APRENDIZAJE NO SUPERVISADO

Maestría en Ciencia de Datos

Lucas Fernández Piana

Primavera 2022

Universidad de San Andrés

APRENDIZAJE SUPERVISADO...

EJEMPLO: REGRESIÓN LINEAL

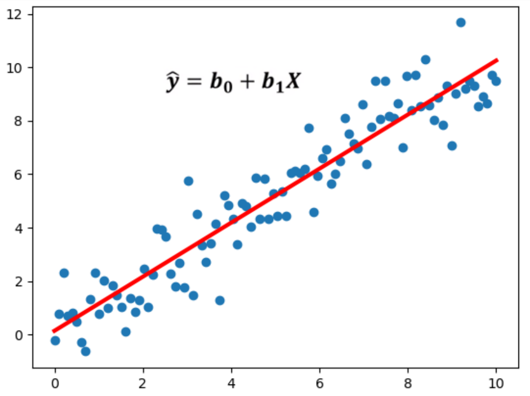


Figura 1: Regresión simple

EJEMPLO: SPLINES

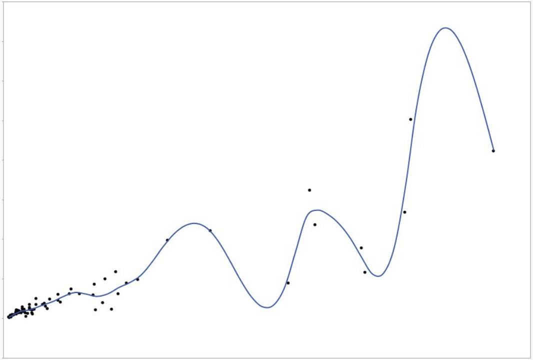


Figura 2: Splines

EJEMPLO: CLASIFICACIÓN

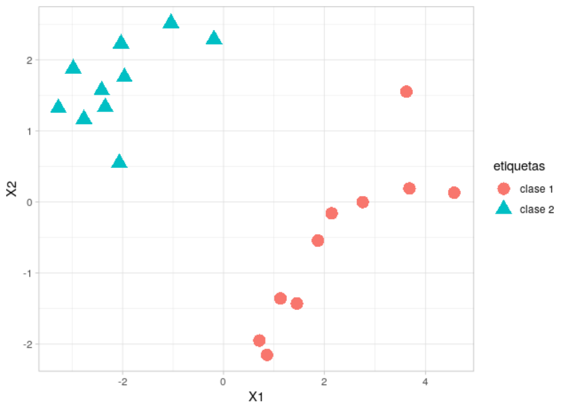


Figura 3: Clasificación binaria.

EJEMPLO: CLASIFICACIÓN

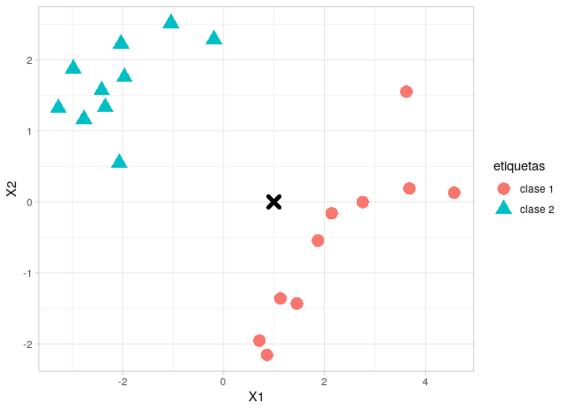


Figura 4: Clasificación binaria.

EJEMPLO: CLASIFICACIÓN BINARIA IMÁGENES

Figura 5: Cat or dog challenge

EJEMPLO: CLASIFICACIÓN BINARIA IMÁGENES

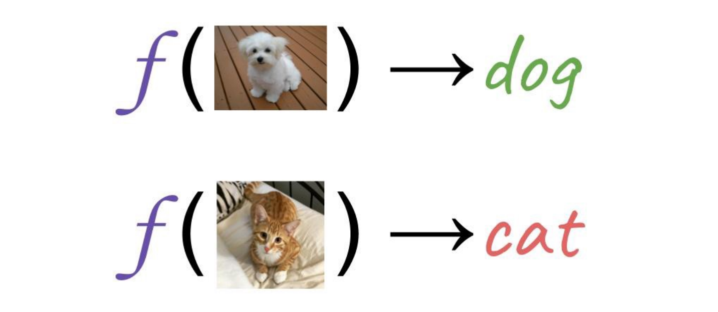


Figura 6: cute algorithm

APRENDIZAJE SUPERVISADO

El aprendizaje supervisado se podría resumir en la siguiente metodología predecir los valores una variable o vector de *respuestas*, $Y = (Y_1, \ldots, Y_l)$, dado un vector predictor $X = (X_1, \ldots, X_p)$.

Dependiendo de la naturaleza de Y podemos determinar las dos clases más conocidas de problemas.

- Y discreto $\longrightarrow$ "**Clasificación**".
- Y continua $\longrightarrow$ "**Regresión**".

Consideremos dos espacio $\mathcal{X}$ (input) y $\mathcal{Y}$ (output). Asumimos que el par $(X, Y) \in \mathcal{X} \times \mathcal{Y}$ son elementos aleatorios (variables, vectores, funciones) con distribución conjunta y desconocida P.

Sea $L : \mathcal{Y} :\to \mathcal{Y}$ decimos que es una *función de pérdida* si cumple:

- $L(y_1, y_2) = 0 \iff y_1 = y_2$.
- L es no negativa.

El objetivo es construir una función $g : \mathcal{X} \to \mathcal{Y}$ tal que $g(X)$ predice a Y. Es decir, El objetivo es construir una función $g : \mathcal{X} \to \mathcal{Y}$ tal que $g(X)$ predice a Y. Es decir, que debe cumplir

$$g(x) = \arg \min_{h:\mathcal{X}\to\mathcal{Y}} E\left[L\left(Y, h(x)\right)|X = x\right].$$

APRENDIZAJE ¡NO! SUPERVISADO...

EJEMPLO: NO SUPERVISADO

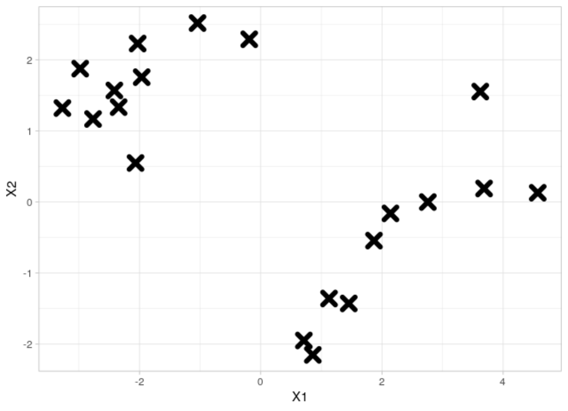


Figura 7: Clustering.

EJEMPLO: NO SUPERVISADO

Figura 8: Imagen satelital

EJEMPLO: NO SUPERVISADO

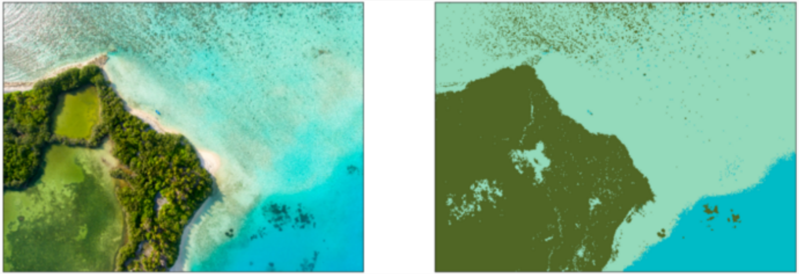


Figura 9: Segmentación de imágenes

NO SUPERVISADO: ¿Con qué no contamos?

Cuando hablamos de aprendizaje no supervisado con qué NO podemos contar,

- El output (Y), es decir, no tenemos respuestas correctas.

NO SUPERVISADO: ¿Con qué no contamos?

Cuando hablamos de aprendizaje no supervisado con qué NO podemos contar,

- El output (Y), es decir, no tenemos respuestas correctas.
- La función de pérdida.

No tenga dudas, es un problema más complicado.

NO SUPERVISADO: ¿Con qué no contamos?

Cuando hablamos de aprendizaje no supervisado con qué NO podemos contar,

- El output (Y), es decir, no tenemos respuestas correctas.
- La función de pérdida.
- En ocaciones ni siquiera una buena definición del problema que queremos resolver.

No tenga dudas, es un problema más complicado.

NO SUPERVISADO: ¿Qué podemos hacer?

Tenemos un conjunto de observaciones $x_1, \ldots, x_n$ que provienen de un elemento aleatorio X que tiene una distribución P_X.

El objetivo es inferir propiedades de P_X que no son los prolemas clásicos de la estadística como aproximar posición o variabilidad.

Nuestro trabajo estará centrado en los siguientes problemas:

- Clustering.

NO SUPERVISADO: ¿Qué podemos hacer?

Tenemos un conjunto de observaciones $x_1, \ldots, x_n$ que provienen de un elemento aleatorio X que tiene una distribución P_X.

El objetivo es inferir propiedades de P_X que no son los prolemas clásicos de la estadística como aproximar posición o variabilidad.

Nuestro trabajo estará centrado en los siguientes problemas:

- Clustering.
- Reducción de dimensión.

NO SUPERVISADO: ¿Qué podemos hacer?

Tenemos un conjunto de observaciones $x_1, \ldots, x_n$ que provienen de un elemento aleatorio X que tiene una distribución P_X.

El objetivo es inferir propiedades de P_X que no son los prolemas clásicos de la estadística como aproximar posición o variabilidad.

Nuestro trabajo estará centrado en los siguientes problemas:

- Clustering.
- Reducción de dimensión.
- Detección de valores atípicos (outliers).

CLUSTERS

CLUSTERS: CONCEPTOS

Dado un conjunto de elementos heterogéneos, los métodos de clustering, son técnicas que tienen como objetivo contruir grupos o clusters más homogéneos.

Se espera que esos nuevos conjuntos estén formados por elementos con características comunes referidas a su naturaleza. Es decir que sean similares entre sí.

CLUSTERS: CONCEPTOS

El enfoque estandar es representar la colección de objetos que tenemos como un conjunto de puntos.

Leitmotiv

Construir una partición del conjunto en clusters tal que puntos en el mismo cluster deben ser “*cercanos*” y puntos en diferente cluster deben estar “*alejados*”.

Todas las técnicas se centran en este objetivo que está vagamente definido.

CLUSTERS: CONCEPTOS

Sea A un conjunto y sean $\mathcal{C} = \{B_1, \dots, B_K\}$ una colección de subconjuntos de A. Decimos que $\mathcal{C}$ es una partición de A, si se cumplen dos condiciones:

- $\mathcal{C}$ es una colección de conjuntos disjuntos, es decir,

$$B_i \cap B_j = \emptyset \quad \forall\, 1 \leq i, j \leq K.$$

- $\mathcal{C}$ es un cubrimiento de A, es decir,

$$A = B_1 \cup B_2 \cup \ \dots \ \cup B_K = \bigcup_{B \in \mathcal{P}} B.$$

Sea A un conjunto y sean $\mathcal{C} = \{B_1, \ldots, B_K\}$ una colección de subconjuntos de A. Decimos que $\mathcal{C}$ es una partición de A, si se cumplen dos condiciones:

- $\mathcal{C}$ es una colección de conjuntos disjuntos, es decir,

$$B_i \cap B_j = \emptyset \quad \forall\, 1 \leq i, j \leq K.$$

- $\mathcal{C}$ es un cubrimiento de A, es decir,

$$A = B_1 \cup B_2 \cup \ \ldots \ \cup B_K = \bigcup_{B \in \mathcal{P}} B.$$

Construir clusters es el arte de construir particiones.

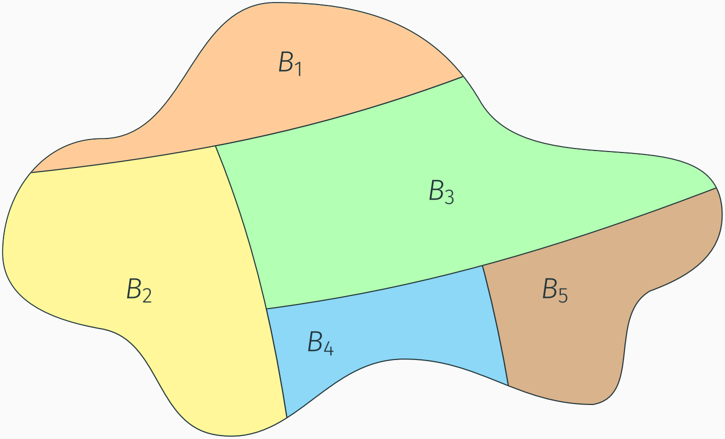
CLUSTERS: CONCEPTOS

Pensemos un ejemplo muy simple, consideremos el conjunto $A = \{1, 2, 3, 4, 7, 8\}$ y construyamos una partición de tres subconjuntos.

- $B_1 = \{1, 2, 4\}$.
- $B_2 = \{3, 7\}$.
- $B_3 = \{8\}$.

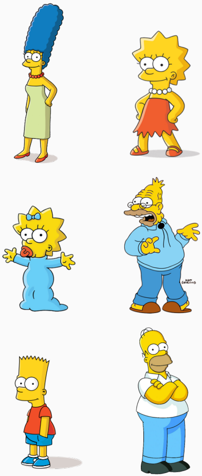
Es claro que $B_1 \cup B_2 \cup B_3 = A$. Además los conjuntos son disjuntos, es decir, no comparten elementos.

CLUSTERS: CONCEPTOS



AGRUPAR LOS SIMPSONS

PARTICIÓN SIMPSONS



Cuadro 1: Familia

Cuadro 2: Escuela

Si tenemos un conjunto de datos $\{x_1, \ldots, x_n\}$ y queremos contruir particiones en relación al leitmotiv de la técnica de clustering vamos a necesitar:

- Entender el espacio ambiente de los datos, es decir, el conjunto más grande que los contiene.
- Formalizar el concepto de similaridad o disimilaridad entre los datos.
- Determinar el número K de particiones convenientes.

En general los conjuntos de datos con los que vamos a trabajar van a estar contenidos en un espacio con alguna noción de métrica o disimilaridad.

En general los conjuntos de datos con los que vamos a trabajar van a estar contenidos en un espacio con alguna noción de métrica o disimilaridad.

Dado un conjunto E, decimos que una función $d : E \times E \to [0, +\infty)$ es una métrica si cumple,

1. $d(x, y) = 0$ si y sólo si $x = y$.
2. $d(x, y) = d(y, x)$ para todo $x, y \in E$.
3. Dados $x, y, z \in E$ se cumple la desigualdad triangular, es decir, $d(x, y) \leq d(x, z) + d(z, y)$.

CLUSTERS: CONCEPTOS

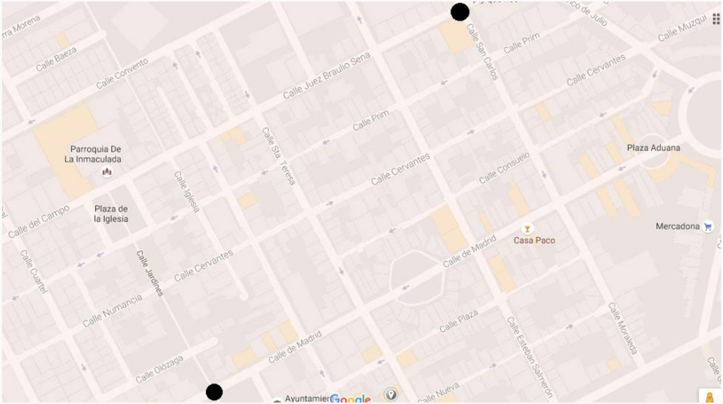
En general los conjuntos de datos con los que vamos a trabajar van a estar contenidos en un espacio con alguna noción de métrica o disimilaridad.

Dado un conjunto E, decimos que una función $d : E \times E \to [0, +\infty)$ es una métrica si cumple,

1. $d(x, y) = 0$ si y sólo si $x = y$.
2. $d(x, y) = d(y, x)$ para todo $x, y \in E$.
3. Dados $x, y, z \in E$ se cumple la desigualdad triangular, es decir, $d(x, y) \leq d(x, z) + d(z, y)$.

Si sólo se cumplen (1) y (2) decimos que d es una disimilaridad.

CLUSTERS: CONCEPTOS



Espacios de dimensión finita.

- $(\mathbb{R}^p, d_2)$, donde $d_2(x, y) = ||x - y||_2 = \sqrt{\sum_{j=1}^{p}(x_j - y_j)^2}$.
- $(\mathbb{R}^p, d_1)$, donde $d_1(x, y) = ||x - y||_1 = \sum_{j=1}^{p} |x_j - y_j|$.
- $(\mathbb{R}^p, d_\infty)$, donde $d_\infty(x, y) = ||x - y||_\infty = \max_{j=1,\dots,p} |x_j - y_j|$.

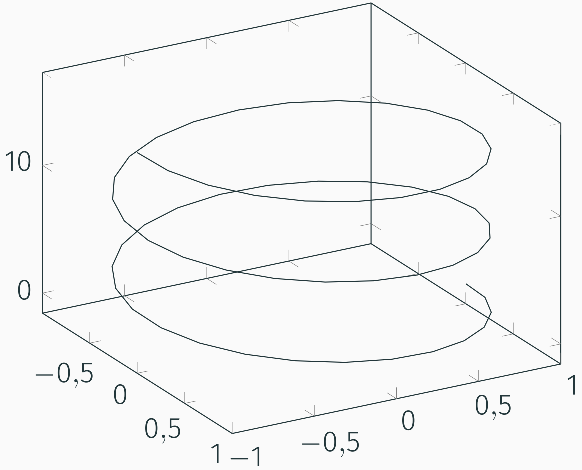
Espacios de dimensión finita.

- $(\mathbb{R}^p, d_2)$, donde $d_2(x, y) = ||x - y||_2 = \sqrt{\sum_{j=1}^{p}(x_j - y_j)^2}$.
- $(\mathbb{R}^p, d_1)$, donde $d_1(x, y) = ||x - y||_1 = \sum_{j=1}^{p} |x_j - y_j|$.
- $(\mathbb{R}^p, d_\infty)$, donde $d_\infty(x, y) = ||x - y||_\infty = \max_{j=1,\ldots,p} |x_j - y_j|$.

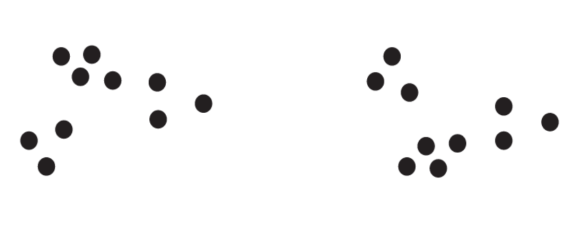
Un ejemplo de un espacio de dimensión infinita,

$$C\left([0, 1], \mathbb{R}^p\right) = \{f : [0, 1] \to \mathbb{R}^3 : f \text{ es continua}\},$$

$$d(f, g) = \sup_{t \in [0,1]} ||f(t) - g(t)||_2.$$


10
0
−0,5
0
0,5
1
−1
−0,5
0
0,5
1

Determinar el número de clusters no es un problema trivial.



CLUSTERS: CONCEPTOS

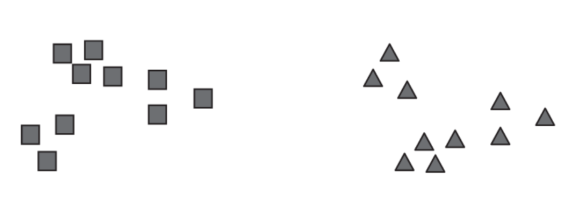


Figura 10: ¿Dos grupos?

Detectar los clusters

Figura 11: ¿Cuatro grupos?

Detectar los clusters

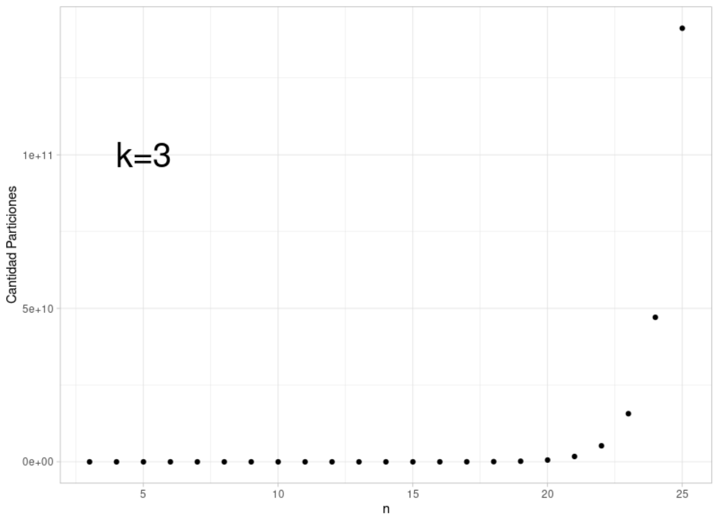
Figura 12: ¿Seis grupos?

Otro problema asosiado a las particiones es la cantidad de particiones que se pueden hacer con pocos datos, aunque sepamos el número de clusters.

Este es un problema muy estudiado por las combinatoria: el número de Stirling que representa la cantidad de particiones no nulas que se pueden obtener con n elementos y k subconjuntos, $S(n, k)$ con $k \leq n$.

$$S(n, k) = S(n - 1, k - 1) + kS(n - 1, k)$$

CLUSTERS: CONCEPTOS



CLUSTERS: CONCEPTOS

A nivel de métodos y algoritmos tenemos enfoques más concretos:

- Centroides.
- Aglomerativos.
- Divisivos.
- Basados en modelos generativos.
- Fundamentados en la teoría de la información.