

Análisis multivariante

Machine Learning

00R Team

2016-17

- 1 Aprendizaje automatizado (ML)
- 2 Clasificación supervisada
- 3 Clasificación no supervisada

Aprendizaje automatizado (ML)

Definiciones

- Conjunto de herramientas para transformar datos en conocimiento.
- Conjunto de técnicas que permitan a las máquinas aprender.
- Conjunto de algoritmos que permitan a la máquina adaptarse a nuevos datos a partir del conocimiento obtenido de datos anteriores.

Pasos para aplicar ML

- Recoger y almacenar datos
- Explorar y preparar los datos
- Entrenar al clasificador con nuestros datos
- Evaluar el desempeño del modelo
- Mejorarlo si fuese necesario

Una vez completados estos pasos, nuestro modelo puede ser implementado en la clasificación de nuevos datos

Clasificación de los algoritmos

- Clasificación supervisada
KNN, Random Forest, SVM, ANN, Naïve Bayes.
- Clasificación no supervisada
K-means, clustering jerárquico

Clasificación supervisada

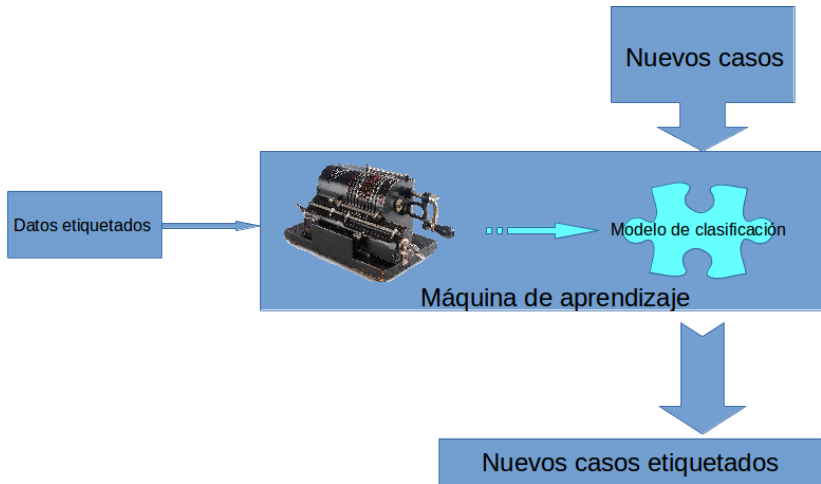


Figura 1

KNN

Introducción

- K-Nearest Neighbours (Los k vecinos más próximos)
- Algoritmo de clasificación supervisada
- Basado en distancias. Generalmente la distancia Euclídea

$$dist(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2}$$

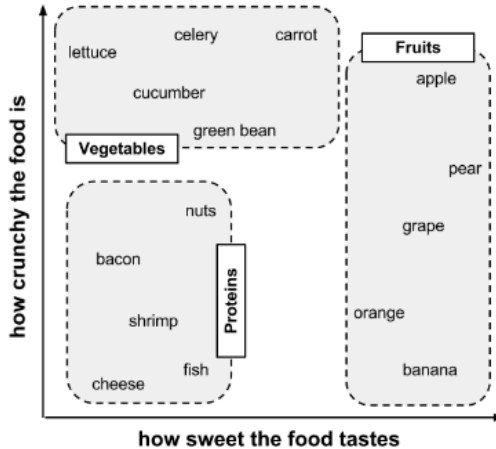


Figura 2

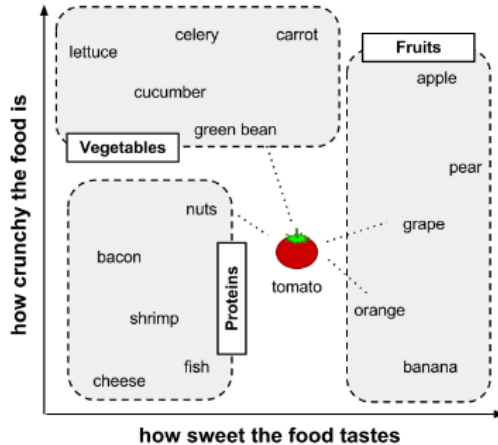


Figura 3

¿Cuántos vecinos?

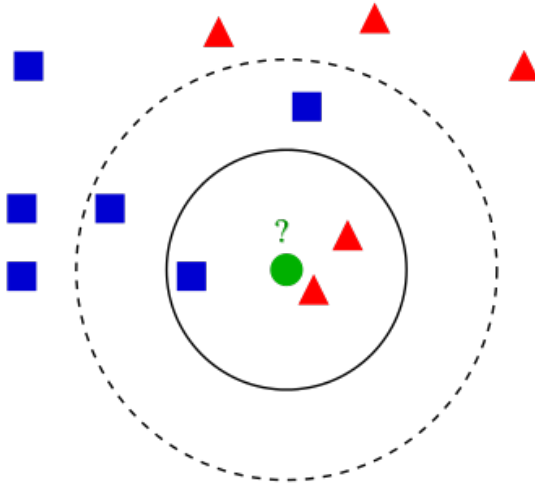


Figura 4

¿Cuántos vecinos?

No hay una única regla:

- $K = n^{\frac{1}{2}}$
- Cross-Validation
- ...

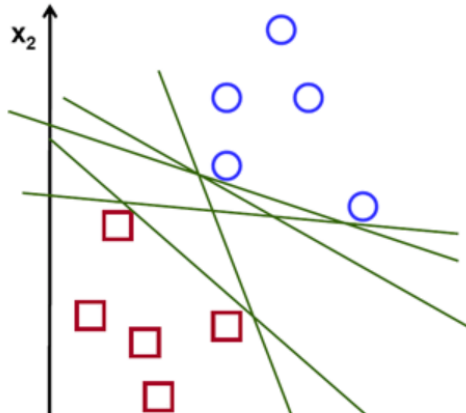
SVM

Introducción

- Conjunto de datos etiquetados
- Construcción de un modelo
- Dado un nuevo dato de etiqueta desconocida
- Modelo es capaz de predecir a que clase pertenece

Introducción

- SVM busca un límite o frontera entre clases \rightarrow Hiperplano
- Separa las diferentes clases en regiones.



¿Qué hiperplano elijo?

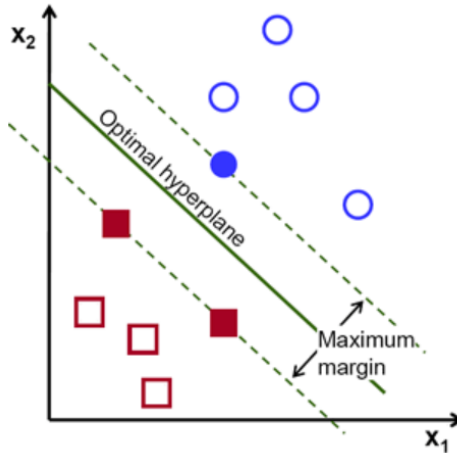


Figura 6

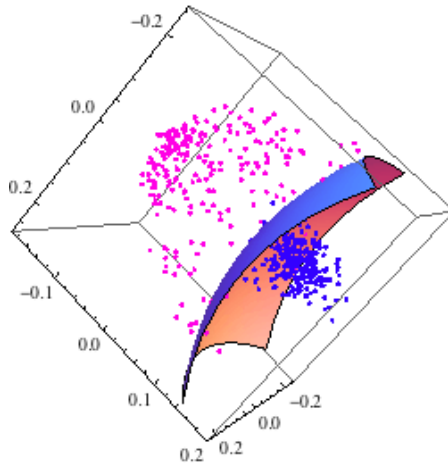


Figura 7

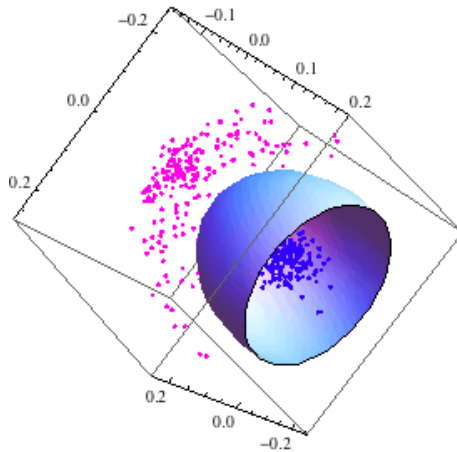
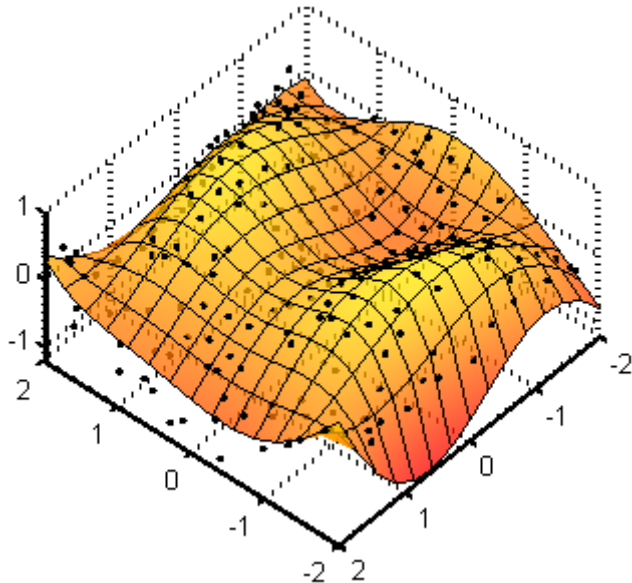


Figura 8



Función kernel

Realiza una transformación sobre los datos para simplificar el aprendizaje

Tipos de funciones Kernel:

- Polinomial: $K(x_i, x_j) = (x_i x_j)^n$
- Perceptrón: $K(x_i, x_j) = \|x_i x_j\|$
- Base radial Gaussiana: $K(x_i, x_j) = e^{-\frac{(x_i - x_j)^2}{2\sigma^2}}$
- Sigmoidal: $K(x_i, x_j) = \tanh(x_i \cdot x_j - \Theta)$

Función kernel

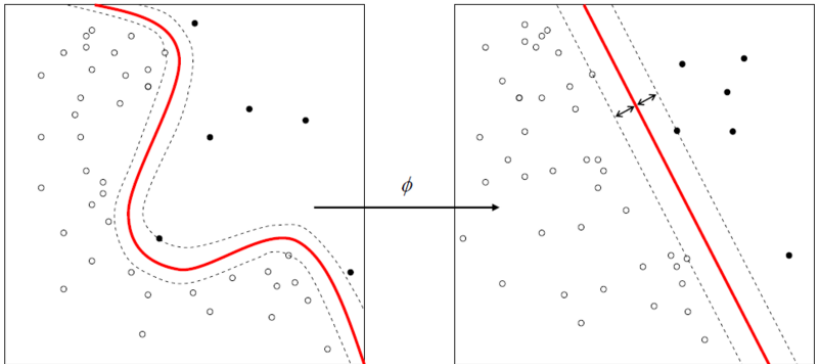


Figura 10

Naïve Bayes

Probabilidad de un evento

$$P(X) = \frac{\textit{Casos favorables}}{\textit{Casos posibles}}$$

Probabilidad de que al tirar un dado salga un 3:

$$P(X = 3) = \frac{1}{6}$$

Probabilidad condicionada

Probabilidad de que suceda un evento cuando ya ha sucedido otro:

$$P(Y | X)$$

Probabilidad de que Y tome un valor determinado cuando X ya ha tomado uno.

Ejemplo: probabilidad de obtener un 10 entre dos dados cuando el primero ha salido 4

$$P(Y = 10 | X = 4)$$

Teorema de Bayes

Basado en el Teorema de Bayes:

$$P(Y | X) = \frac{P(X | Y) P(Y)}{P(X)}$$

En castellano

$$P(\textit{posteriori}) = \frac{P(\textit{probabilidad condicional}) P(\textit{a priori})}{P(\textit{total})}$$

Teorema de bayes

Ejemplo: determinar si jugaremos al tenis en función del tiempo.

Cielo	Temperatura	Humedad	Viento	Jugar
Soleado	Calor	Alta	Si	No
Soleado	Calor	Alta	No	Si
Soleado	Templado	Baja	Si	Si
Nublado	Frio	Baja	No	Si
Nublado	Calor	Normal	No	Si
Nublado	Frio	Alta	Si	No
LLuvia	Templado	Alta	No	No

Probabilidades

¿Jugaremos al tenis (Y) un día nublado, frío, con humedad alta y sin viento (X)?

$$P(Y = Si \mid X_{v1,v2,v3,v4}) = \frac{P(X \mid Y)_{vi} \cdot P(Y)}{P(total)}$$

Probabilidades

$$P(SI_{jugar}) = 4/7 \quad P(NO_{jugar}) = 3/7$$

$$P(nublado \mid SI_{jugar}) = 2/4 \quad P(nublado \mid NO_{jugar}) = 1/3$$

$$P(frio \mid SI_{jugar}) = 1/4 \quad P(frio \mid NO_{jugar}) = 1/3$$

$$P(Alta_{humedad} \mid SI_{jugar}) = 1/4 \quad P(Alta_{humedad} \mid NO_{jugar}) = 3/3$$

$$P(Si_{viento} \mid SI_{jugar}) = 1/4 \quad P(Si_{viento} \mid NO_{jugar}) = 2/3$$

La hora de la verdad

¿Jugaremos?

$$P(Y = Si \mid X_{v1,v2,v3,v4}) = \frac{2/4 \cdot 1/4 \cdot 1/4 \cdot 1/4 \cdot 4/7 = 0,024}{0,024 + 0,21} = 0,103$$

$$P(Y = No \mid X_{v1,v2,v3,v4}) = \frac{1/3 \cdot 1/3 \cdot 3/3 \cdot 2/3 \cdot 3/7 = 0,21}{0,024 + 0,21} = 0,897$$

Redes Neuronales artificiales

Introducción

Paradigma de aprendizaje automatizado inspirado en sistemas nerviosos biológicos

Introducción

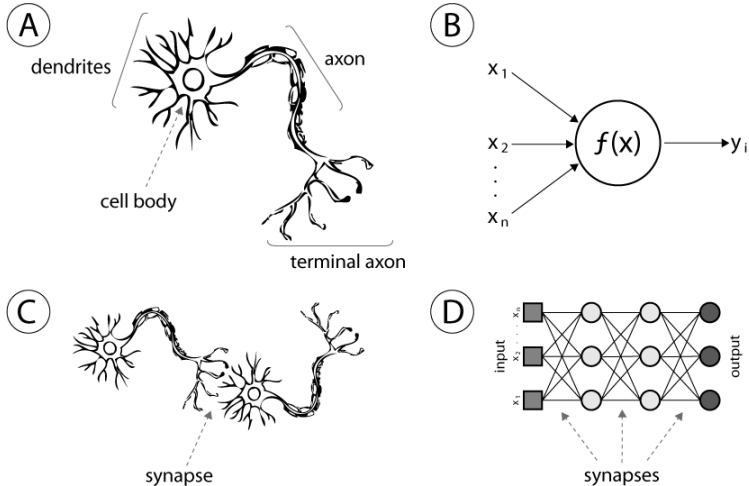
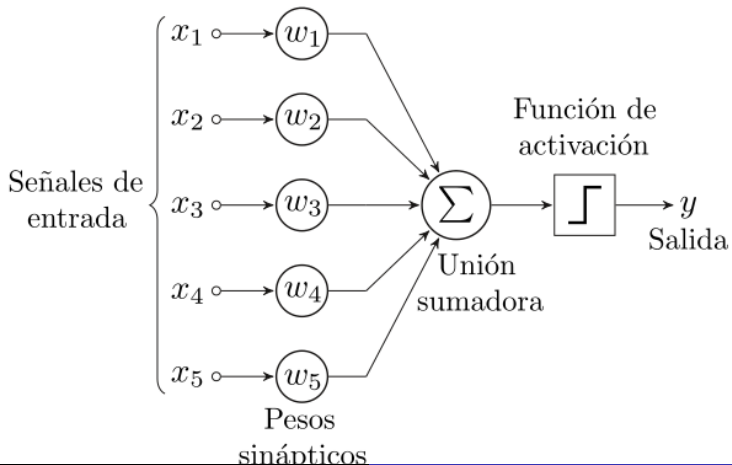


Figura 11

Perceptrón

- La unidad básica es el perceptrón



Componentes

- N entradas, x_1, \dots, x_n
- Cada entrada con un peso x, x_1, \dots, x_n
- Un nodo de entrada extra llamada bias (intercepto)
- Suma de las entradas ponderada por sus pesos:

$$y = \sum x_0 w_0, \dots, x_n w_n$$

- Función de activación p.e.:

$$f_a(x) = 1 \text{ si } y > 0, f_a(x) = -1 \text{ si } x \leq 0$$

Componentes

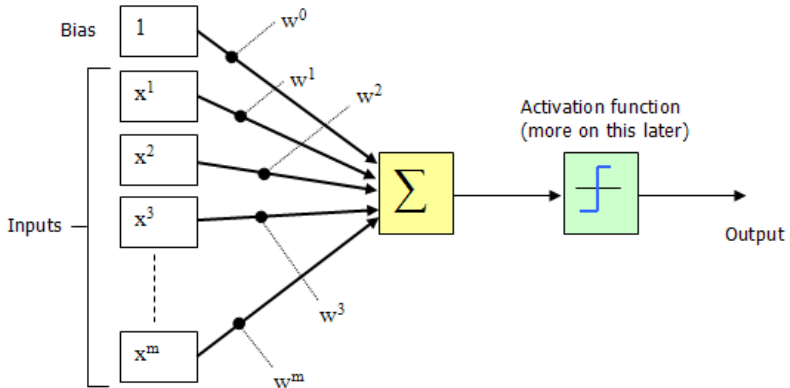


Figura 13

Red neuronal de internet.

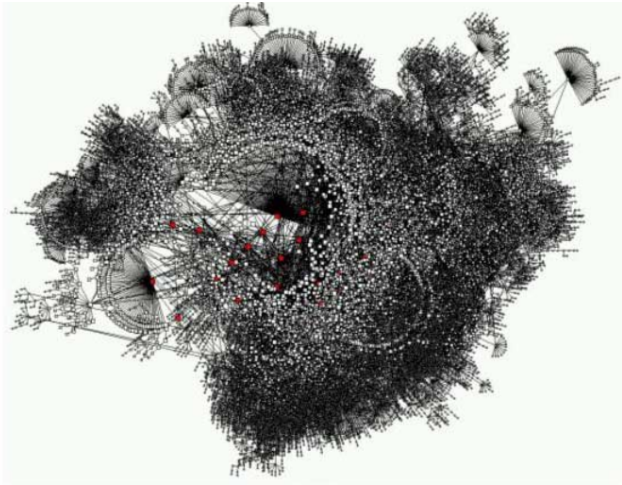


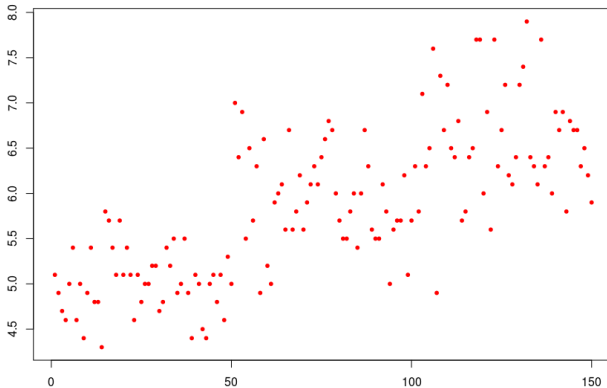
Figura 14

Clasificación no supervisada

K-means

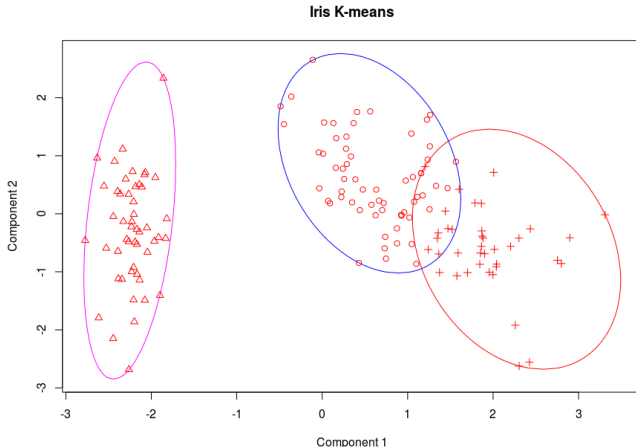
Clasificador no supervisado

- No necesita datos etiquetados



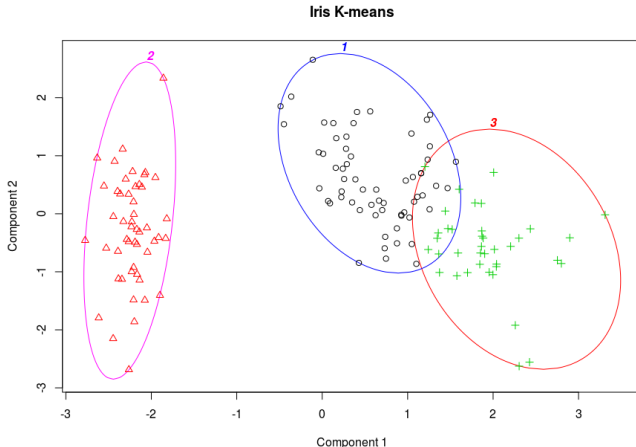
Clasificador no supervisado

- Crea grupos internamente homogéneos



Clasificador no supervisado

- Etiqueta los grupos



Clasificador no supervisado

- Asigna nuevas observaciones al grupo de mayor similitud

