

Clase 22

Machine Learning

ECON. Y CIENCIA DE DATOS - EAE 253B

I SEM 2019

A solid orange horizontal bar spanning the width of the slide at the bottom.

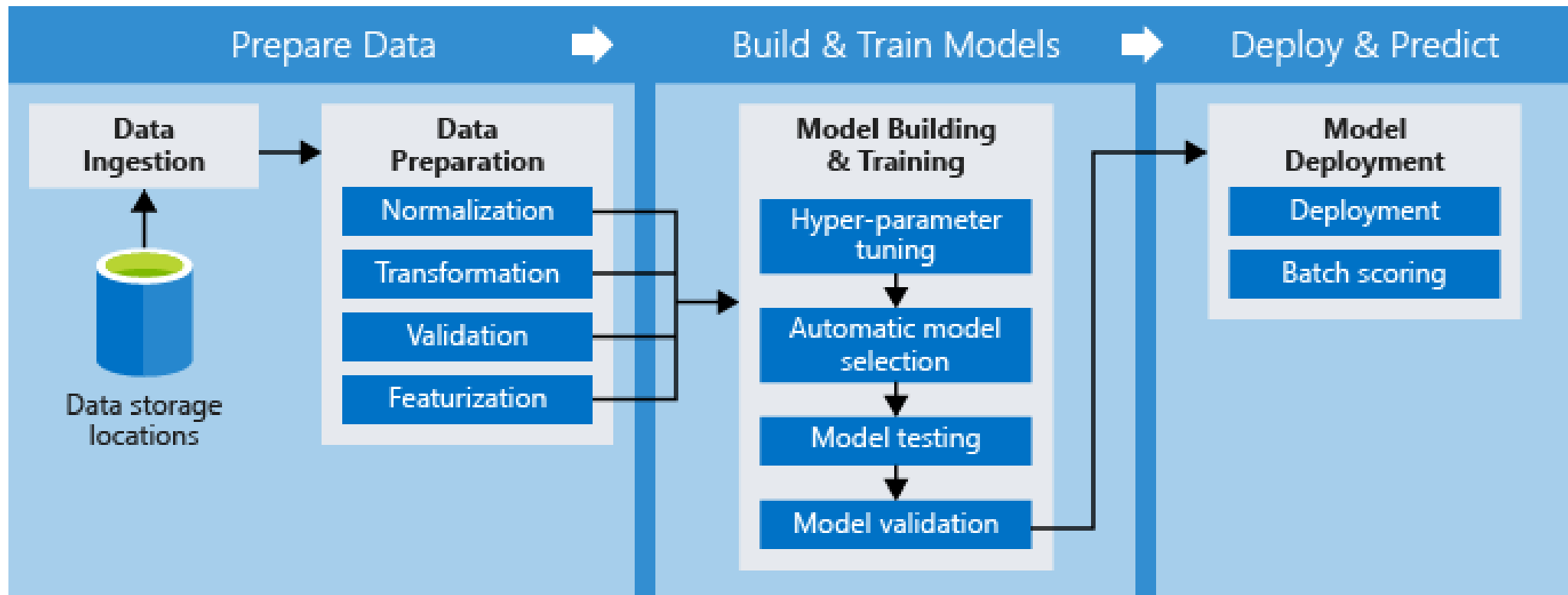
Machine Learning

Rama de la Ciencia de Datos
que busca resolver problemas
implementando programas
que “aprendan” de experiencias (*datos*)

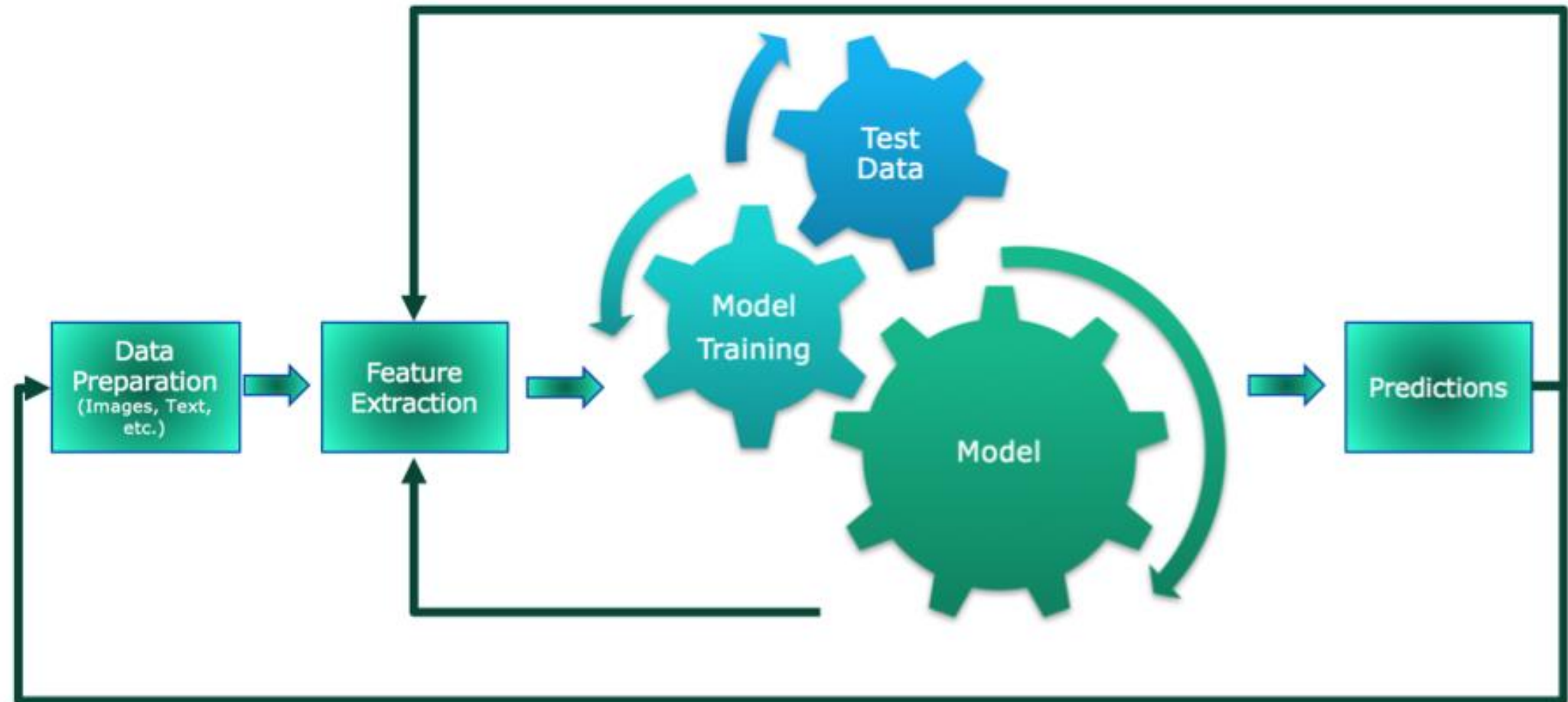
¿qué quiere decir “aprender”?

“Se dice que un programa aprende
de una **experiencia (E)**
con respecto a alguna **tarea (T)**
y **medida de eficacia (P)**
si su eficacia (P) en realizar la tarea (T)
mejora con cada experiencia (E)”

ML Pipeline



A Standard Machine Learning Pipeline



<https://www.datanami.com/2018/09/05/how-to-build-a-better-machine-learning-pipeline/>

Elementos de ML

Experiencia (E) → Datos

Tarea (T) → Objetivo buscado por el humano

Clasificación, Predicción, Detección, Recomendación,
Transformación, Agrupación, ...

Medida de Eficacia (P) → Evaluación de resultados

Modelos de ML

No-Supervisados

Modelos de ML: No-Supervisados

Problema a resolver es usualmente de reducción de dimensionalidad:

- Transformación de datos (continuo)
- Agrupación de datos (discreto)

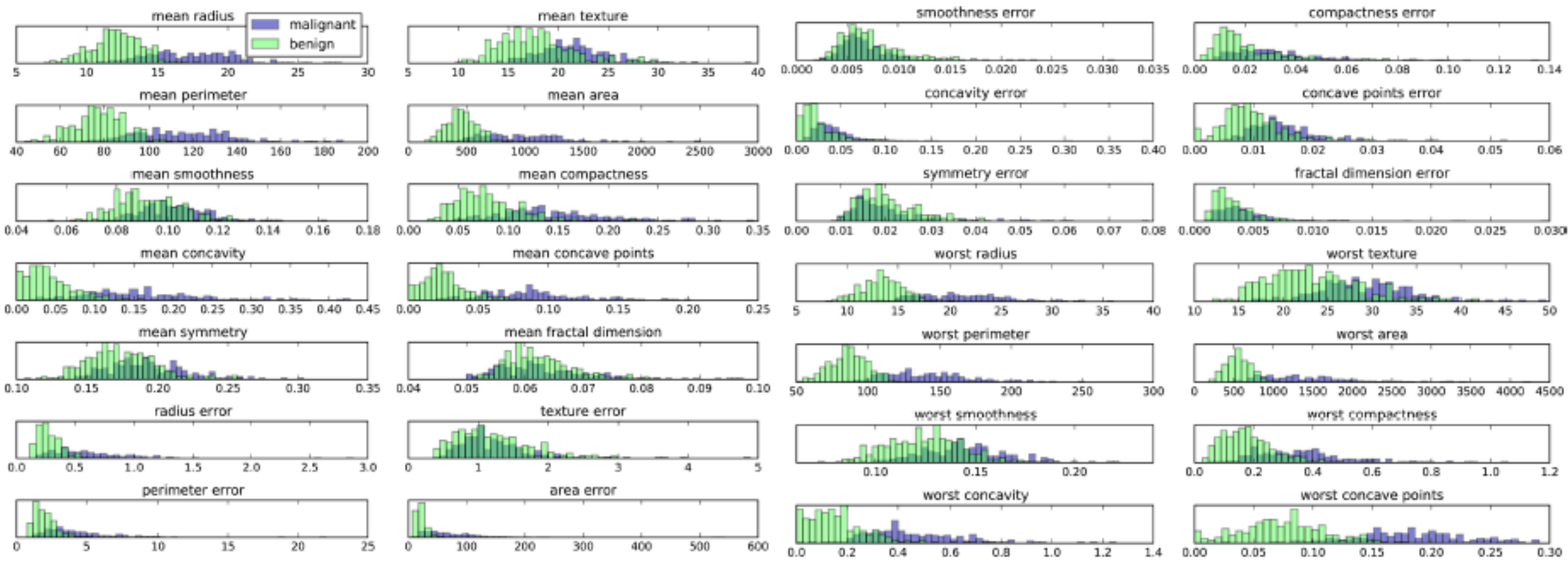
Transformación de datos

- Motivación: visualizar, comprimir información o encontrar una representación más informativa para procesamiento posterior
- Algunos métodos:
 - **PCA** (*principal component analysis*)
 - NMF (*non-negative matrix factorization*)
 - t-SNE (*t-distributed Stochastic Neighbor Embedding*)

PCA

- Método para transformar / agrupar un set de variables y reducir su dimensionalidad
- Sirve cuando tenemos muchas variables que tratan de medir algo, y queremos capturar las más relevantes (las que “aportan más información”) o resumirlas
- También ayuda a visualizar mejor nuestro problema

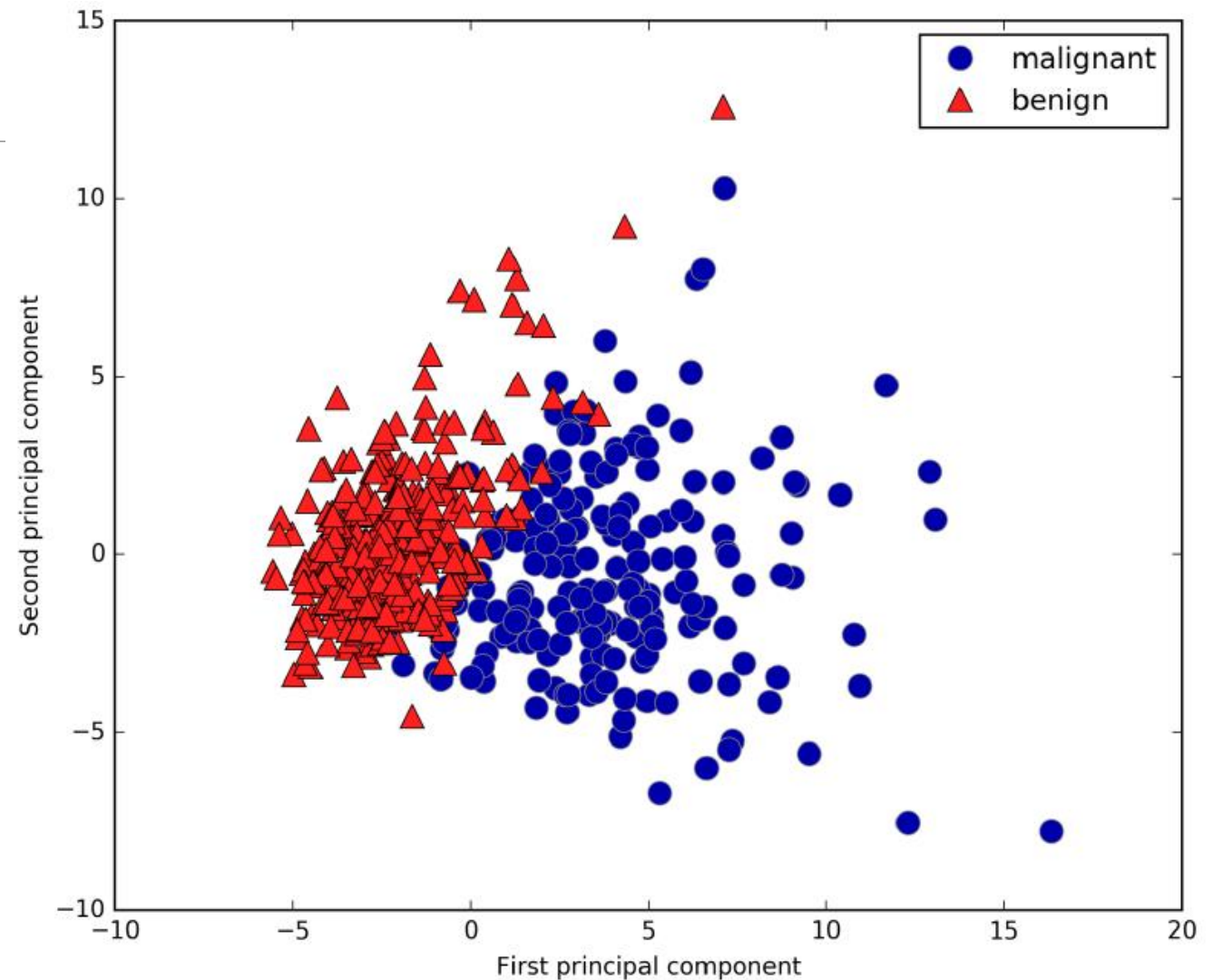
Ejemplo



Ejemplo

```
from sklearn.decomposition import PCA
# keep the first two principal components of the data
pca = PCA(n_components=2)
# fit PCA model to breast cancer data
pca.fit(X_scaled)

# transform data onto the first two principal components
X_pca = pca.transform(X_scaled)
print("Original shape: {}".format(str(X_scaled.shape)))
print("Reduced shape: {}".format(str(X_pca.shape)))
```

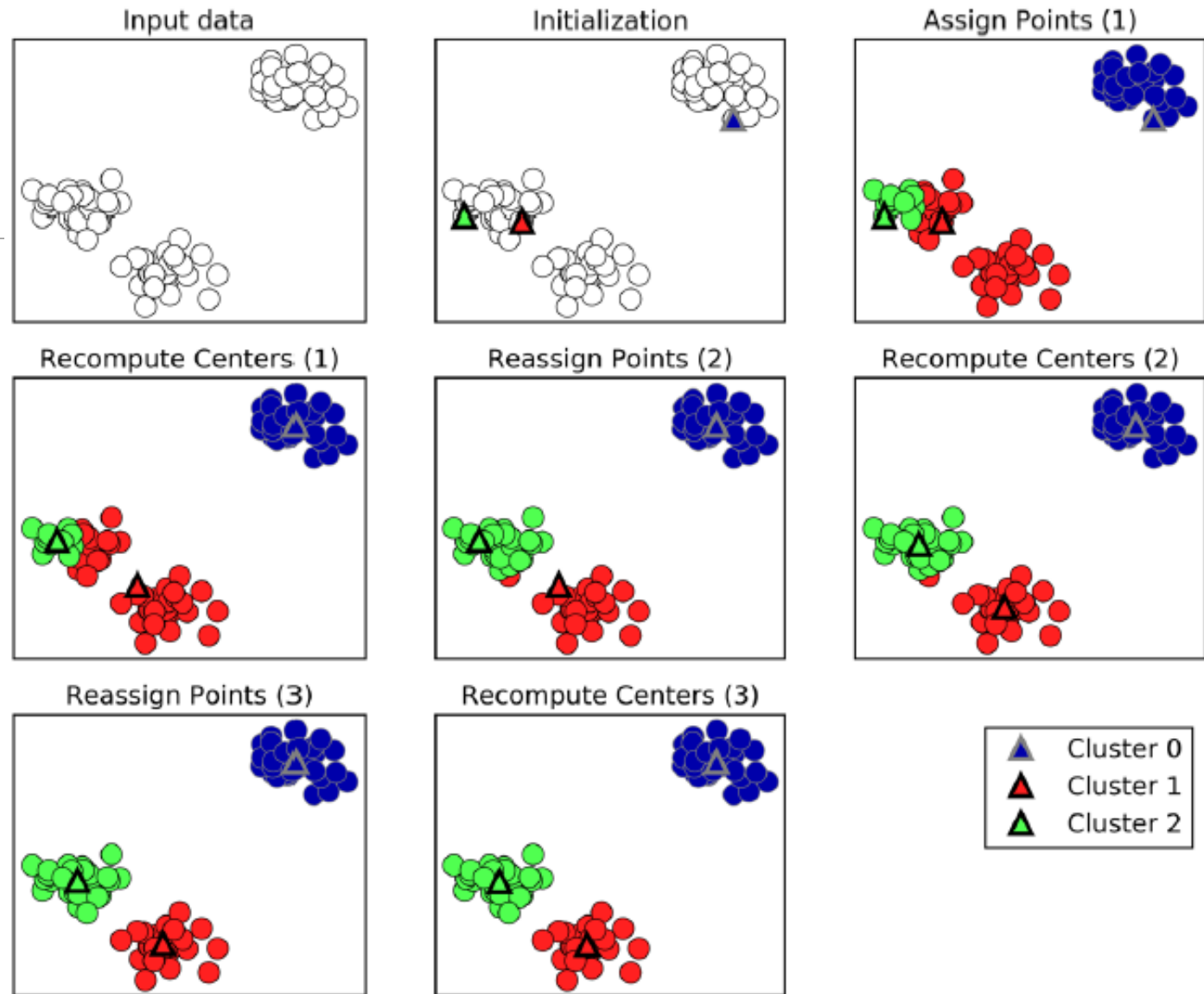


Clustering o agrupación

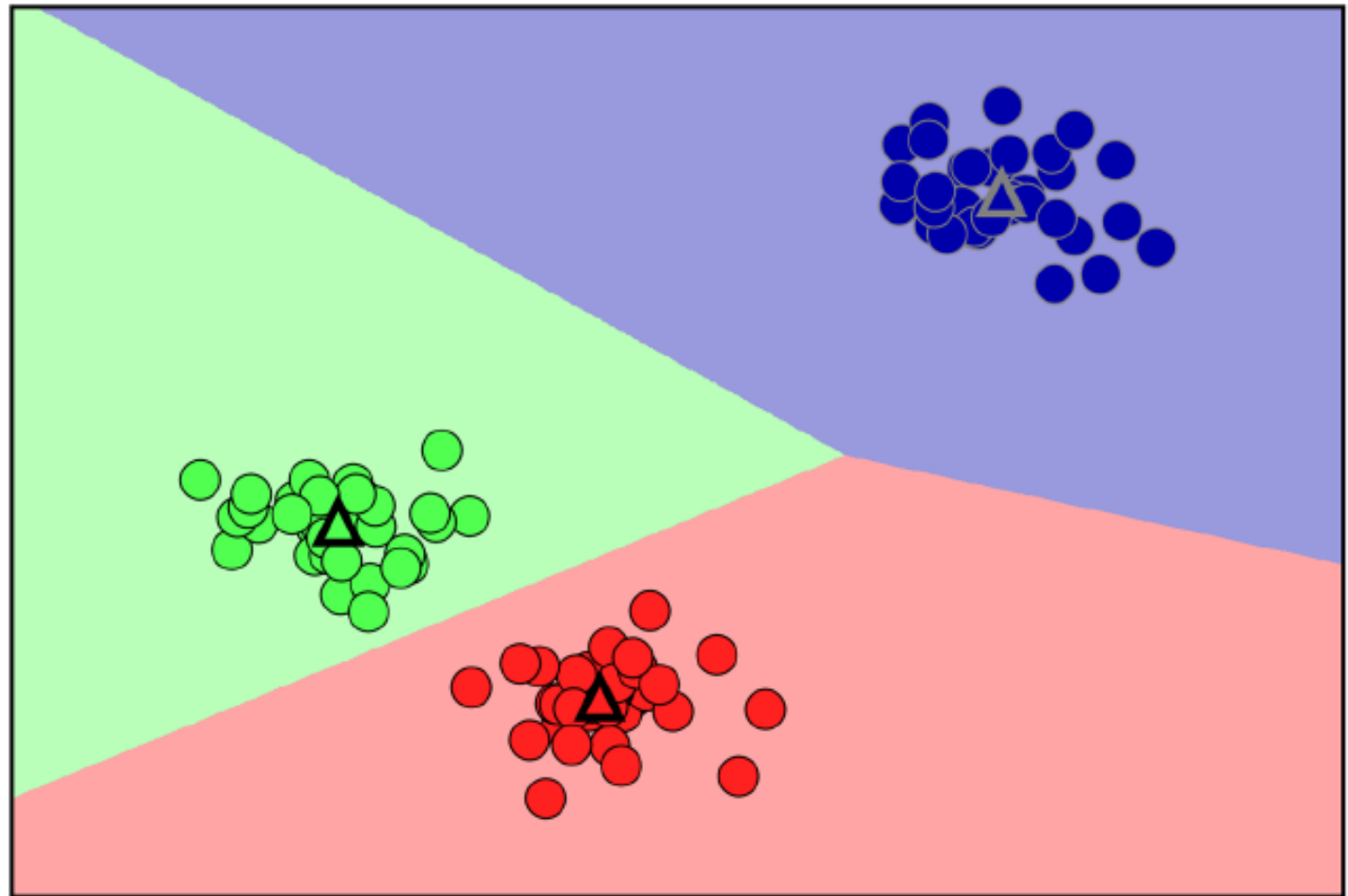
Existen varios algoritmos de agrupación, por ejemplo:

- **K-medias**
- **Agrupamiento jerárquico**
- **DBSCAN**

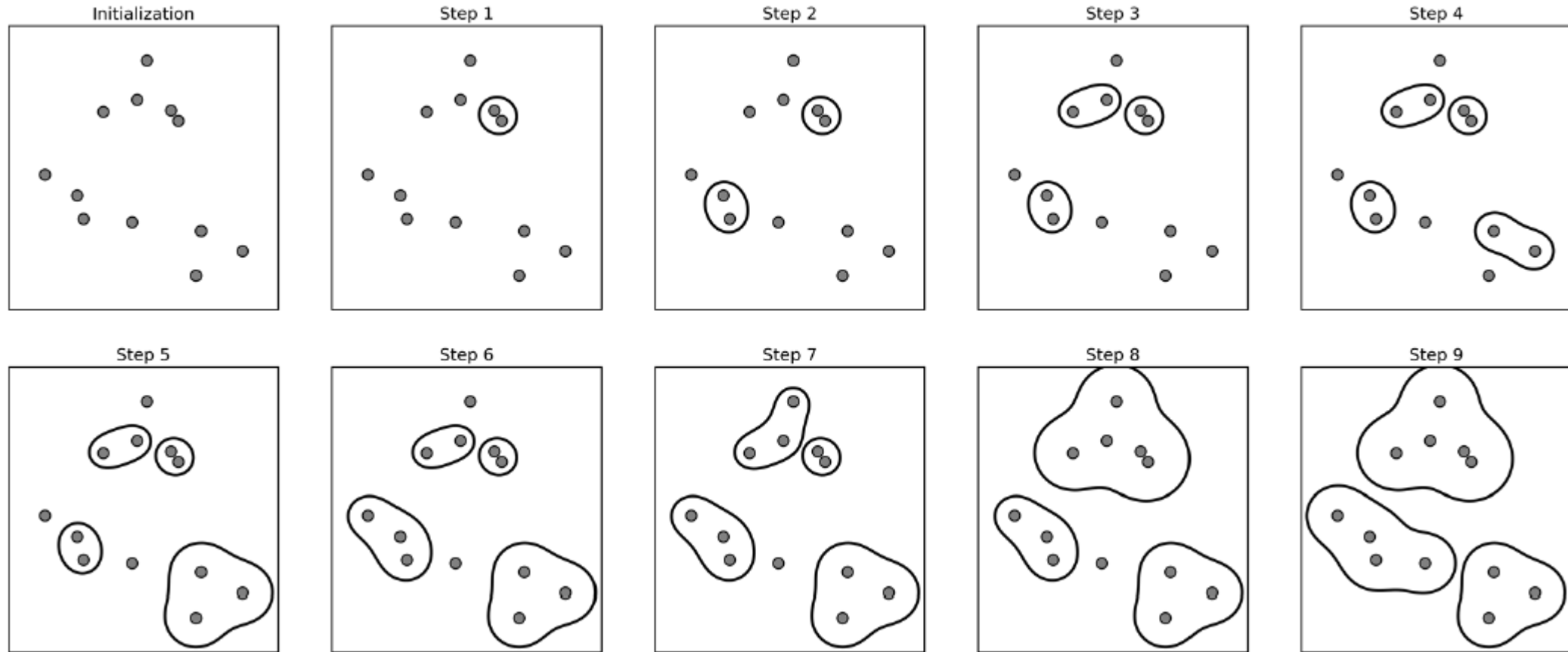
K-medias



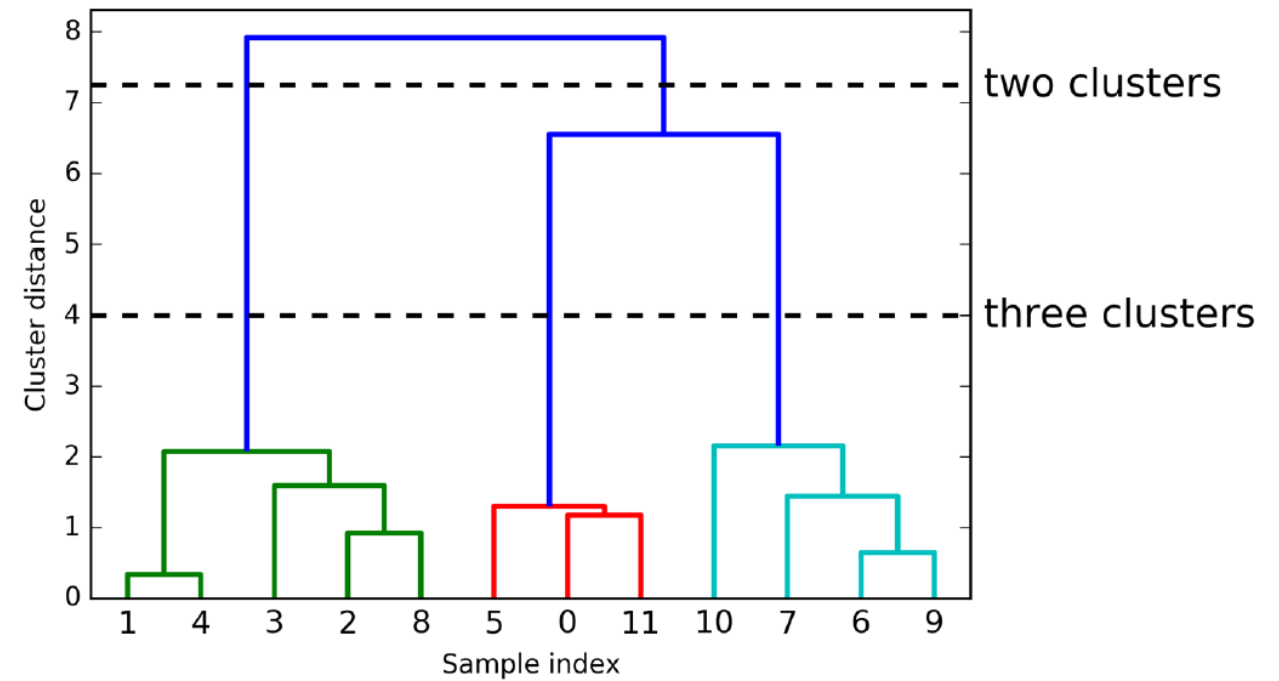
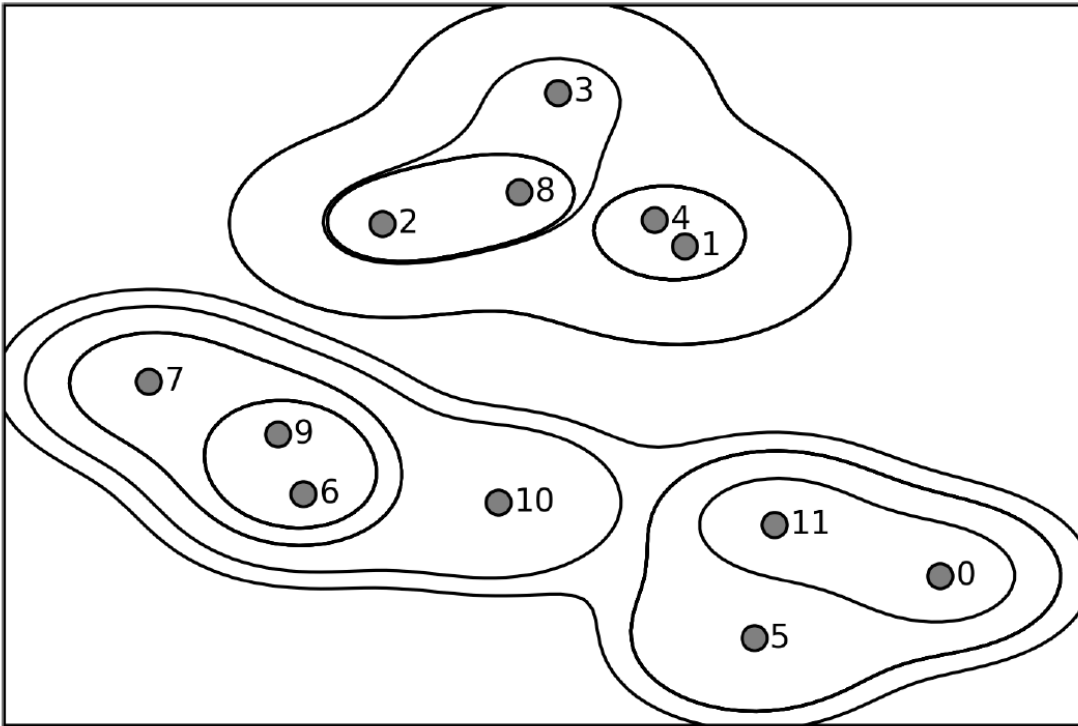
K-medias



Agrupamiento jerárquico



Agrupamiento jerárquico



Representación de Dendrograma