



# No Supervisado Dimensionality Reduction





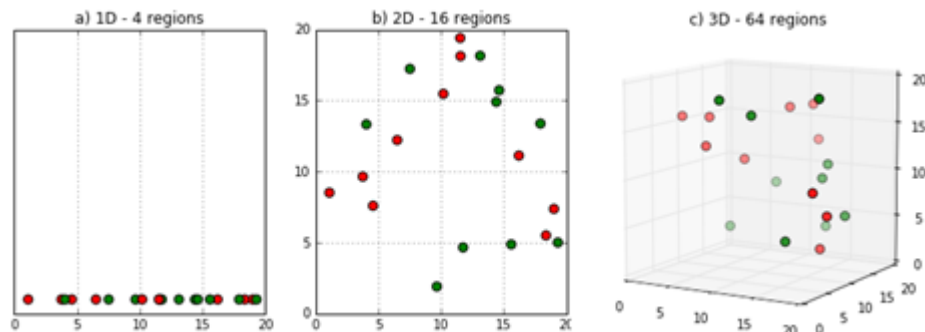
# Curse of Dimensionality



# Curse of Dimensionality

El número de muestras que se necesitan para estimar una función arbitraria (un target de ML, por ejemplo) con un cierto nivel de precisión crece exponencialmente con el número de inputs/dimensiones/variables de la función.

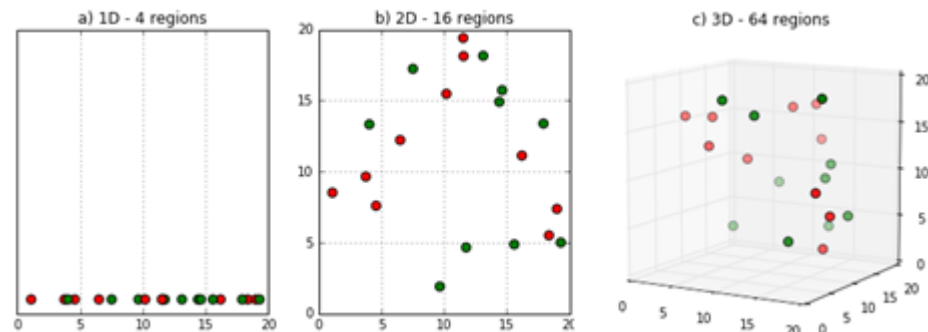
Este fenómeno afecta mucho a la dispersión y la cercanía de los datos. Según vamos añadiendo dimensiones, se van diferenciando mejor.



# Curse of Dimensionality

El número de muestras que se necesitan para estimar una función arbitraria (un target de ML, por ejemplo) con un cierto nivel de precisión crece exponencialmente con el número de inputs/dimensiones/variables de la función.

Este fenómeno afecta mucho a la dispersión y la cercanía de los datos. Según vamos añadiendo dimensiones, se van diferenciando mejor.



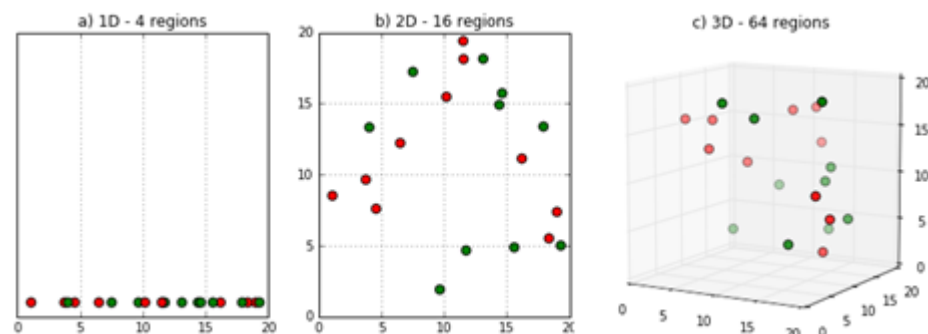
Cuando estamos ante pocas dimensiones, tenemos datos que pueden resultar muy parecidos, pero según vamos añadiendo características y dimensiones nuevas, esto cambia



# Curse of Dimensionality

El número de muestras que se necesitan para estimar una función arbitraria (un target de ML, por ejemplo) con un cierto nivel de precisión crece exponencialmente con el número de inputs/dimensiones/variables de la función.

Este fenómeno afecta mucho a la dispersión y la cercanía de los datos. Según vamos añadiendo dimensiones, se van diferenciando mejor.



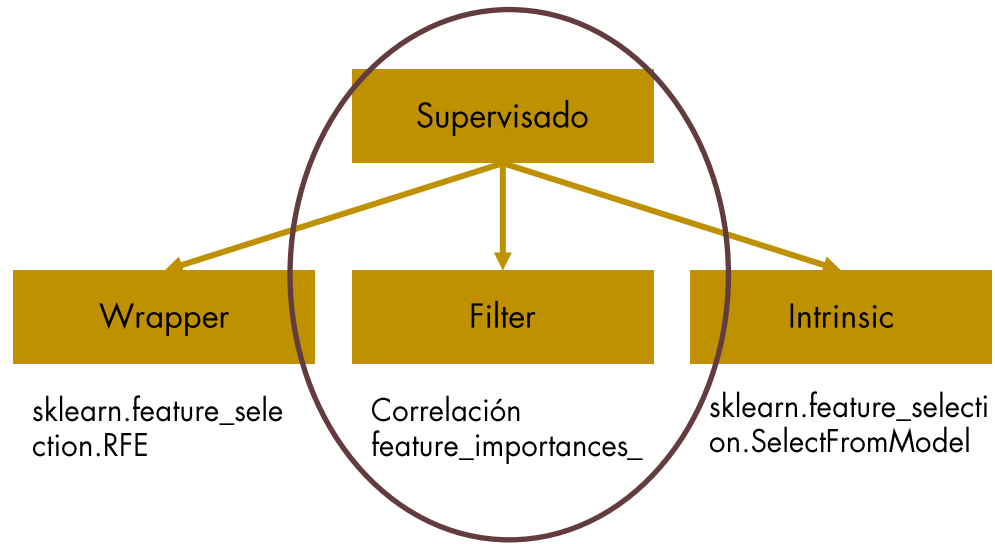
Cuando estamos ante pocas dimensiones, tenemos datos que pueden resultar muy parecidos, pero según vamos añadiendo características y dimensiones nuevas, esto cambia



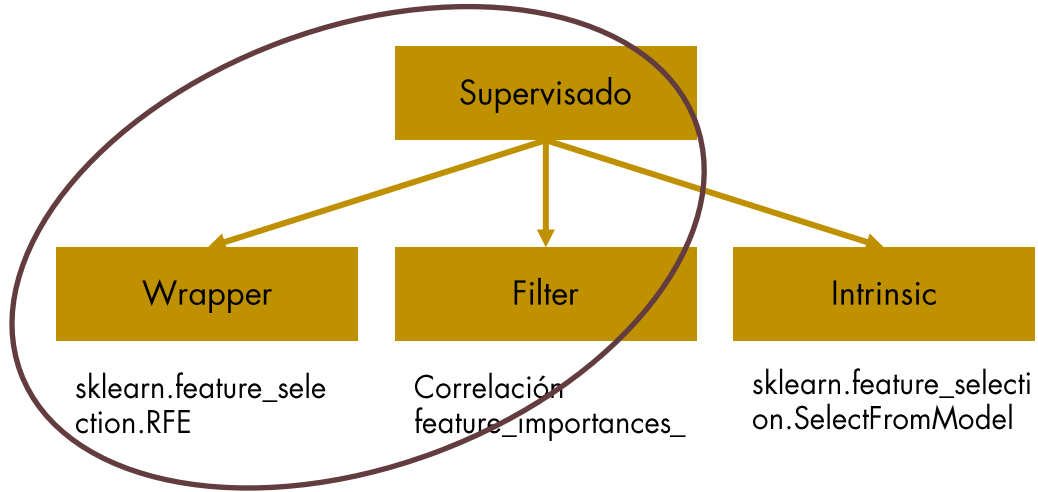
Datasets con muchas dimensiones serán muy dispersos y con mucha distancia entre los puntos, lo cual es bueno para clasificar. El problema es que nuevas observaciones estarán también muy lejanas de las originales (**overfitting**), produciendo predicciones menos fiables que datasets con pocas dimensiones. La solución sería incrementar el conjunto de train.



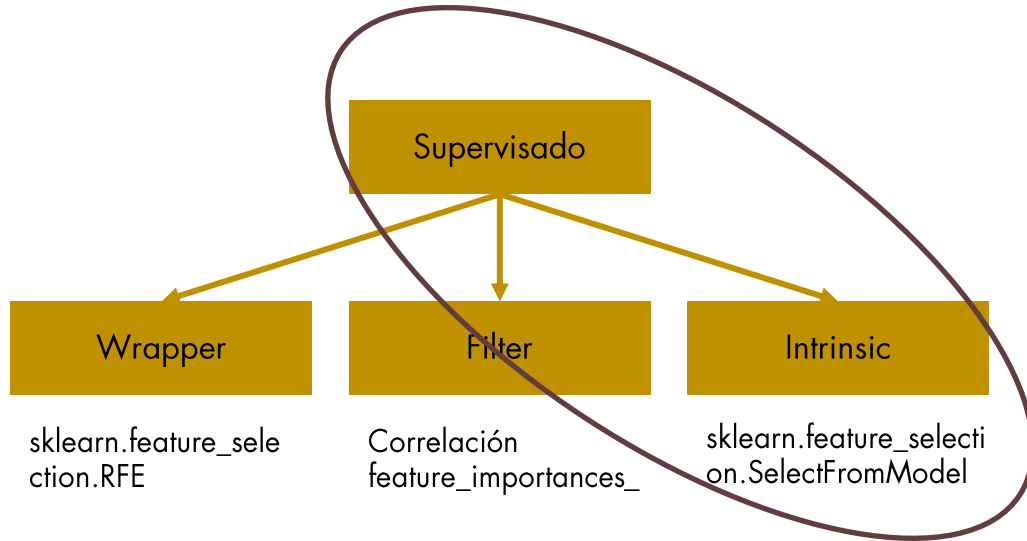
# Feature Reduction



# Feature Reduction

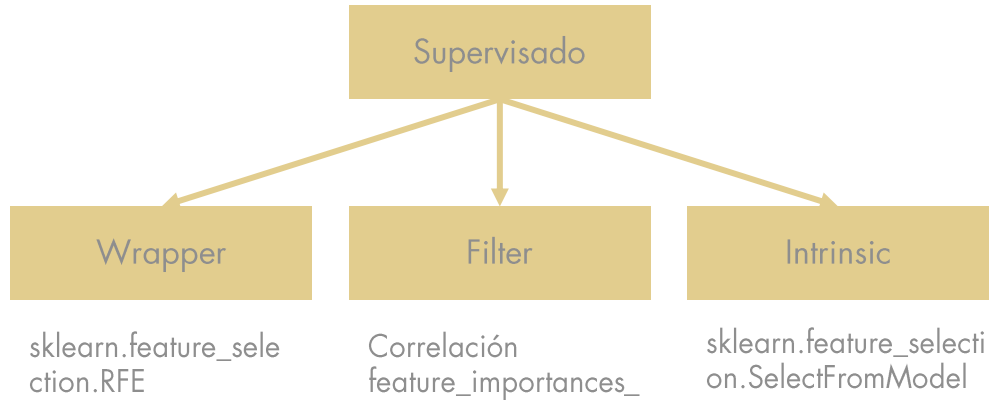


# Feature Reduction



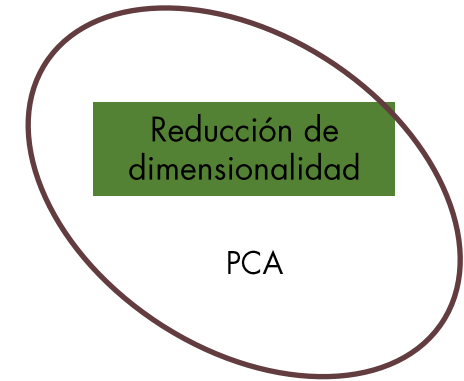


# Feature Reduction

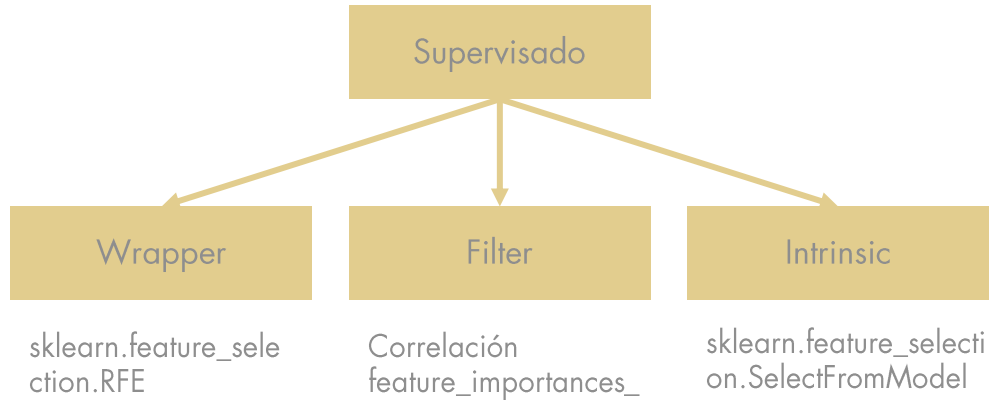


No Supervisado

sklearn.feature\_selection.VarianceThreshold



# Feature Reduction



No Supervisado

sklearn.feature\_selection.VarianceThreshold

Reducción de dimensionalidad

PCA

## Aplicaciones

- Mejora computacional
- Detección de features discriminantes
- Eliminación de features irrelevantes
- Compresión de la información (imágenes)
- Visualización (PCA)



