



Ayudantía 11

Aprendizaje no supervisado

Por Martín Vial y Stephen Pugh

10 de junio 2024

¿Qué es el aprendizaje no supervisado?



- Aprendizaje automático que aprende de datos no etiquetados.
- Los datos no tienen etiquetas o categorías preexistentes.
- El objetivo del aprendizaje no supervisado es descubrir patrones y relaciones en los datos sin ninguna guía explícita.
- La tarea de la máquina es agrupar información no ordenada según similitudes, patrones y diferencias sin ningún entrenamiento previo de los datos.

¿Qué es el aprendizaje no supervisado?



- A diferencia del aprendizaje supervisado, no se proporciona una etiqueta, lo que significa que no se dará ningún entrenamiento al algoritmo.
- El algoritmo está limitado a encontrar la estructura oculta en los datos no etiquetados por sí misma, bajo sus métodos establecidos.



Clustering

- En resumen, el clustering es **agrupar datos** similares, especialmente en el contexto de datos no etiquetados, donde no hay un *label* fijo al que predecir
- En clases se vió *k-means*, clustering jerárquico y DBSCAN, que tienen distintas funciones



Clustering

- En resumen, el clustering es **agrupar datos** similares, especialmente en el contexto de datos no etiquetados, donde no hay un *label* fijo al que predecir
- En clases se vió *k-means*, clustering jerárquico y DBSCAN, que tienen distintas funciones
 - **K-means:** Encontrar un número adecuado de clusters



Clustering

- En resumen, el clustering es **agrupar datos** similares, especialmente en el contexto de datos no etiquetados, donde no hay un *label* fijo al que predecir
- En clases se vió *k-means*, clustering jerárquico y DBSCAN, que tienen distintas funciones
 - **K-means:** Encontrar un número adecuado de clusters
 - **Jerárquico:** Administrar adecuadamente la cantidad de clusters



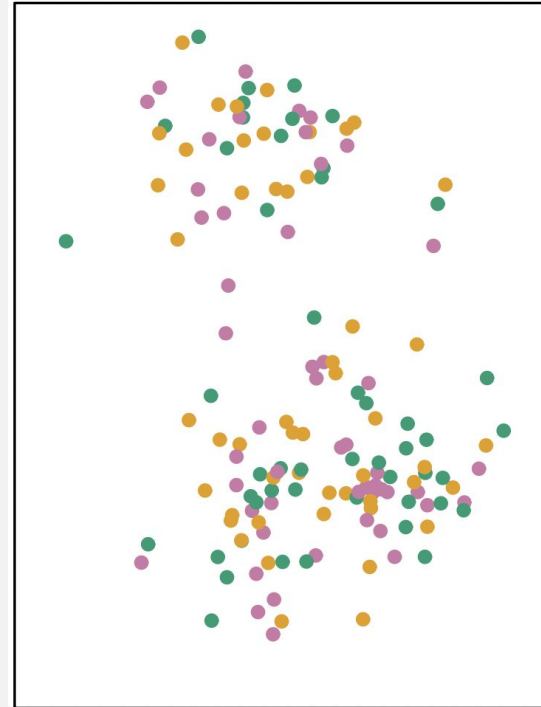
Clustering

- En resumen, el clustering es **agrupar datos** similares, especialmente en el contexto de datos no etiquetados, donde no hay un *label* fijo al que predecir
- En clases se vió *k-means*, clustering jerárquico y DBSCAN, que tienen distintas funciones
 - **K-means:** Encontrar un número adecuado de clusters
 - **Jerárquico:** Administrar adecuadamente la cantidad de clusters
 - **DBSCAN:** Hacer clusters por densidad en vez de distancia



Algoritmo K-means

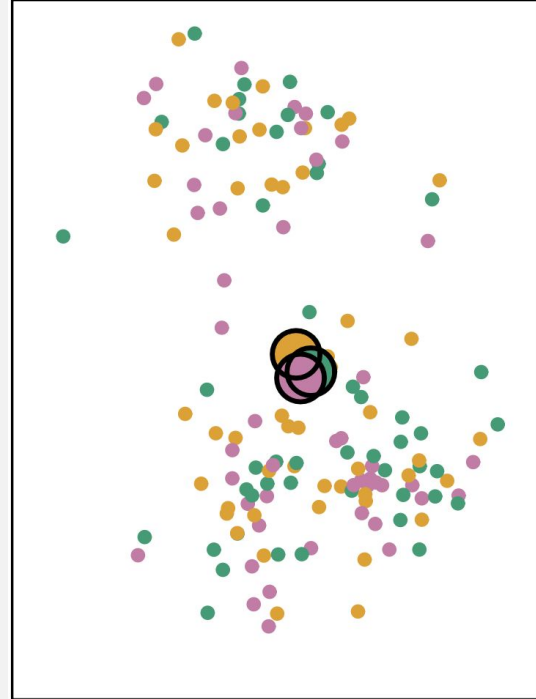
1. **Asignar** aleatoriamente una clase k a cada dato





Algoritmo K-means

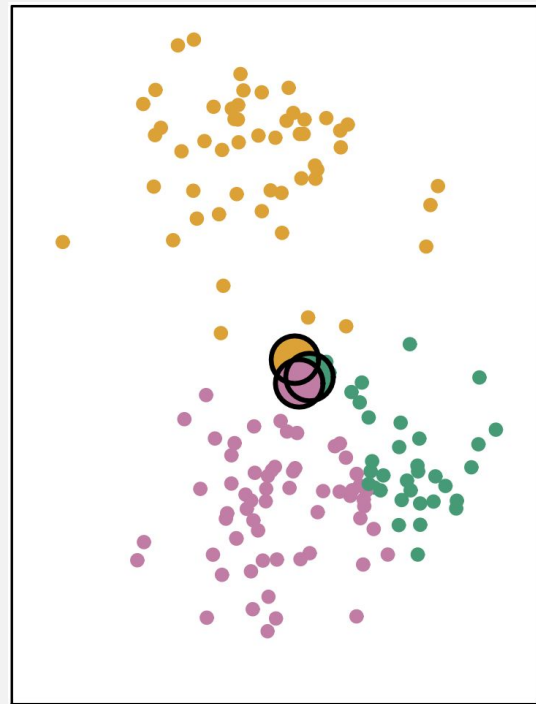
1. Asignar aleatoriamente una clase k a cada dato
2. **Inicializar** k clusters en una posición aleatoria del plano





Algoritmo K-means

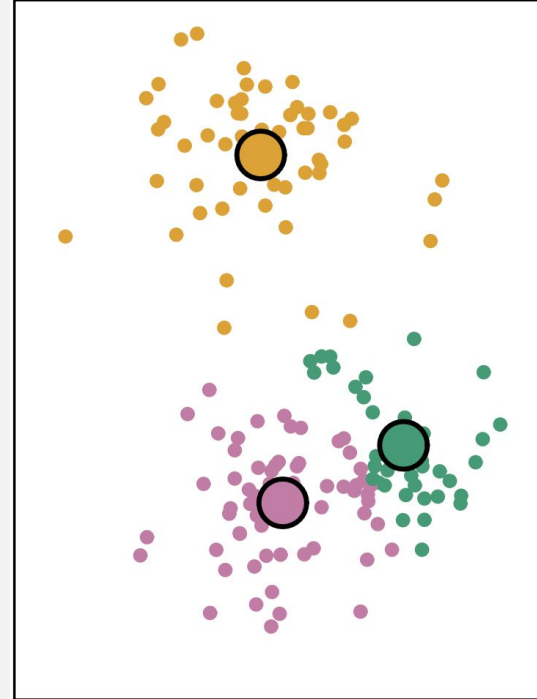
1. Asignar aleatoriamente una clase k a cada dato
2. Inicializar k clusters en una posición aleatoria del plano
3. **Cambiar** la clase de cada dato a su cluster más cercano





Algoritmo K-means

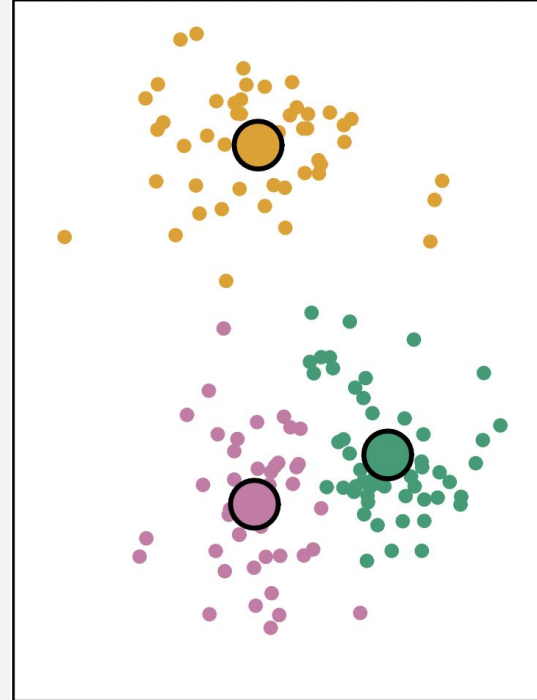
1. Asignar aleatoriamente una clase k a cada dato
2. Inicializar k "centroides" en una posición aleatoria del plano
3. Cambiar la clase de cada dato a su centroide más cercano
4. **Mover** el centroide efectivamente el centro de cada cluster





Algoritmo K-means

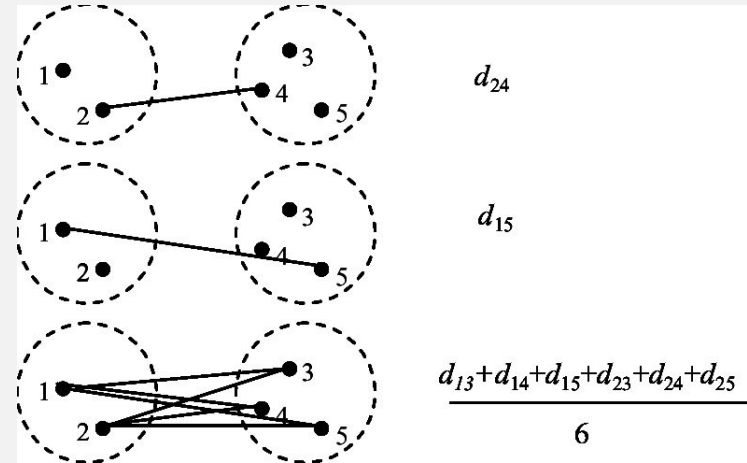
1. Asignar aleatoriamente una clase k a cada dato
2. Inicializar k "centroides" en una posición aleatoria del plano
3. Cambiar la clase de cada dato a su centroide más cercano
4. Mover el centroide efectivamente el centro de cada cluster
5. **Repetir** 3 y 4 hasta que el centroide no se mueva (converge)





Clustering Jerárquico

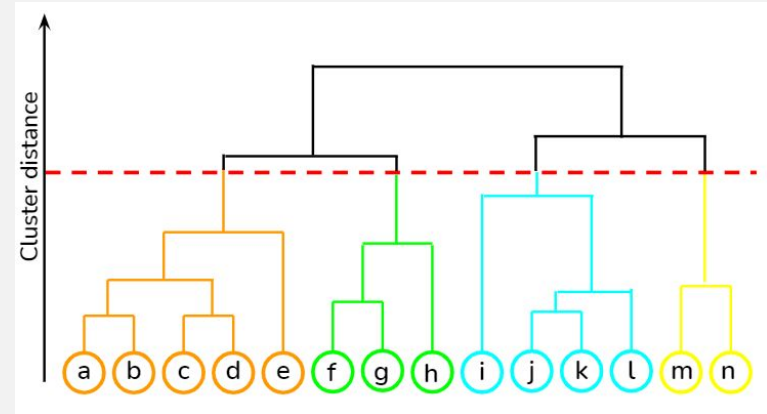
- Una vez que tenemos clusters, medimos la **distancia** entre ellos
 - Single link (distancia mínima)
 - Complete link (distancia máxima)
 - Average
 - Distancia entre centroides





Clustering Jerárquico

- Una vez que tenemos clusters, medimos la **distancia** entre ellos
 - Single link (distancia mínima)
 - Complete link (distancia máxima)
 - Average
 - Distancia entre centroides
- Comparar distancias entre clusters nos sirve para decidir si elegimos muy pocos, demasiados y cómo se relacionan entre sí los clusters



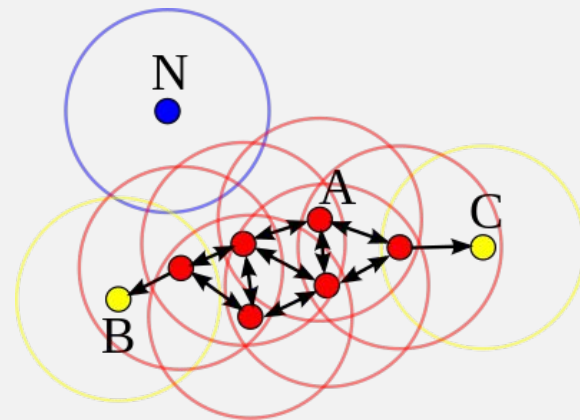


DBSCAN

Density Based Spatial Clustering of Applications of Noise

Considerando un radio (parámetro *Eps*) y un mínimo de puntos por región (parámetro *MinPts*) clasificamos a los puntos como

- **Core:** Hay más puntos que *MinPts* en un área con radio *Eps*
- **Border:** Menos de *MinPts* puntos, pero con un vecino core en su radio *Eps*
- **Noise:** Ninguna de las anteriores





Características de cada algoritmo

K-means	
Pros	Contras
Fácil de implementar	Problemas con alta densidad
Escalable a un <i>dataset</i> grande	Hay que determinar k
Buena forma de “explorar”	Puede caer en máximos locales



Características de cada algoritmo

Clustering Jerárquico	
Pros	Contras
Determina k automáticamente	Poco escalable
Describe el dataset en lo macro y lo micro	Sensible a los <i>outliers</i>



Características de cada algoritmo

DBSCAN	
Pros	Contras
Determina k automáticamente	Problemas con densidad variada
Clusters de forma arbitraria (ojo humano)	Poco escalable
Maneja bien los <i>outliers</i>	



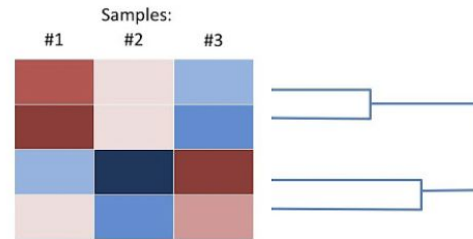
Material

Videos prácticos: StatQuest

https://www.youtube.com/watch?v=Gv9_4yMHFhI&list=PLblh5JKOoLUICTaGLRoHQDuF_7q2GfuJF

(playlist Machine Learning)

Hierarchical Clustering, Clearly Explained!!!!



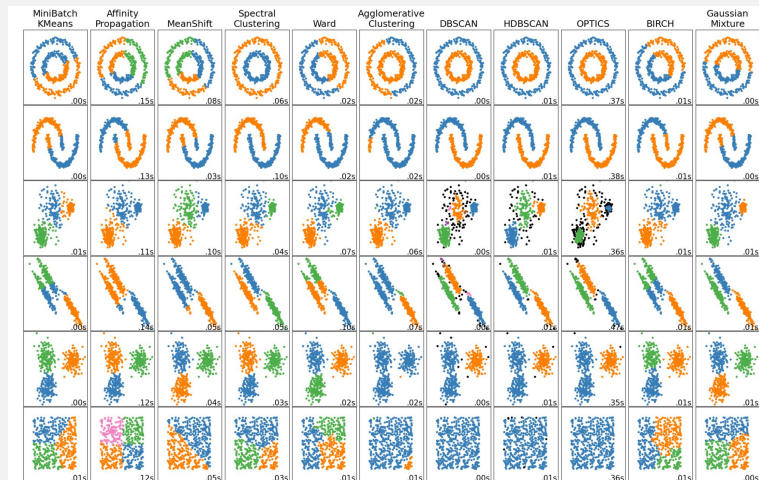
11:19



Material

Documentación: Scikit-learn

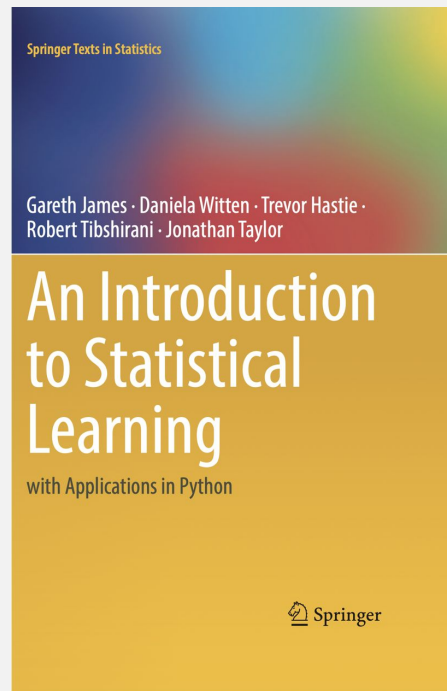
<https://scikit-learn.org/stable/modules/clustering>



Material

Materia: An Introduction to Statistical Learning (with Applications in Python)

<https://www.statlearning.com>





¡Vamos al código!



Transformación de características

- Proceso de modificar o convertir las características de input en otro conjunto de datos para mejorar el rendimiento de un modelo de aprendizaje automático.
- Se aplican operaciones matemáticas a las características para hacerlas más adecuadas para el algoritmo de aprendizaje.
- El objetivo es crear un nuevo conjunto de características que capturen mejor los patrones y relaciones subyacentes en los datos, mejorando en última instancia la capacidad del modelo para hacer predicciones o clasificaciones precisas.

PCA (Principal Components Analysis)



- Algoritmo no supervisado que se utiliza para reducir la dimensionalidad de un conjunto de datos y resaltar las características más importantes.
- Se basa en la idea de que la mayor parte de la variabilidad de los datos puede ser representada por un conjunto más pequeño de características, conocidas como componentes principales
- La idea es transformar los datos originales X en datos nuevos Y que no tengan correlación entre ellos. Y es una transformación lineal de X .
- Se ordenan las nuevas características según las que aporten más varianza.



PCA - Ventajas

1. La representación de muchas características es mucho más sencilla ya que se reduce la dimensionalidad.
2. La clasificación usando las primeras componentes es más simple y más rápida para el clasificador.
3. Es posible reconstruir X a partir de las principales columnas de Y .

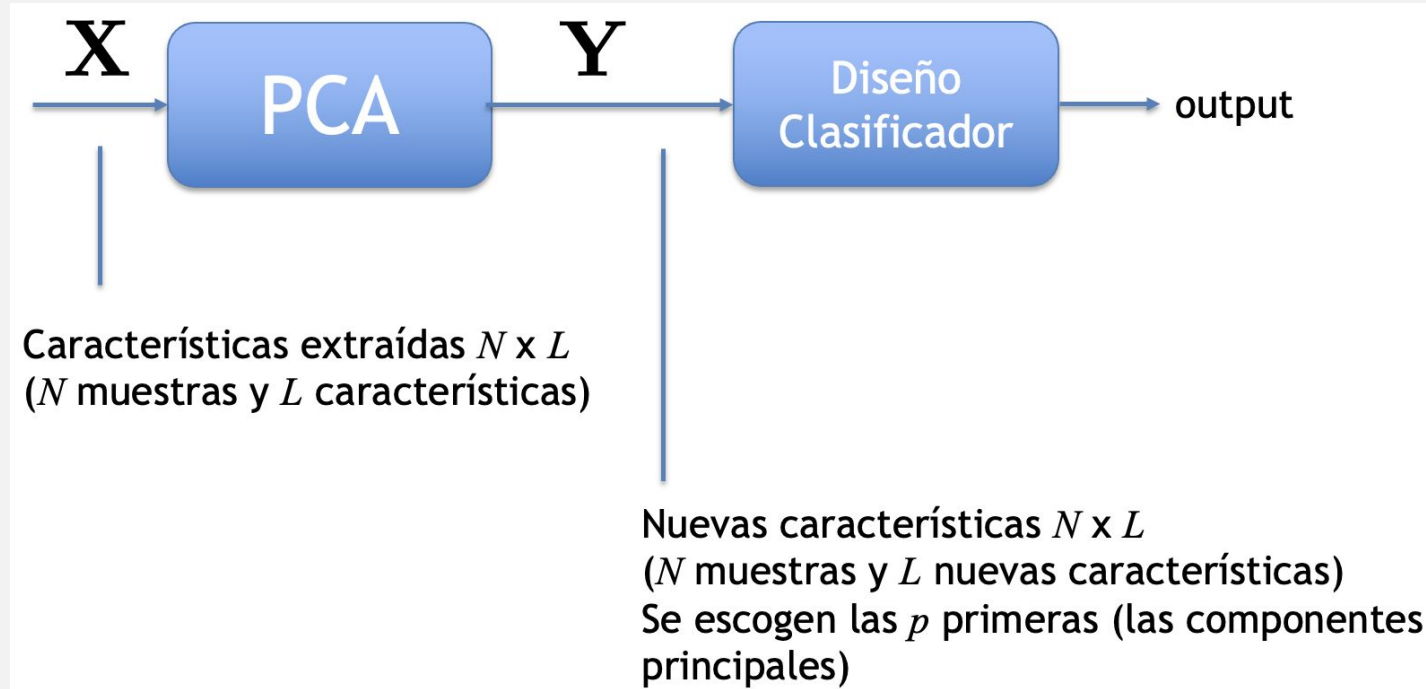


PCA - Desventajas

1. Las características de PCA dependen de todas las características extraídas (es necesario extraer todas las características X para obtener Y , no hay ahorro de cómputo en la extracción).
2. Al ser no supervisado no asegura que las características transformadas tengan una buena separabilidad.
3. No se pueden interpretar analíticamente las nuevas características.

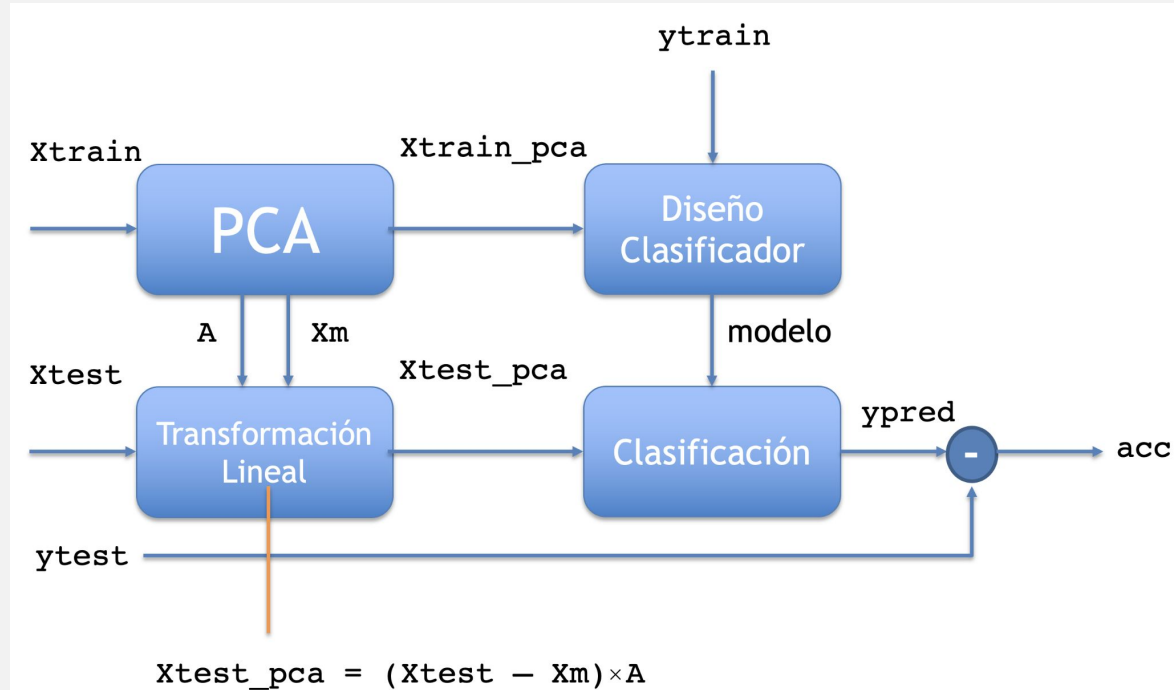


PCA - Aprendizaje no supervisado





PCA - Aprendizaje supervisado





Selección de características

1. Proceso que elige un subconjunto de características de las características originales.
2. El espacio de características se reduce de manera óptima según un criterio determinado.
3. Se crea un nuevo dataset con las características seleccionadas.



Objetivos

1. Evitar características no discriminatorias
2. Evitar características correlacionadas
3. Simplificar la etapa de prueba
4. Evitar correlaciones falsas
5. Evitar alta dimensionalidad innecesaria



SFS (Sequential Feature Selection)

- Algoritmo que agrega o elimina iterativamente características de un conjunto de datos con el fin de mejorar el rendimiento de un modelo predictivo.



SFS - Funcionamiento

1. El selector se inicializa con un modelo predictivo, el número de características a seleccionar, la métrica de evaluación y la tolerancia para la mejora.



SFS - Funcionamiento

1. El selector se inicializa con un modelo predictivo, el número de características a seleccionar, la métrica de evaluación y la tolerancia para la mejora.
2. El conjunto de características seleccionadas se inicia vacío.



SFS - Funcionamiento

1. El selector se inicializa con un modelo predictivo, el número de características a seleccionar, la métrica de evaluación y la tolerancia para la mejora.
2. El conjunto de características seleccionadas se inicia vacío.
3. Se añade una característica al conjunto seleccionado.



SFS - Funcionamiento

1. El selector se inicializa con un modelo predictivo, el número de características a seleccionar, la métrica de evaluación y la tolerancia para la mejora.
2. El conjunto de características seleccionadas se inicia vacío.
3. Se añade una característica al conjunto seleccionado.
4. El selector ajusta el modelo predictivo al conjunto de características seleccionadas y se evalúa el modelo con la selección.



SFS - Funcionamiento

1. El selector se inicializa con un modelo predictivo, el número de características a seleccionar, la métrica de evaluación y la tolerancia para la mejora.
2. El conjunto de características seleccionadas se inicia vacío.
3. Se añade una característica al conjunto seleccionado.
4. El selector ajusta el modelo predictivo al conjunto de características seleccionadas y se evalúa el modelo con la selección.
5. La característica que más mejora la puntuación de la validación cruzada del modelo se añade al conjunto de características seleccionadas.



SFS - Funcionamiento

1. El selector se inicializa con un modelo predictivo, el número de características a seleccionar, la métrica de evaluación y la tolerancia para la mejora.
2. El conjunto de características seleccionadas se inicia vacío.
3. Se añade una característica al conjunto seleccionado.
4. El selector ajusta el modelo predictivo al conjunto de características seleccionadas y se evalúa el modelo con la selección.
5. La característica que más mejora la puntuación de la validación cruzada del modelo se añade al conjunto de características seleccionadas.
6. El selector repite los pasos 3-5 hasta que se haya seleccionado el número deseado de características.



¡Vamos al código!