensional scaling
- t-SNE: t-distributed Stochastic Neighbor Embedding
- Auto-Encoders (Se hace con Redes Neuronales)
- LDA: Linear Discriminant Analysis (si hay etiquetas de clases)

Aprendizaje No Supervisado • Reducción de la dimensionalidad

¿Para qué sirve?

Reducir la cantidad de features en un dataset puede servir para:

- Reducir el input en un modelo de regresión o clasificación
- Compresión de archivos
- Visualización
- Detectar features relevantes en datasets
- Muchísimas mas cosas

¿Cómo se hace?

Algunos de los métodos de reducción de dimensionalidad son:

- **PCA: Principal Component Analysis (usa SVD)**
- MDS: Multidimensional scaling
- t-SNE: t-distributed Stochastic Neighbor Embedding
- Auto-Encoders (Se hace con Redes Neuronales)
- LDA: Linear Discriminant Analysis (si hay etiquetas de clases)

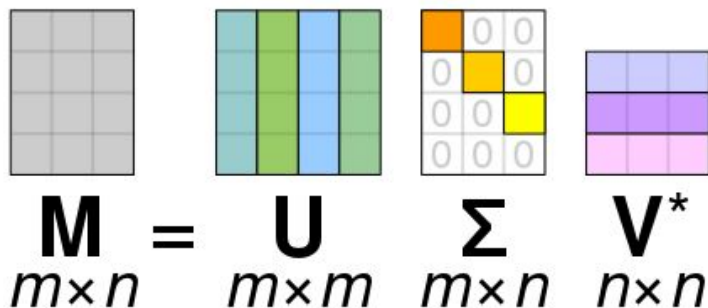
Aprendizaje No Supervisado

SVD (Singular Value Decomposition)



SVD • Definición

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



The diagram illustrates the SVD decomposition of a matrix M into three matrices U , Σ , and V^* . Each matrix is represented by a grid of colored squares:

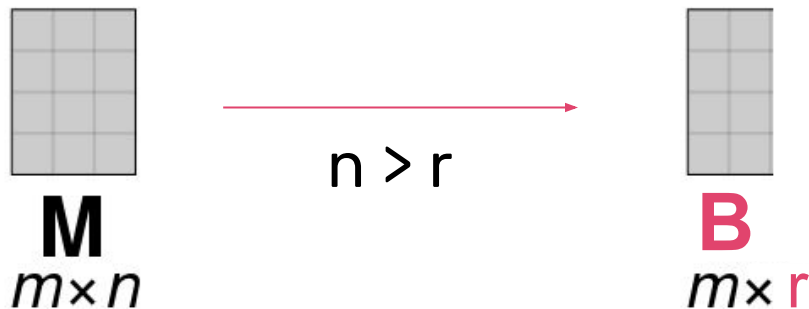
- M is a 4×4 matrix represented by a gray grid.
- U is a 4×4 matrix represented by a grid with four vertical columns of different colors: light blue, green, light blue, and green.
- Σ is a 4×4 matrix represented by a grid with three non-zero elements on the diagonal: orange, yellow, and yellow. The other elements are white.
- V^* is a 4×4 matrix represented by a grid with four horizontal rows of different colors: light blue, purple, purple, and light blue.

The equation is shown as:

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

SVD • ¿Para qué sirve?

Para MUCHAS COSAS. Es parte del corazón de muchos algoritmos numéricos (solución sis. lineal, pseudoinversa, etc.). En este contexto vamos a usarlo para “reducir” adecuadamente la matriz M (pasar de tener muchos features a tener menos, pero que sean buenos).



SVD • Álgebra

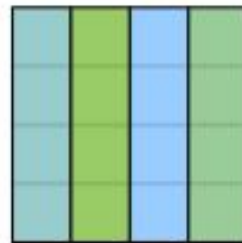
Se puede demostrar que a toda matriz M la podemos escribir como :

Matriz de Datos
(m instancias,
 n features)



$$\mathbf{M}_{m \times n}$$

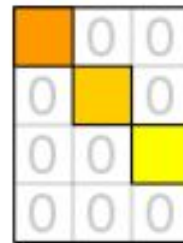
Matriz de
vectores
singulares por
izquierda



$$\mathbf{U}_{m \times m}$$

Matriz
Unitaria

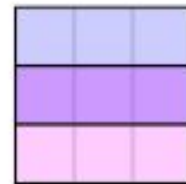
Matriz de los
valores
singulares



$$\mathbf{\Sigma}_{m \times n}$$

Matriz
Diagonal

Matriz de
vectores
singulares por
derecha



$$\mathbf{V}^*_{n \times n}$$

Matriz
Unitaria

¿Y qué tiene que ver esto con todo lo que venimos hablando?



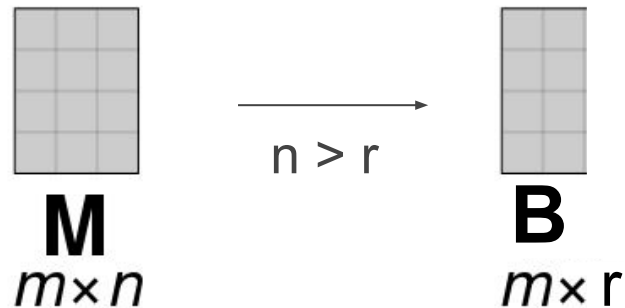
Aprendizaje No Supervisado

SVD truncado



SVD truncado

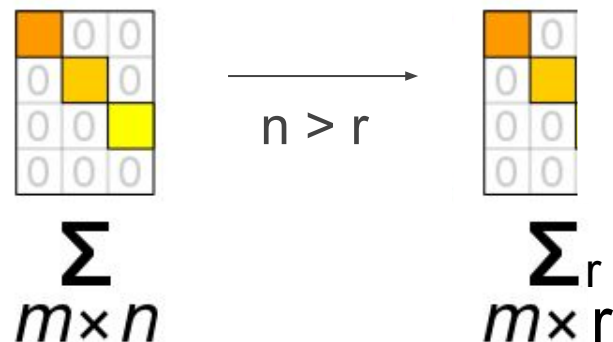
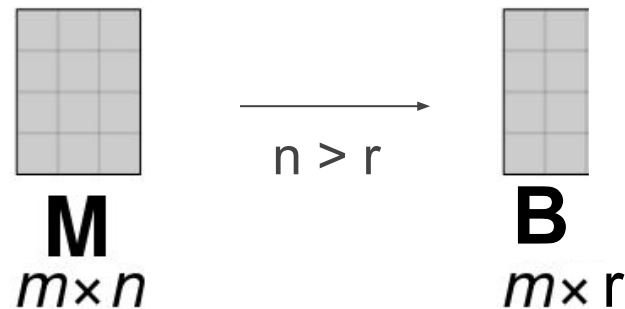
Objetivo: queremos una nueva matriz B que reemplace a M , que tenga menos columnas (menos features).



SVD truncado

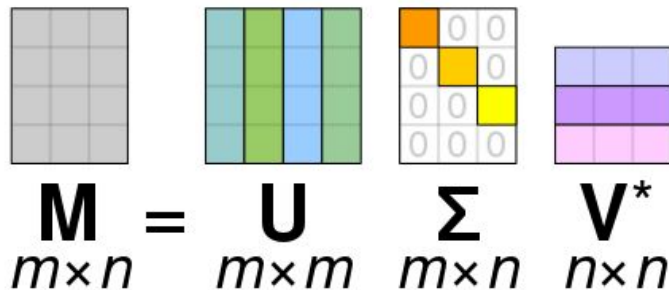
Objetivo: queremos una nueva matriz B que reemplace a M , que tenga menos columnas (menos features).

Idea de cómo lograrlo: si tomamos solo los r valores principales (elementos en la diagonal de Sigma) de valor más grande, podemos construir una matriz B que sea una “buena” reducción de M .



SVD truncado

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



The diagram illustrates the SVD decomposition of a matrix M into three components: U , Σ , and V^* . The equation is shown as $M = U \Sigma V^*$. Above each matrix symbol is a grid representing its structure:

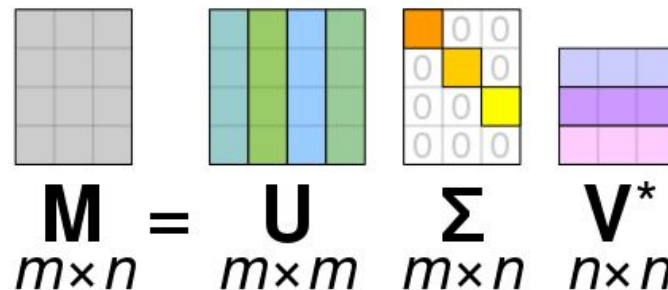
- M is a 4x4 grid of gray squares.
- U is a 4x4 grid with columns colored light blue, green, light blue, and green.
- Σ is a 4x4 grid with a diagonal of colored squares (orange, yellow, yellow, light blue) and zeros elsewhere.
- V^* is a 4x4 grid with rows colored light blue, purple, purple, and pink.

Below each grid is its label and dimensions:

- M is labeled $m \times n$.
- U is labeled $m \times m$.
- Σ is labeled $m \times n$.
- V^* is labeled $n \times n$.

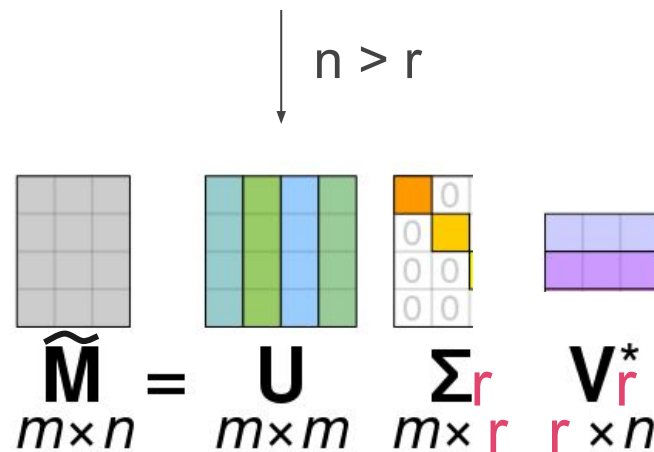
SVD truncado

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


$$\begin{matrix} \text{4x4} & & \text{4x4} & & \text{4x4} & & \text{4x4} \\ \mathbf{M} & = & \mathbf{U} & \mathbf{\Sigma} & \mathbf{V}^* \\ m \times n & & m \times m & & m \times n & & n \times n \end{matrix}$$


Matriz truncada: perdimos información. Pero si tomamos un valor de r adecuado, \hat{M} moño es muy parecida a M . Construimos una matriz B mas chica que M , esta es la matriz con la que vamos a trabajar.

$\downarrow n > r$


$$\begin{matrix} \text{4x4} & & \text{4x4} & & \text{4x4} & & \text{3x4} \\ \hat{\mathbf{M}} & = & \mathbf{U} & \mathbf{\Sigma}_r & \mathbf{V}_r^* \\ m \times n & & m \times m & & m \times r & & r \times n \end{matrix}$$

SVD truncado


Parecidas



The diagram illustrates the full SVD decomposition of a matrix M of size $m \times n$. Matrix M is represented by a 4x4 grid of gray squares. It is equal to the product of three matrices: U (size $m \times m$), Σ (size $m \times n$), and V^* (size $n \times n$). Matrix U is shown as a 4x4 grid with four colored columns: teal, green, blue, and green. Matrix Σ is a 4x4 grid with a diagonal of four colored squares (orange, yellow, yellow, yellow) and zeros elsewhere. Matrix V^* is a 4x4 grid with four colored rows: light blue, purple, purple, and pink.

$$\begin{matrix} \text{4x4 grid} \\ \mathbf{M} \\ m \times n \end{matrix} = \begin{matrix} \text{4x4 grid} \\ \mathbf{U} \\ m \times m \end{matrix} \begin{matrix} \text{4x4 grid} \\ \mathbf{\Sigma} \\ m \times n \end{matrix} \begin{matrix} \text{4x4 grid} \\ \mathbf{V}^* \\ n \times n \end{matrix}$$

$n > r$




The diagram illustrates the truncated SVD decomposition of a matrix M of size $m \times n$. Matrix M is represented by a 4x4 grid of gray squares. It is equal to the product of three matrices: U (size $m \times m$), Σ_r (size $m \times r$), and V_r^* (size $r \times n$). Matrix U is shown as a 4x4 grid with four colored columns: teal, green, blue, and green. Matrix Σ_r is a 4x3 grid with a diagonal of three colored squares (orange, yellow, yellow) and zeros elsewhere. Matrix V_r^* is a 3x4 grid with three colored rows: light blue, purple, and purple.

$$\begin{matrix} \text{4x4 grid} \\ \widetilde{\mathbf{M}} \\ m \times n \end{matrix} = \begin{matrix} \text{4x4 grid} \\ \mathbf{U} \\ m \times m \end{matrix} \begin{matrix} \text{4x3 grid} \\ \mathbf{\Sigma}_r \\ m \times r \end{matrix} \begin{matrix} \text{3x4 grid} \\ \mathbf{V}_r^* \\ r \times n \end{matrix}$$

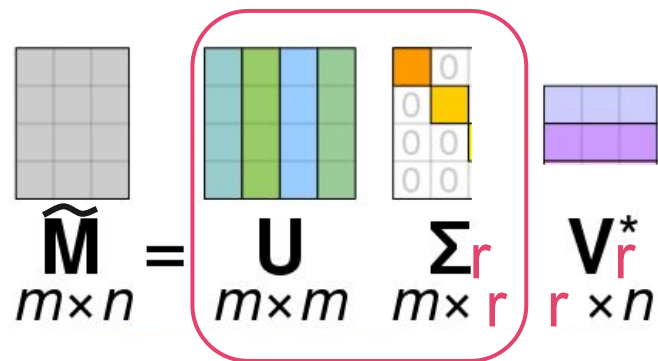
SVD truncado

Parecidas



$$\begin{matrix} \text{4x4} & & \text{4x4} & & \text{4x4} & & \text{4x4} \\ \mathbf{M} & = & \mathbf{U} & & \mathbf{\Sigma} & & \mathbf{V}^* \\ m \times n & & m \times m & & m \times n & & n \times n \end{matrix}$$

$n > r$



$$\begin{matrix} \text{4x4} & & \text{4x4} & & \text{4x4} & & \text{4x4} \\ \widetilde{\mathbf{M}} & = & \mathbf{U} & & \mathbf{\Sigma}_r & & \mathbf{V}_r^* \\ m \times n & & m \times m & & m \times r & & r \times n \end{matrix}$$

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

SVD truncado

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

B
 $m \times r$



Matriz con la que vamos a trabajar
en vez de M, tiene la misma
información que M moño.

Esta matriz funciona
como un diccionario
para pasar del mundo
de B al mundo de M.



Aprendizaje No Supervisado

Ejemplo conceptual SVD



SVD • Ejemplo 1

Tenemos un dataset de 7 usuarios y 5 películas. Cada usuario puso un valor entre 0 a 5 a cada película.

SVD • Ejemplo 1

Tenemos un dataset de 7 usuarios y 5 películas. Cada usuario puso un valor entre 0 a 5 a cada película.

$$\mathbf{M}_{7 \times 5} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix}$$

SVD • Ejemplo 1

Tenemos un dataset de 7 usuarios y 5 películas. Cada usuario puso un valor entre 0 a 5 a cada película.

$$\mathbf{M}_{7 \times 5} = \begin{array}{c} \begin{array}{c} \text{Matrix} \\ \text{Alien} \\ \text{Avatar} \\ \text{Titanic} \\ \text{Amelie} \end{array} \left[\begin{array}{ccccc} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{array} \right] \begin{array}{c} \text{Usuario 1} \\ \text{Usuario 2} \\ \dots \end{array} \left. \begin{array}{c} \text{Ciencia Ficción} \\ \text{Románticas} \end{array} \right\}$$

SVD • Ejemplo 1

Buscamos una matriz B más con menos columnas que M. Proponemos usar un valor de $r=2$ es decir que B será de 7×2 . Veamos como quedaría:

The diagram illustrates the SVD decomposition of a matrix M into three components: U , Σ , and V^* . The matrix M is shown as a 7×7 grid. The matrix U is a 7×7 grid with columns colored green, blue, and green. The matrix Σ is a 7×7 grid with diagonal elements colored orange, yellow, and yellow. The matrix V^* is a 7×7 grid with rows colored purple, purple, and purple. The matrix M is equal to the product of U , Σ , and V^* .

An arrow points to the right, where the matrix \hat{M} is shown as a 7×7 grid. The matrix \hat{M} is equal to the product of U , Σ_r , and V_r^* . The matrix U is a 7×7 grid with columns colored green, blue, and green. The matrix Σ_r is a 7×2 grid with diagonal elements colored orange and yellow. The matrix V_r^* is a 2×7 grid with rows colored purple and purple. The matrix \hat{M} is equal to the product of U , Σ_r , and V_r^* .

The matrix B is shown as a 7×2 grid, which is the product of U , Σ_r , and V_r^* .

Esta vez usaremos solo los 2 valores singulares más grandes de Sigma.

SVD • Ejemplo 1

$$U_r = \begin{bmatrix} 0.13 & 0.02 \\ 0.41 & 0.07 \\ 0.55 & 0.09 \\ 0.68 & 0.11 \\ 0.15 & -0.59 \\ 0.07 & -0.73 \\ 0.07 & -0.29 \end{bmatrix} \quad \Sigma_r = \begin{bmatrix} 12.4 & 0 \\ 0 & 9.5 \end{bmatrix} \quad V_r^* = \begin{bmatrix} 0.56 & 0.59 & 0.56 & 0.09 & 0.09 \\ 0.12 & -0.02 & -0.12 & -0.69 & -0.69 \end{bmatrix}$$

SVD • Ejemplo 1

$$U_r = \begin{bmatrix} \boxed{0.13} & \boxed{0.02} \\ \boxed{0.41} & \boxed{0.07} \\ \boxed{0.55} & \boxed{0.09} \\ \boxed{0.68} & \boxed{0.11} \\ \boxed{0.15} & \boxed{-0.59} \\ \boxed{0.07} & \boxed{-0.73} \\ \boxed{0.07} & \boxed{-0.29} \end{bmatrix}$$

$$\Sigma_r = \begin{bmatrix} 12.4 & 0 \\ 0 & 9.5 \end{bmatrix}$$

$$V_r^* = \begin{array}{c} \begin{array}{cc} \text{Ciencia Ficción} & \text{Románticas} \end{array} \\ \begin{array}{cc} \overbrace{\hspace{1.5cm}} & \overbrace{\hspace{1.5cm}} \\ \begin{array}{cc} \text{Matrix} & \text{Alien} & \text{Avatar} & \text{Titanic} & \text{Amelie} \end{array} \end{array} \\ \begin{bmatrix} 0.56 & 0.59 & 0.56 & 0.09 & 0.09 \\ 0.12 & -0.02 & -0.12 & -0.69 & -0.69 \end{bmatrix} \end{array}$$

Pesos: X Z

SVD • Ejemplo 1

$$U_r = \begin{bmatrix} \boxed{0.13} & \boxed{0.02} \\ \boxed{0.41} & \boxed{0.07} \\ \boxed{0.55} & \boxed{0.09} \\ \boxed{0.68} & \boxed{0.11} \\ \boxed{0.15} & \boxed{-0.59} \\ \boxed{0.07} & \boxed{-0.73} \\ \boxed{0.07} & \boxed{-0.29} \end{bmatrix} \quad \Sigma_r = \begin{bmatrix} 12.4 & 0 \\ 0 & 9.5 \end{bmatrix} \quad V_r^* = \begin{bmatrix} 0.56 & 0.59 & 0.56 & 0.09 & 0.09 \\ 0.12 & -0.02 & -0.12 & -0.69 & -0.69 \end{bmatrix}$$

Ciencia Ficción			Románticas	
Matrix	Alien	Avatar	Titanic	Amelie

Pesos: X Z

- Ahora cada Usuario estará identificado por dos features X y Z. Notemos que los primeros 4 usuarios tienen un valor alto de X y bajo de Z. En los otros 3, se da al revés.
- Los features encontrados corresponden a los géneros.

SVD • Ejemplo 1

Pasamos de identificar a cada usuario con un puntaje al género de las películas en lugar de a las películas en sí, pasamos de 5 a 2 features.

Cuanta información perdemos por usar B en lugar de M?

The diagram illustrates the SVD decomposition of a matrix \hat{M} into three matrices: U , Σ_r , and V_r^* . Above each matrix is a grid representing its structure and dimensions:

- \hat{M} is represented by a 4x4 grid of gray squares, with dimensions $m \times n$ below it.
- U is represented by a 4x4 grid of colored squares (teal, green, blue, green), with dimensions $m \times m$ below it.
- Σ_r is represented by a 4x2 grid of squares, with dimensions $m \times r$ below it. The top-left square is orange, the top-right is white with a '0', the second row is white with '0' and yellow, and the bottom two rows are white with '0'.
- V_r^* is represented by a 2x4 grid of colored squares (light blue, light blue, purple, purple), with dimensions $r \times n$ below it.

The equation is shown as:

$$\hat{M}_{m \times n} = U_{m \times m} \Sigma_r_{m \times r} V_r^*_{r \times n}$$


SVD • Ejemplo 1

Pasamos de identificar a cada usuario con un puntaje al género de las películas en lugar de a las películas en sí, pasamos de 5 a 2 features.

Cuanta información perdemos por usar B en lugar de M?

$$M = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 2 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 1 & 0 & 2 & 2 \end{bmatrix}$$

$$\tilde{M} = \begin{bmatrix} 0.92 & 0.95 & 0.92 & 0.01 & 0.01 \\ 2.91 & 3.01 & 2.91 & -0.01 & -0.01 \\ 3.90 & 4.04 & 3.90 & 0.01 & 0.01 \\ 4.82 & 5.00 & 4.82 & 0.03 & 0.03 \\ 0.70 & 0.53 & 0.70 & 4.11 & 4.11 \\ -0.69 & 1.34 & -0.69 & 4.78 & 4.78 \\ 0.32 & 0.23 & 0.32 & 2.01 & 2.01 \end{bmatrix}$$



The diagram shows the SVD decomposition of matrix M into three matrices: U, Σ_r, and V_r^*. Matrix M is represented by a 7x5 grid of gray squares. Matrix U is represented by a 7x5 grid of colored squares (teal, green, blue, green, teal, teal, teal). Matrix Σ_r is represented by a 7x5 grid of colored squares (orange, yellow, yellow, yellow, yellow, yellow, yellow). Matrix V_r^* is represented by a 5x5 grid of colored squares (purple, purple, purple, purple, purple). Below the matrices, the equation is written as:

$$\tilde{M}_{m \times n} = U_{m \times m} \Sigma_r_{m \times r} V_r^*_{r \times n}$$

Estamos muy cerca!!

Hiperparámetro r

SVD • Hiperparámetro r

¿Cómo podríamos elegir el valor de r ?

SVD • Hiperparámetro r

¿Cómo podríamos elegir el **valor de r**?

Una posibilidad es **mirar la distancia entre M y M moño.**

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

El método de SVD nos GARANTIZA que elegimos los mejores r vectores (combinaciones de features) para minimizar esta norma!

SVD • Hiperparámetro r

¿Cómo podríamos elegir el **valor de r**?

Una posibilidad es **mirar la distancia entre M y M moño.**

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

El método de SVD nos GARANTIZA que elegimos los mejores r vectores (combinaciones de features) para minimizar esta norma!

Full-Rank Dog



Rank 200 Dog



Rank 100 Dog



Rank 50 Dog



Rank 30 Dog



Rank 20 Dog



Rank 10 Dog



Rank 3 Dog



SVD • Hiperparámetro r

¿Cómo podríamos elegir el **valor de r** ?

Una posibilidad es **mirar la distancia entre M y M moño.**

Otra posibilidad es **tener algún criterio sobre el peso relativo** de los valores singulares seleccionados respecto a la suma de todos. (Es más costoso, hay que calcular todos los valores singulares)

¿Y no hay algo un poco más visual?

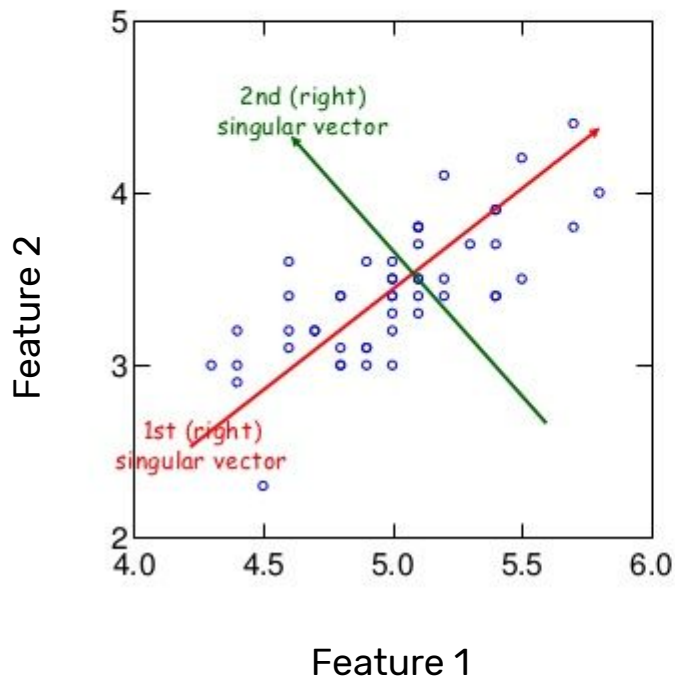


Aprendizaje No Supervisado

Representación gráfica SVD

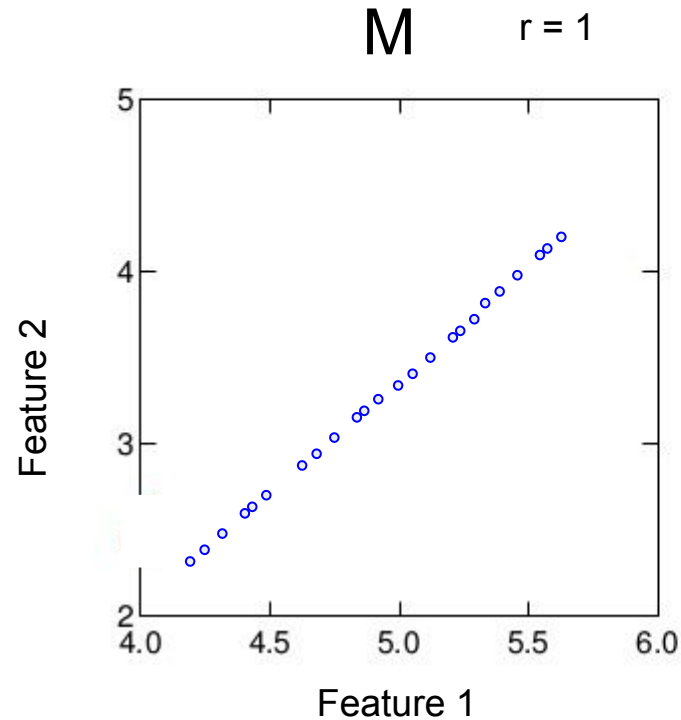
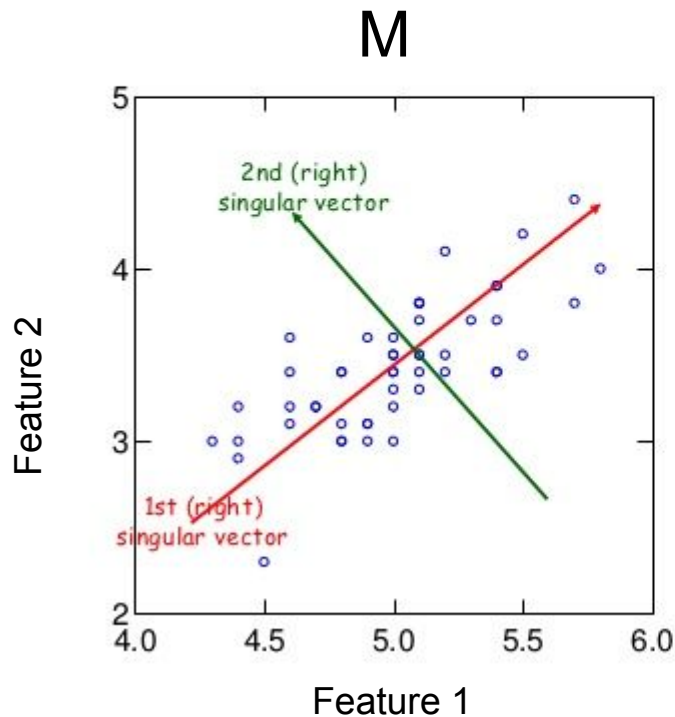


SVD • Representación gráfica



- El espacio original tiene 2 coordenadas, 2 features. Esto sirve para definir la posición de todas las instancias del dataset (cada punto azul).
- SVD nos da dos nuevos vectores, el 1er y 2do vector singular. Si usamos ambos como coordenadas, podemos definir perfecto la posición de cada punto.
- Veamos qué pasa si ahora sólo usamos el primer vector singular para definir los puntos.

SVD • Representación gráfica



A close-up photograph of a white ceramic cup filled with a latte. The surface of the milk is decorated with intricate latte art, featuring a central heart shape surrounded by concentric, wavy lines. The cup is placed on a matching white saucer. In the background, a white napkin and a silver fork are visible, though they are out of focus. The overall lighting is soft and even, highlighting the textures of the coffee and the smooth surface of the cup.

¡BREAK!



Hands-on training



Hands-on training



DS_Encuentro_38_SVD.ipynb

Para la próxima: Data Science en mi vida



Data Science en mi vida

En 10/15 minutos, tendrán que contar a sus compañeros y equipo docente lo siguiente:

- a) En qué problemas estoy aplicando lo aprendido en DS y cómo lo estoy encarando.
- b) Contar algún tema que me interese o que proyecto aplicar relacionado con lo que vimos.

Para la próxima

1. Terminar de ver los videos de Reducción de Dimensionalidad.
2. Completar los notebooks de hoy y atrasados.
3. Si están leyendo sobre PCA, pueden jugar con esta página:
<http://setosa.io/ev/principal-component-analysis/>.
4. Preparar el relato “Data Science en mi vida”.

ACÀMICA