ACÁMICA

¡Bienvenidas/os a Data Science!

Agenda

Actividad: Data Science en mi vida

Explicación: SVD - Compresión de imágenes

Break

Notebook: SVD sklearn

Explicación: PCA - Notebook de hoy

Hands-on training

Cierre

¿Dónde estamos?

BLOQUE 2 (Parte 2)	Procesamiento del lenguaje natural	Semana 13	Modelos avanzados - SVM Sesgo y Varianza
		Semana 14	Ensambles, Bagging, Random forest Ensambles, Boosting
		Semana 15	Redes Neuronales: Descenso por gradiente Redes Neuronales: Perceptrón
		Semana 16	Redes Neuronales: Perceptrón Multicapa Redes Neuronales: Repaso
		Semana 17	Procesamiento del lenguaje natural (NLP)
		Semana 18	Trabajo sobre el proyecto Intro aprendizaje no supervisado + Clustering
	Sistema de recomendación	Semana 19	Métricas de evaluación para clustering Reducción de dimensionalidad: SVD
		Semana 20	PCA Sistemas de recomendación
		Semana 21	Sistemas de recomendación Ecosistema digital Trabajo sobre el proyecto
		Semana 22	Ecosistema digital Puesta en producción

Actividad:
Data Science en mi vida

Data Science en mi vida

¡Preparen sus charlas relámpago!
En 7 minutos con 7 slides comparte con tus compañeros:

En qué problemas estás aplicando lo aprendido en Data Science y cómo lo estás haciendo.
O bien, en qué problemas te gustaría aplicar Data Science y cómo lo harías.

¡Elige algún tema o proyecto que te interese y relaciónalo con lo aprendido!

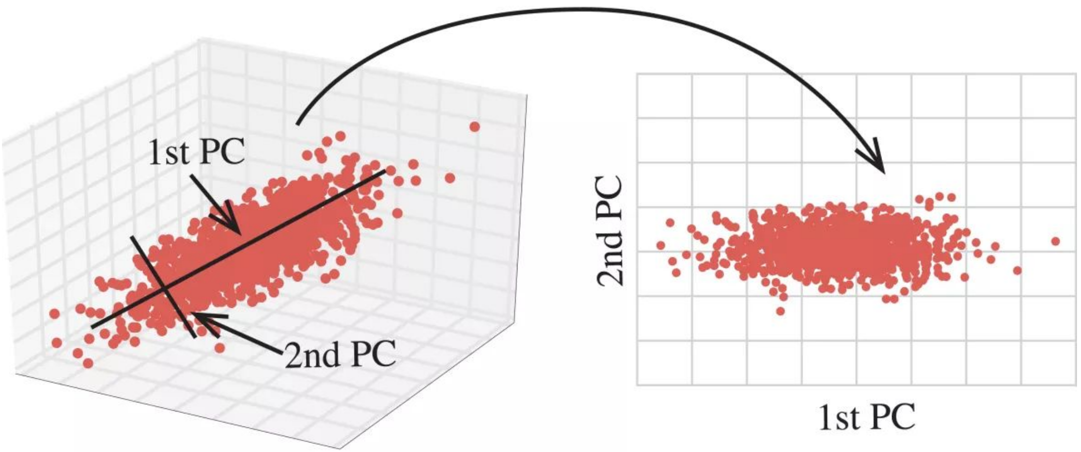
Reducción de la dimensionalidad

- 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.

Reducción



3 features originales

2 features <u>nuevas</u>

¿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
- Reducir overfitting
- Muchísimas mas cosas

¿Cómo se hace?

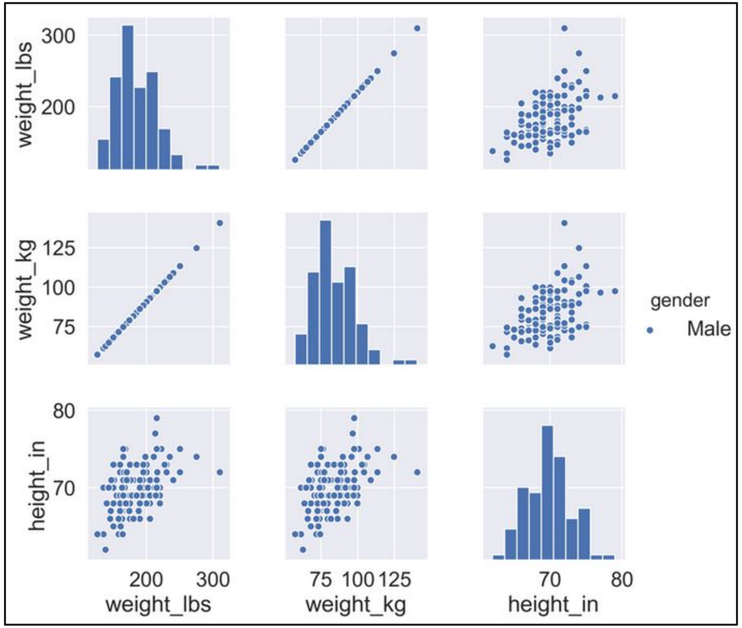
Algunos de los métodos de reducción de dimensionalidad son:

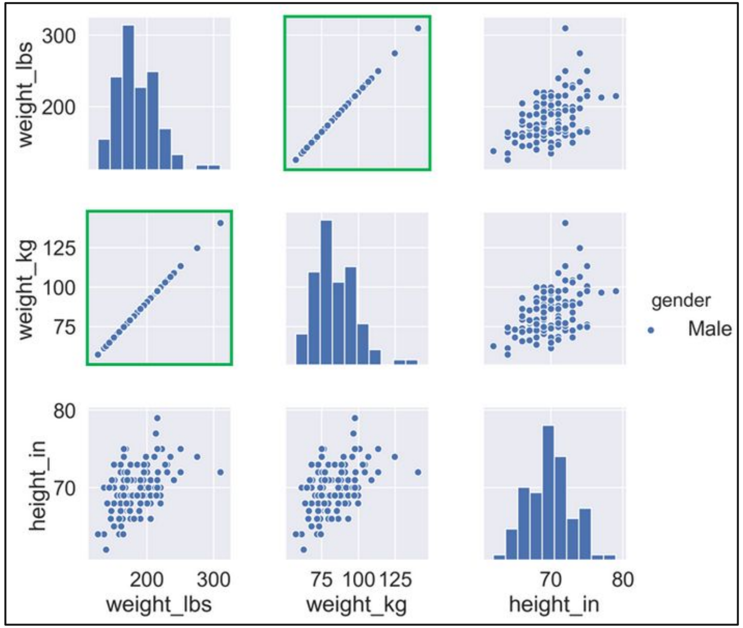
- PCA: Principal Component Analysis (usa SVD)
- MDS: Multidimensional scaling
- t-SNE: t-distributed Stochastic Neighbor Embedding
- Auto-Encoders (Se hace con Redes Neuronales)
- LDA: Linear Discriminant Analysis (si hay etiquetas de clases)

Reducción MANUAL

Usando el dataset ANSUR (anthropometric survey), que contiene los datos antropométricos de la Armada Estadounidense

```
sns.pairplot(ansur_df, hue="gender", diag_kind='hist')
```



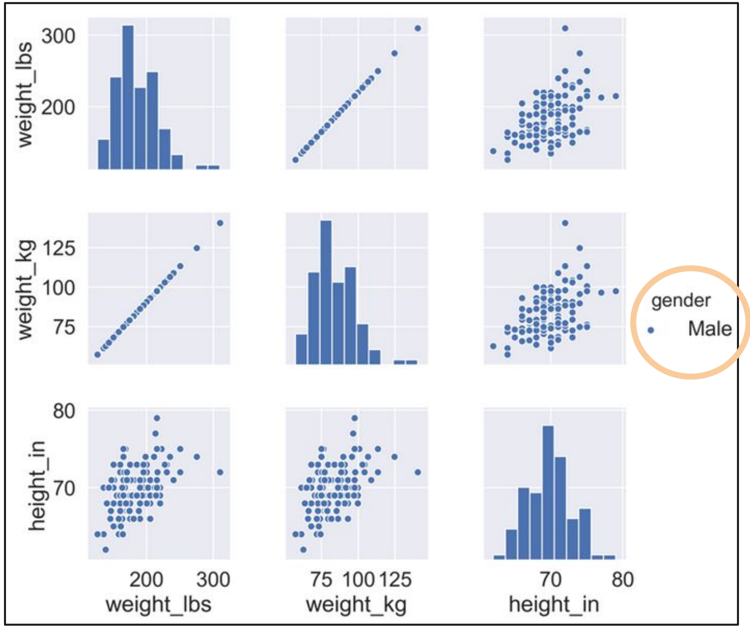


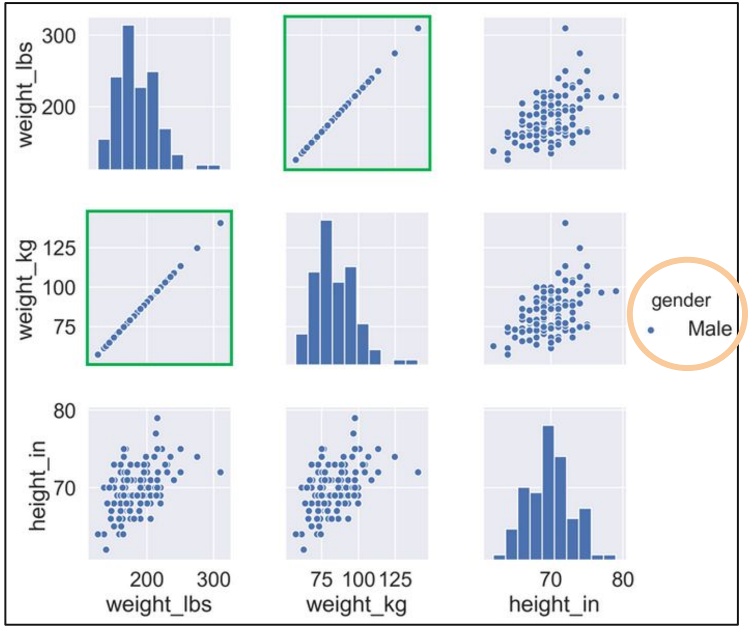
A simple vista, aparece otra feature que también podríamos eliminar.

¿Cuál es?

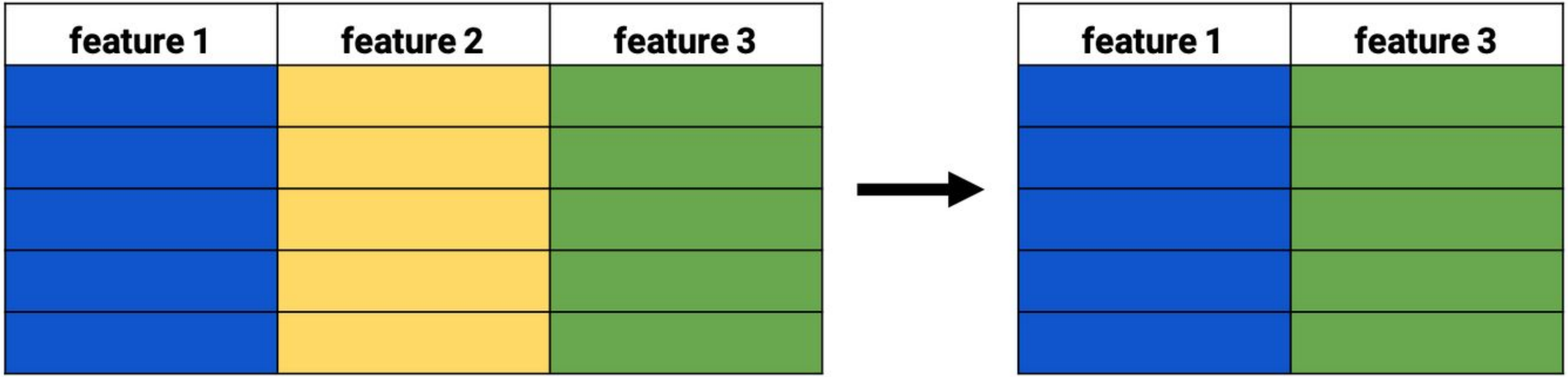
A simple vista, aparece otra feature que también podríamos eliminar.

¿Cuál es?



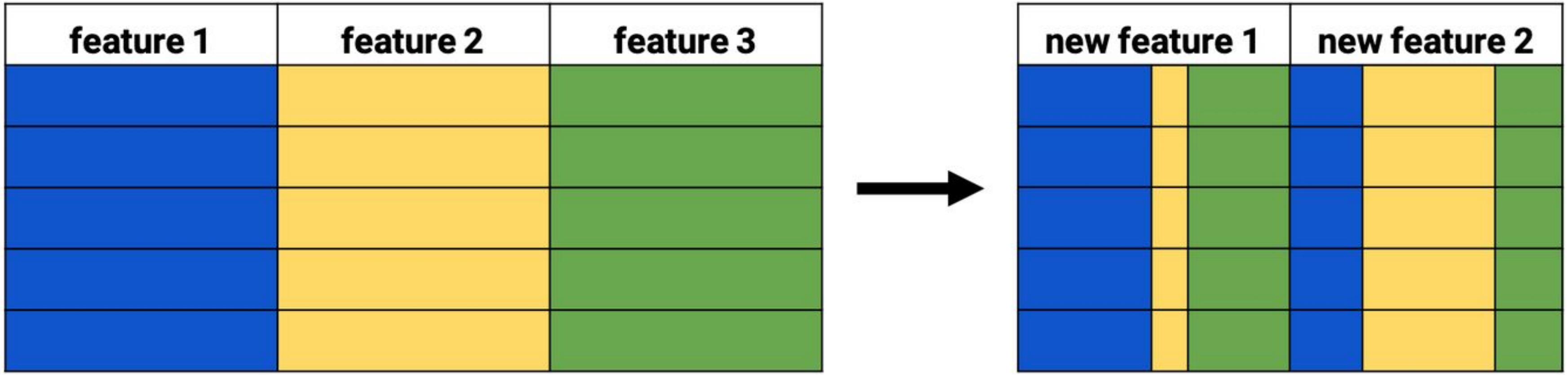


Esto que hicimos, también es conocido como **FEATURE SELECTION**



Elegimos, bajo cierto criterio, las features que van a formar parte del dataset “final”.

Otra forma de reducir la dimensionalidad es haciendo **FEATURE EXTRACTION**. Es un enfoque muy distinto al anterior, pero que busca cumplir el mismo objetivo.



“Extraemos” nuevas features a partir de las originales. Estas nuevas features tienen la menor redundancia de información posible, por lo tanto, son menos cantidad.

¿Se les ocurre una desventaja al trabajar las features de esta manera?

Aprendizaje No Supervisado

SVD (Singular Value Decomposition)

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:

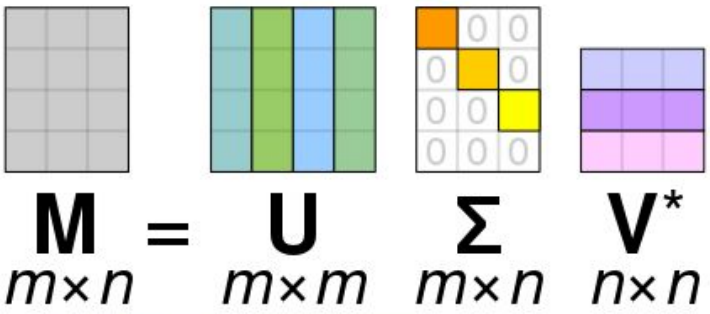
- **PCA: Principal Component Analysis (usa SVD)**
- MDS: Multidimensional scaling
- t-SNE: t-distributed Stochastic Neighbor Embedding
- Auto-Encoders (Se hace con Redes Neuronales)
- LDA: Linear Discriminant Analysis (si hay etiquetas de clases)

Aprendizaje No Supervisado

SVD truncado

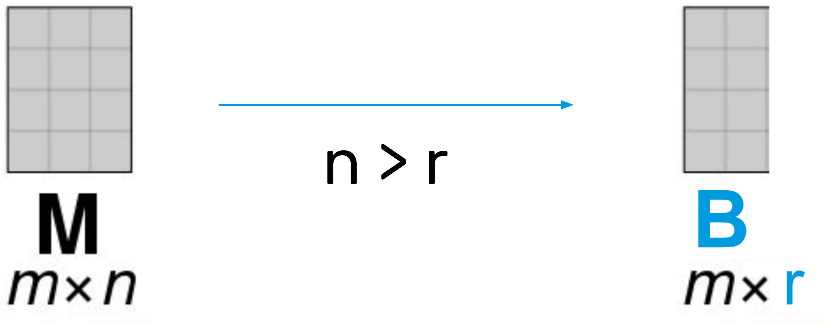
SVD • Definición

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



SVD • ¿Para qué sirve?

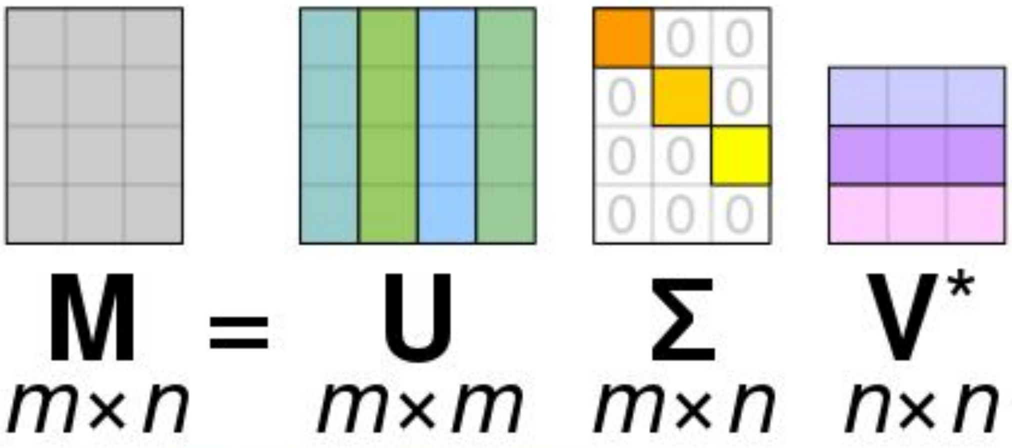
Para MUCHAS COSAS. Es parte del corazón de muchos algoritmos numéricos (solución sis. lineal, pseudoinversa,etc.). En este contexto vamos a usarlo para “reducir” adecuadamente la matriz M (pasar de tener muchos features a tener menos, pero que sean buenos).



SVD • Álgebra

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

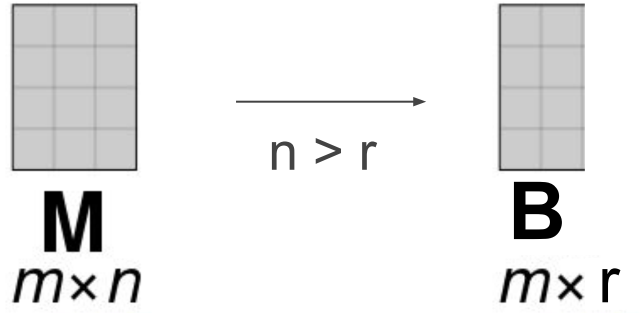
Matriz de Datos (m instancias, n features)	Matriz de vectores singulares por izquierda	Matriz de los <u>valores singulares</u>	Matriz de vectores singulares por derecha



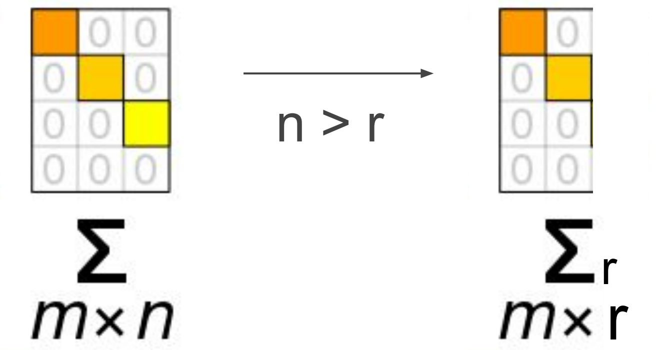
	Matriz Unitaria	Matriz Diagonal	Matriz Unitaria

SVD **truncado**

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



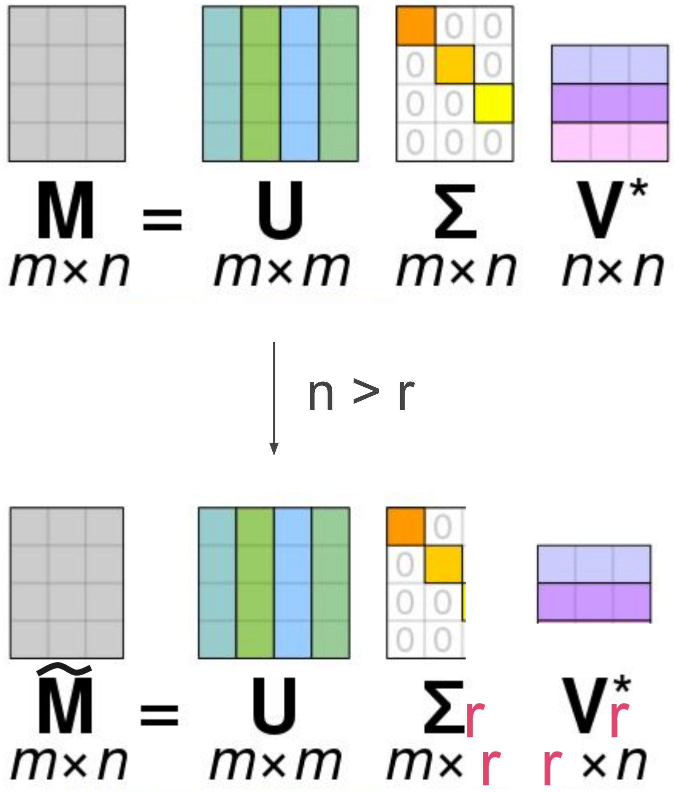
Idea de cómo lograrlo: si tomamos solo los r valores principales (elementos en la diagonal de Sigma) de valor más grande, podemos construir una matriz B que sea una "buena" reducción de M.



SVD **truncado**

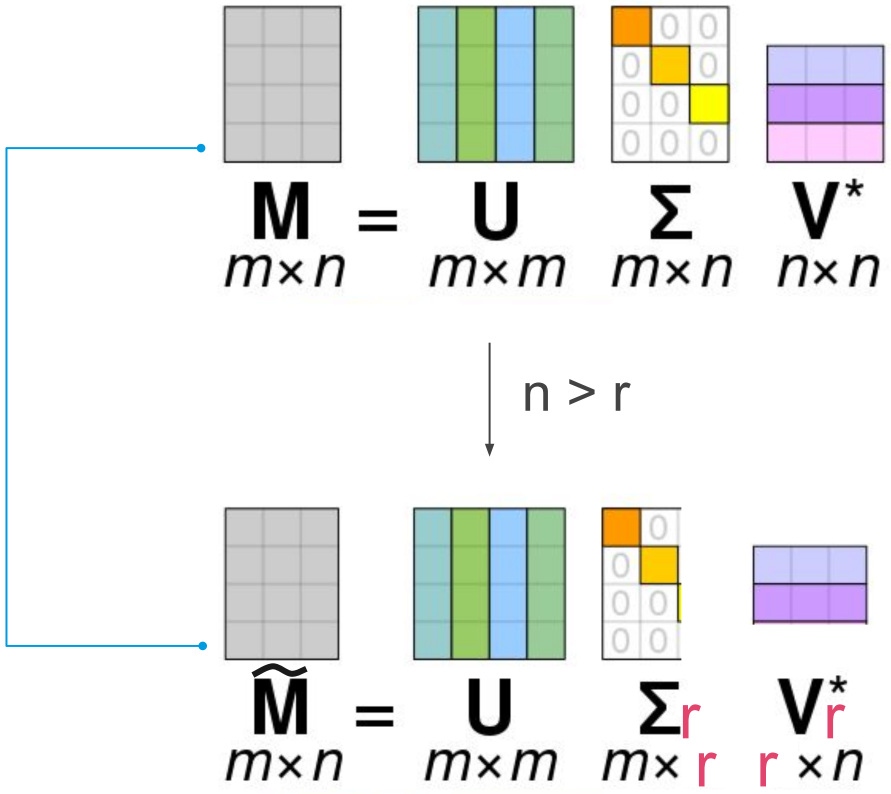
Matriz completa: es la M original, tiene toda la información.

Matriz truncada: perdimos información. Pero si tomamos un valor de r adecuado, 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.

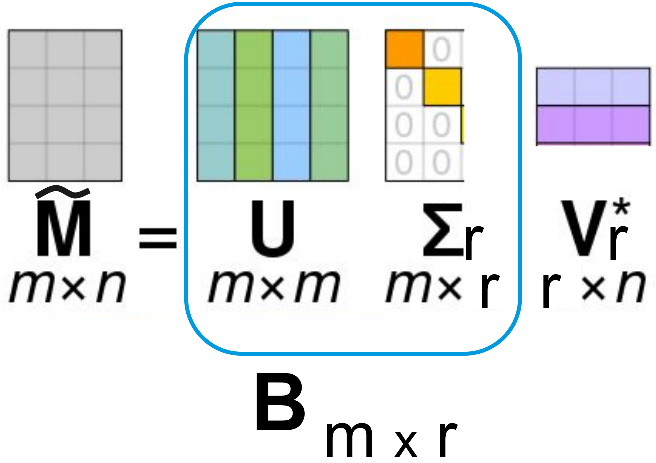


SVD **truncado**

Parecidas



SVD truncado



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

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

SVD • **Hiperparámetro r**

¿Cómo podríamos elegir el valor de r?

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

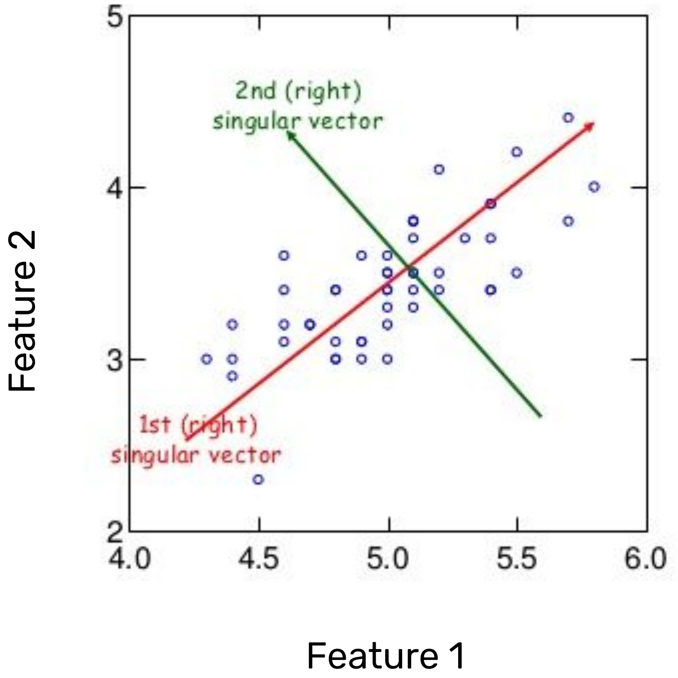
$$\| M - \widetilde{M} \|_F = \sqrt{\sum_{ij} (M_{ij} - \widetilde{M}_{ij})^2}$$

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

Aprendizaje No Supervisado

Representación gráfica SVD

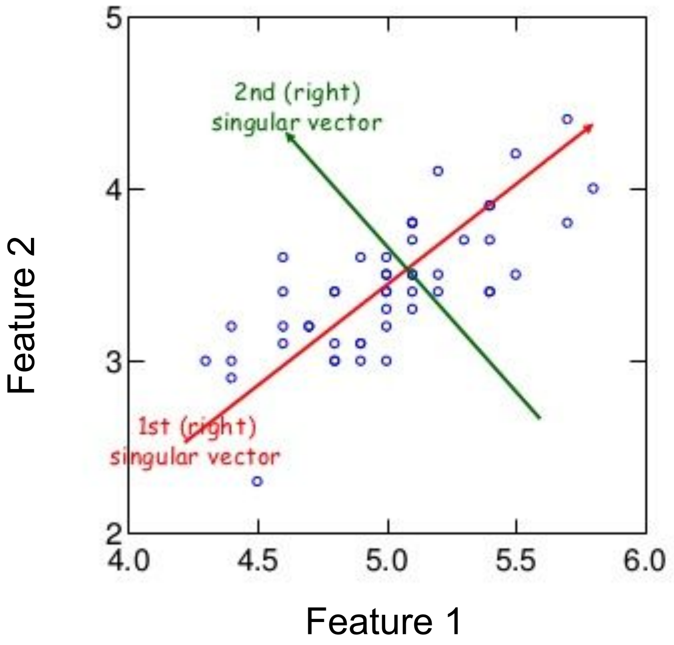
SVD • Representación gráfica



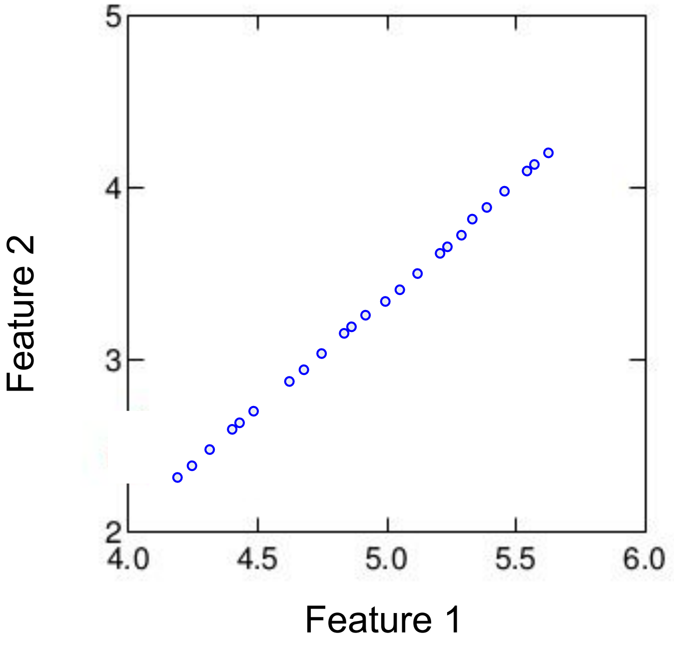
- El espacio original tiene 2 coordenadas, 2 features. Esto sirve para definir la posición de todas las instancias del dataset (cada punto azul).
- SVD nos da dos nuevos vectores, el 1er y 2do vector singular. Si usamos ambos como coordenadas, podemos definir perfecto la posición de cada punto.
- Veamos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

SVD • Representación gráfica

M



M r = 1

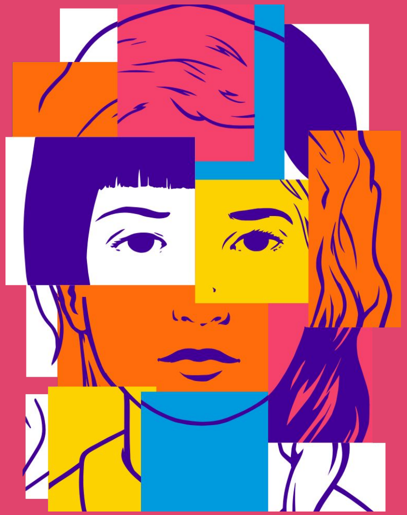


Hands-on training

DS_Encuentro_38_SVD.ipynb

Aprendizaje No Supervisado

PCA (Principal Component Analysis)



“Por lo que vimos en clase y lo que vi en los videos, SVD y PCA parecen ser lo mismo…”

Anónimo

“Casi, pero no, no son lo mismo.”

(otro) Anónimo

PCA • Definición

PCA es el método de reducción de dimensionalidad más utilizado.

¿Y PCA es muy distinto a usar un SVD truncado?

PCA • Definición

PCA y SVD truncado son casi iguales, solo que existe una diferencia:

PCA = Centrar datos + SVD truncado

PCA • Definición

PCA y SVD truncado son casi iguales, solo que existe una diferencia:

PCA = Centrar datos + SVD truncado

Debemos sustraer la media de cada columna de Features antes de aplicar SVD truncado.

¿Por qué PCA es tan relevante? ¿Por qué no se lo ve solo como un caso particular de SVD?

PCA

PCA • Importancia

- Matemáticamente, se puede llegar por otro camino (Matriz de covarianza)
- Tiene una interpretación muy intuitiva:

Componentes Principales → **Direcciones de máxima varianza**

La primer componente principal está en la dirección donde los datos presentan varianza máxima.

La segunda componente principal está la segunda dirección en términos de la varianza, y así sucesivamente.

¿Cuándo uso PCA y cuándo SVD?

Comparación • PCA vs. SVD

Para muchos casos, el resultado de usar uno u otro método va a ser muy parecido. Dependiendo del problema, puede que restar la media y trabajar con la distancia de cada feature a la media en el dataset sea mejor (si no saben, lo más común es usar PCA).

Analogías

PCA	SVD
Numero de componentes	Rango R
Componentes principales	Vectores singulares por derecha
Autovalores	Valores singulares
Maximiza Varianza	Minimiza Distancia

¡BREAK!

Ph. Credit: Drew Coffmann

Hands-on training

Hands-on training

DS_Encuentro_39_PCA.ipynb

Para la próxima

1. Ver los videos de la plataforma “Aprendizaje No Supervisado: Sistemas de recomendación”.
2. Completar el notebook de hoy y atrasados.
3. Bajar el siguiente dataset https://www.kaggle.com/netflix-inc/netflix-prize-data

ACÁMICA