

Intelixencia Artificial

[G4012328]

P3.2. Aprendizaxe non supervisada (Clustering)

Curso 2023 / 2024

3º Grao en Enx. Informática

José María Alonso Moral, Nicolás Vila Blanco

josemaria.alonso.moral@usc.es

<https://citius.gal/team/jose-maria-alonso-moral>

nicolas.vila@usc.es

<https://citius.gal/es/team/nicolas-vila-blanco>

Escola Técnica Superior de Enxeñaría (ETSE)

Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS)

Universidade de Santiago de Compostela (USC)

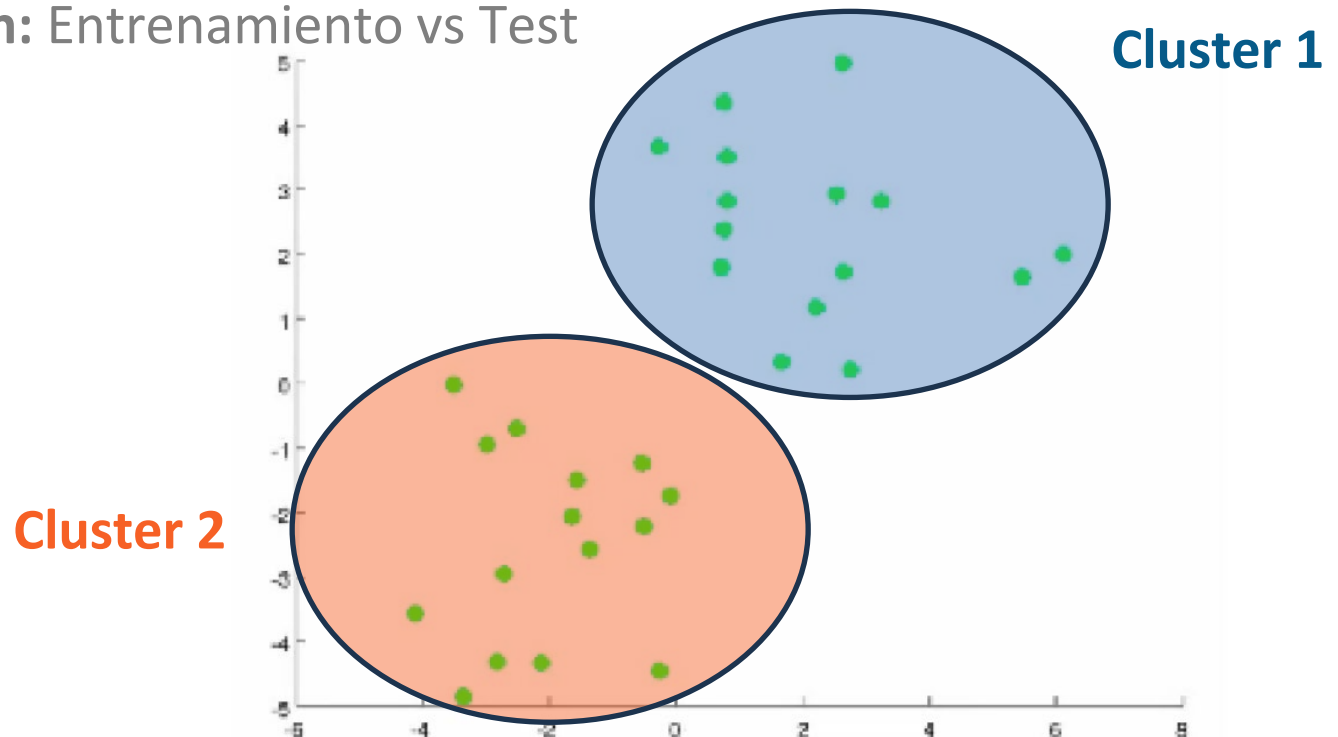
Aprendizaje no Supervisado: Clustering

□ Aprender la función que mejor agrupe los datos disponibles

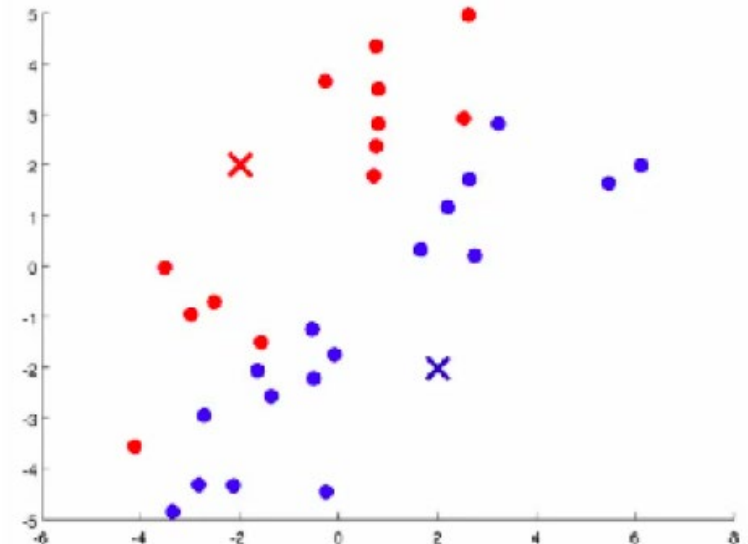
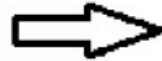
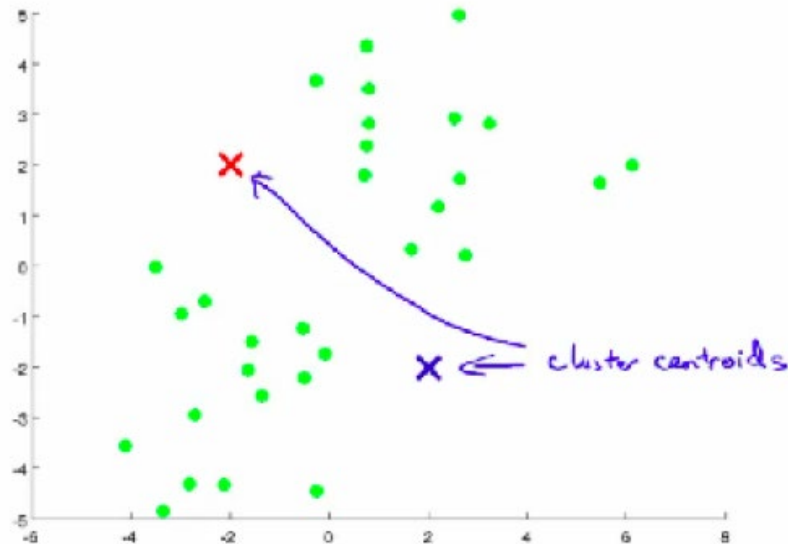
➤ Datos no etiquetados

- ❖ Análisis exploratorio: máxima cohesión intra-cluster y máxima separación extra-clúster
- ❖ El diseñador establece el criterio de agrupación (distancias)

➤ Validación: Entrenamiento vs Test



Aprendizaje no Supervisado: K-medias



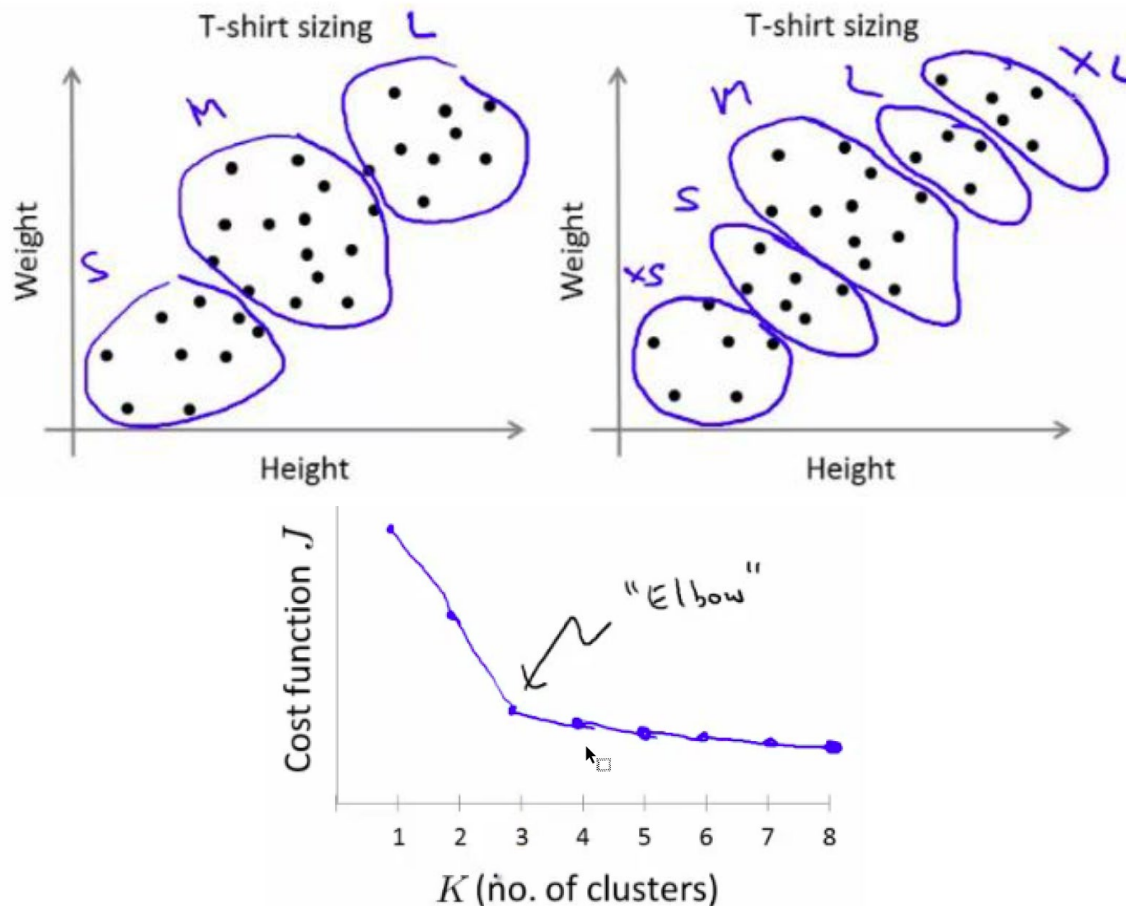
$$J(c^{(1)}, \dots, c^{(m)}, \mu_1, \dots, \mu_K) = \frac{1}{m} \sum_{i=1}^m \|x^{(i)} - \mu_{c^{(i)}}\|^2$$

$$\min_{\substack{c^{(1)}, \dots, c^{(m)}, \\ \mu_1, \dots, \mu_K}} J(c^{(1)}, \dots, c^{(m)}, \mu_1, \dots, \mu_K)$$

Aprendizaje no Supervisado: K-medias

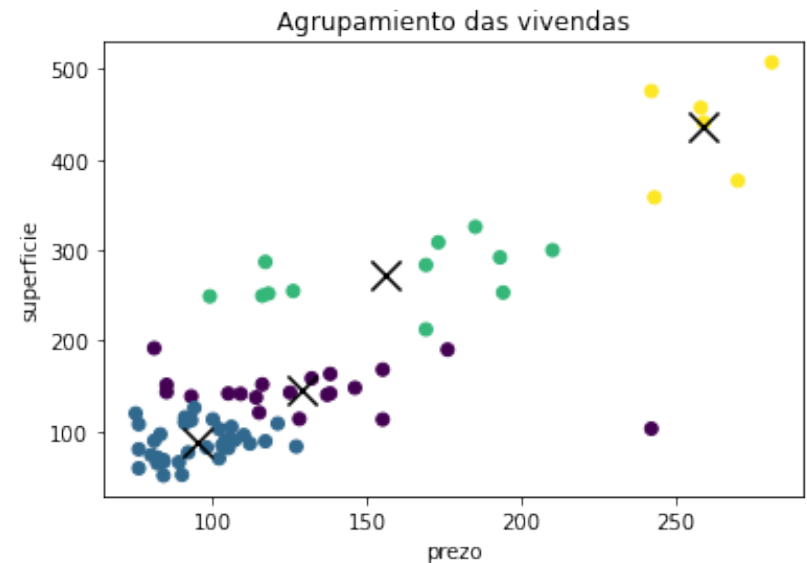
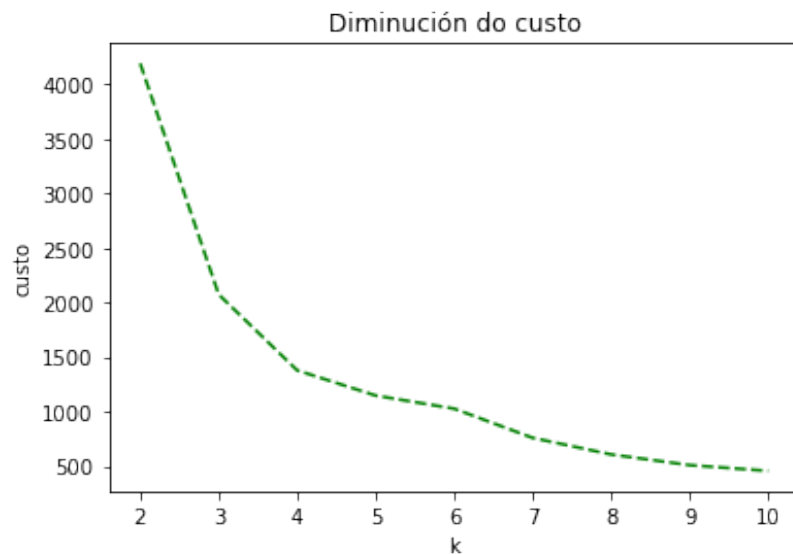
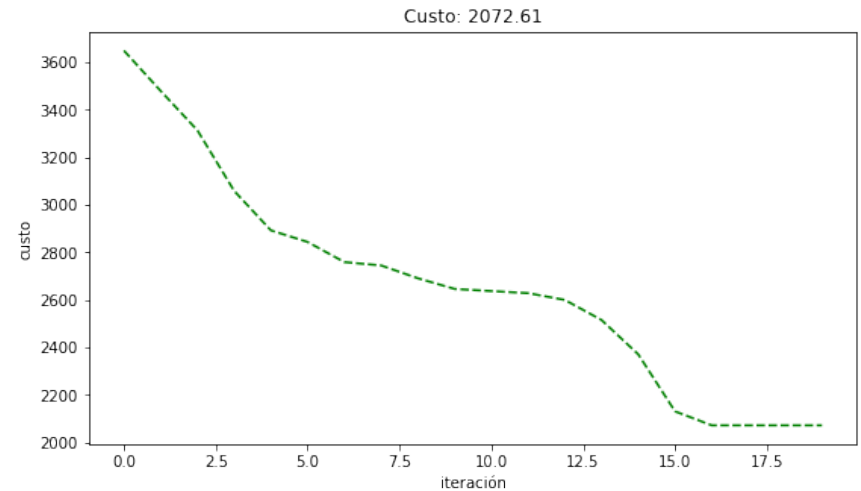
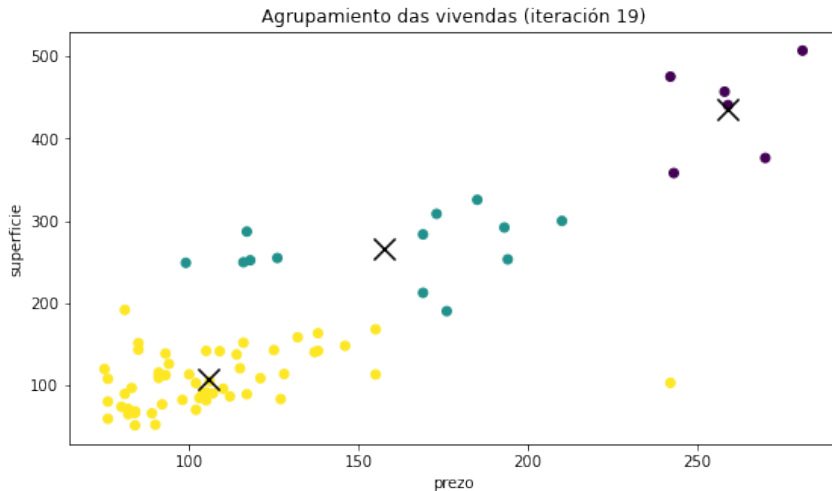
□ ¿Número de Clústers?

- Establecido manualmente a priori (conocimiento experto)
- Calculado dinámicamente (método del codo)




Aprendizaje no Supervisado: Ejemplo

□ Viviendas



Aprendizaje no Supervisado: K-medias (sklearn)

[Install](#) [User Guide](#) [API](#) [Examples](#) [Community](#) [More ▾](#)

[Prev](#) [Up](#) [Next](#)

scikit-learn 1.3.2
[Other versions](#)

Please **cite us** if you use the software.

sklearn.cluster.KMeans
[KMeans](#)
[Examples using sklearn.cluster.KMeans](#)

sklearn.cluster.KMeans

```
class sklearn.cluster.KMeans(n_clusters=8, *, init='k-means++', n_init='warn',
                             max_iter=300, tol=0.0001, verbose=0, random_state=None, copy_x=True,
                             algorithm='lloyd')
```

[\[source\]](#)

K-Means clustering.

Read more in the [User Guide](#).

<https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>

Aprendizaje no Supervisado: Heart Disease

<https://www.kaggle.com/datasets/kingabzpro/heart-disease-patients>

Variables

- **id:** identificador del paciente
- **Age:** edad del paciente
- **Sex:** género del paciente
- **CP:** tipo de dolor en el pecho
- **Trestbps:** presión sanguínea en reposo
- **Chol:** colesterol
- **Fbs:** nivel de azúcar en sangre
- **Restecg:** electrocardiograma
- **Thalach:** máximo ritmo cardiaco
- **Exang:** angina inducida (1=yes, 0=no)
- **Oldpeak:** depression inducida
- **Slope:** pendiente en el pico de ejercicio

Realizar un buen agrupamiento de los pacientes e interpretar los grupos resultantes

Referencias

- ❑ Kaggle (<https://www.kaggle.com/>)
- ❑ Stuart Russell and Peter Norvig, “**Artificial Intelligence: A Modern Approach**”, 4th US Ed.
 - <https://aima.cs.berkeley.edu/>
 - <https://github.com/aimacode/aima-python>
- ❑ **Aula Virtual:** “Aprendizaxe automática / Fundamentos e Regresión Lineal” (Senén Barro Ameneiro)