El aprendizaje de métricas de distancia

Trabajo de Fin de Grado

Juan Luis Suárez Díaz

Tutores:
Francisco Herrera Triguero
Salvador García López

Objetivos

- Conocer la disciplina del aprendizaje de métricas de distancia.
- Estudiar los fundamentos matemáticos del aprendizaje de métricas de distancia.
- Analizar los principales algoritmos de aprendizaje de métricas de distancia.
- Desarrollar un software que integre los algoritmos de aprendizaje estudiados.

El aprendizaje de métricas de distancia

¿Qué es?

Es una rama del aprendizaje automático cuya finalidad es apender distancias a partir de los datos.

Definición

Sea X un conjunto no vacío. Una distancia sobre X es una aplicación $d\colon X \times X \to \mathbb{R}$, verificando:

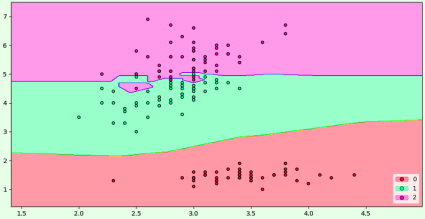
1. $d(x, y) = 0 \iff x = y$ *para cualesquiera* $x, y \in X$ *(coincidencia)*
2. $d(x, y) = d(y, x)$ *para cualesquiera* $x, y \in X$ *(simetría)*
3. $d(x, z) \leq d(x, y) + d(y, z)$ *para cualesquiera* $x, y, z \in X$ *(desigualdad triangular).*

- ***Pseudodistancias:*** *Exigen solo* $d(x, x) = 0$ *en* 1.

¿Por qué distancias?

Ejemplo

Los clasificadores de vecinos cercanos



Problema

- Los algoritmos basados en distancias suelen utilizar distancias fijas.
- **Solución**: Aprender distancias.

¿Cómo aprender una distancia?

Definición (*Distancias de Mahalanobis*)

Sea $M \in \mathscr{M}_d(\mathbb{R})$ *semidefinida positiva. Entonces,* $d_M\colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, *dada por*

$$d(x, y) = \sqrt{(x-y)^T M (x-y)}$$

es una (pseudo-)distancia, denominada distancia de Mahalanobis.

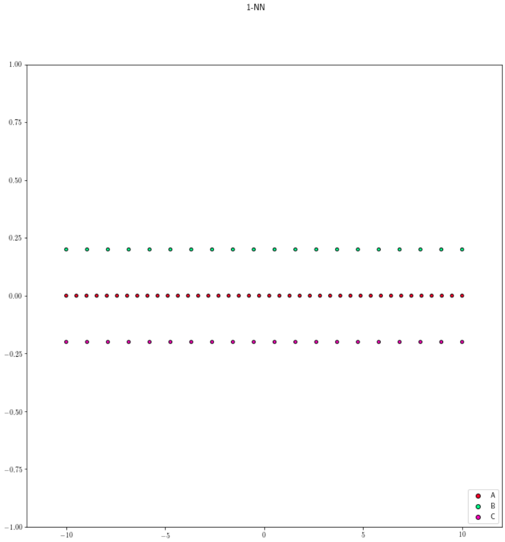
Enfoque principal del aprendizaje de métricas de distancia

Aprender distancias de Mahalanobis sobre espacios vectoriales d-dimensionales.

Dos opciones:

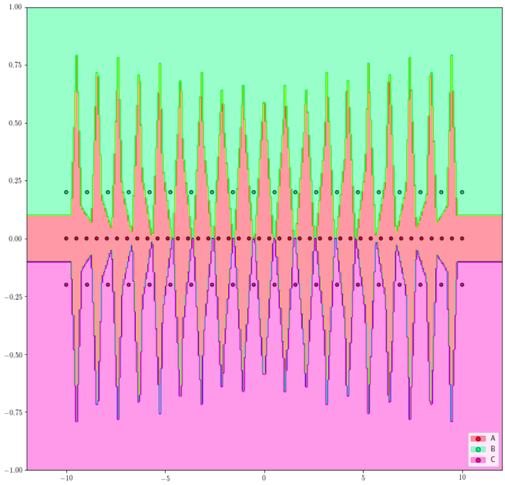
1. Aprender M.
2. Aprender una aplicación lineal L.
   Entonces, $M = L^T L$ y $d_M(x, y)^2 = \|L(x-y)\|_2^2$.

Mejora de clasificadores basados en distancias



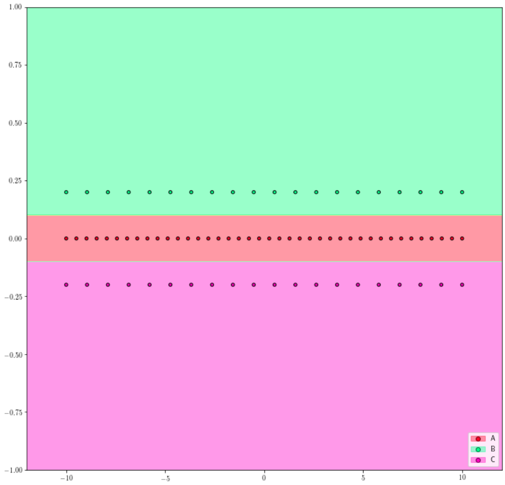
Mejora de clasificadores basados en distancias

$$M = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$



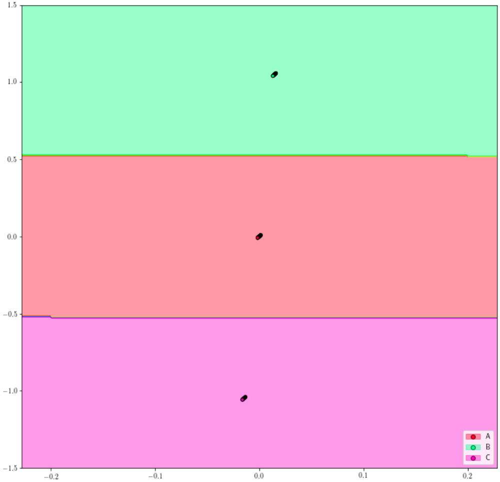
Mejora de clasificadores basados en distancias

$$M \approx \begin{pmatrix} 0 & -0.004 \\ -0.004 & 27.5 \end{pmatrix}$$

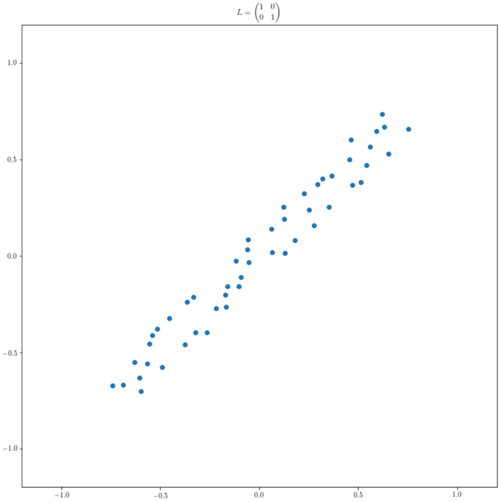


Mejora de clasificadores basados en distancias

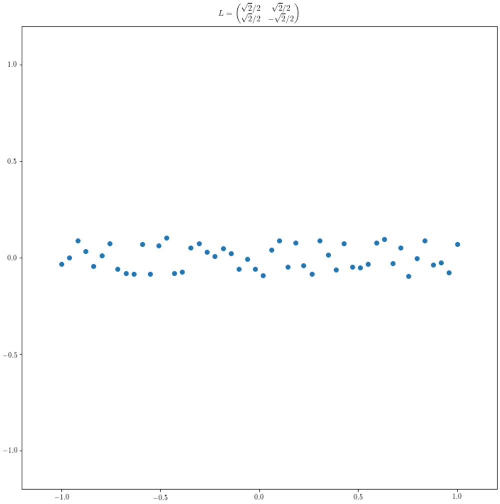
$$L \approx \begin{pmatrix} -0.0001 & 0.073 \\ -0.0008 & 5.24 \end{pmatrix}$$



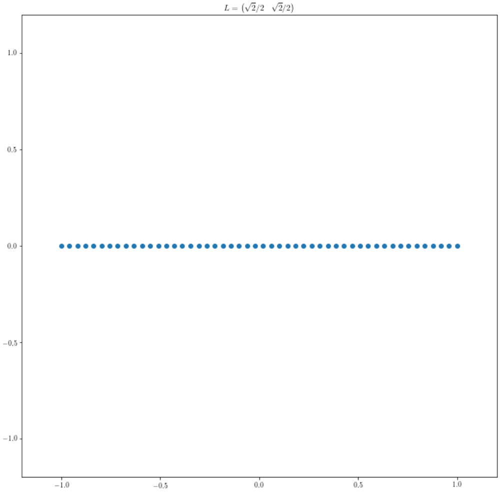
Organización de datos y reducción de dimensionalidad



Organización de datos y reducción de dimensionalidad



Organización de datos y reducción de dimensionalidad



Las matemáticas bajo el aprendizaje de métricas de distancia

1. **Análisis convexo.** De gran importancia en la mayoría de algoritmos de aprendizaje de métricas de distancia.
2. **Análisis matricial.** Las matrices son la herramienta fundamental para modelar el problema.
3. **Teoría de la información.** Presente en algunos de los algoritmos.

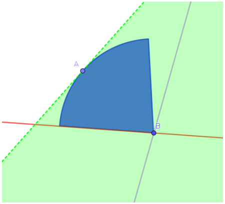
Hiperplanos soporte

Definición (*Hiperplano soporte*)

Sean $T\colon \mathbb{R}^d \to \mathbb{R}$ *lineal,* $\alpha \in \mathbb{R}$ *y* $P = \{x \in \mathbb{R}^d \colon T(x) = \alpha\}$ *hiperplano.*
Definimos $P^+ = \{x \in \mathbb{R}^d \colon T(x) \geq \alpha\}$ *y* $P^- = \{x \in \mathbb{R}^d \colon T(x) \leq \alpha\}$.
P *es un hiperplano soporte para* $K \subset \mathbb{R}^d$ *si* $P \cap \overline{K} \neq \emptyset$ *y* $K \subset P^+$ *o* $K \subset P^-$.

Teorema (*Teorema del hiperplano soporte*)

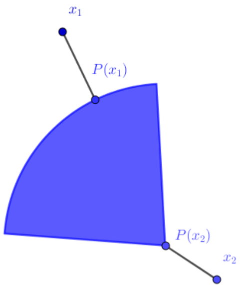
1. *Si* $K \subset \mathbb{R}^d$ *es convexo y cerrado, para cada* $x_0 \in \operatorname{Fr} K$ *existe un hiperplano soporte* P *de* K *tal que* $x_0 \in P$.
2. *Todo conjunto convexo cerrado y propio de* $\mathbb{R}^d$ *es la intersección de todos sus semiespacios soporte.*
3. *Sea* $K \subset \mathbb{R}^d$ *un conjunto cerrado con interior no vacío. Entonces,* K *es convexo si y solo si para todo* $x \in \operatorname{Fr} K$ *existe un hiperplano soporte* P *de* K *con* $x \in P$.



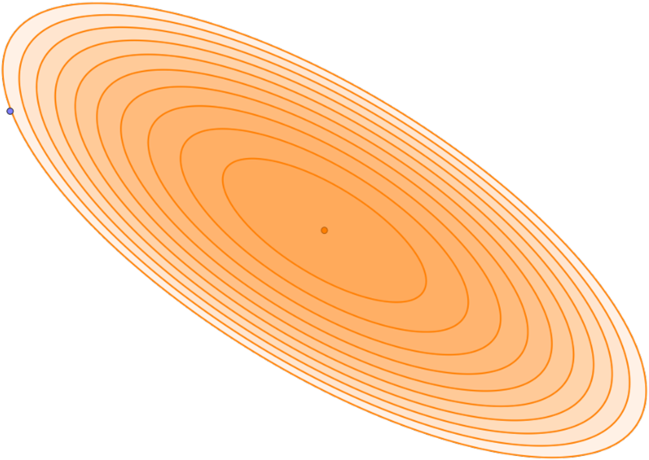
Proyecciones convexas

Teorema (*Proyección convexa*)

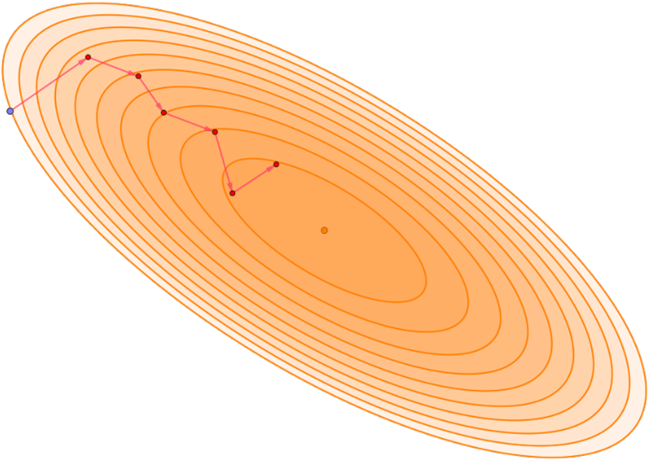
Si $K \subset \mathbb{R}^d$ es no vacío, cerrado y convexo, entonces, para cada $x \in \mathbb{R}^d$ existe un único punto $P_K(x) \in K$ tal que $d(x, K) = d(x, P_K(x))$: la proyección convexa de x sobre K.



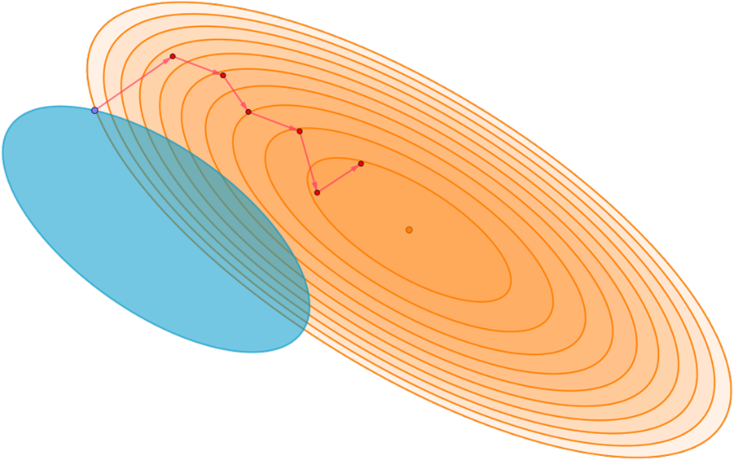
Métodos de optimización: gradiente descendente y gradiente con proyecciones



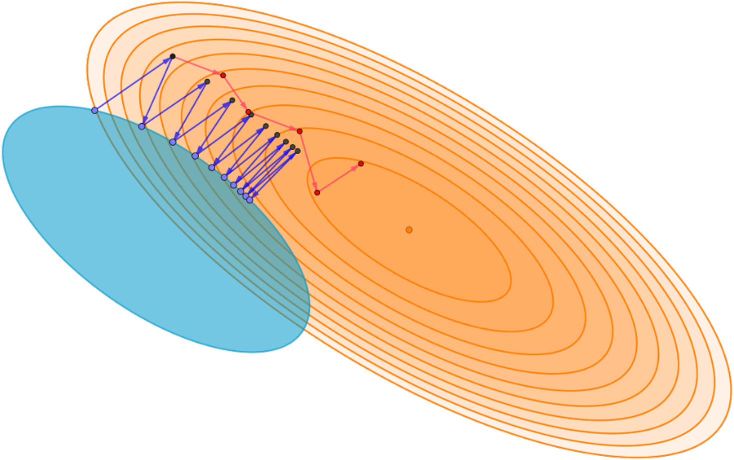
Métodos de optimización: gradiente descendente y gradiente con proyecciones



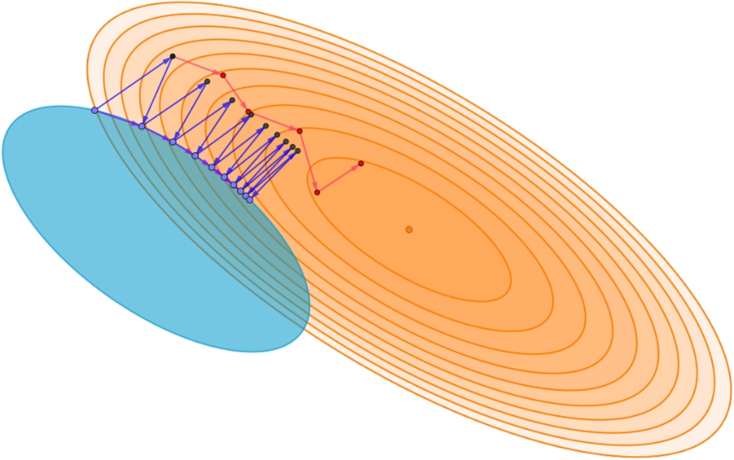
Métodos de optimización: gradiente descendente y gradiente con proyecciones



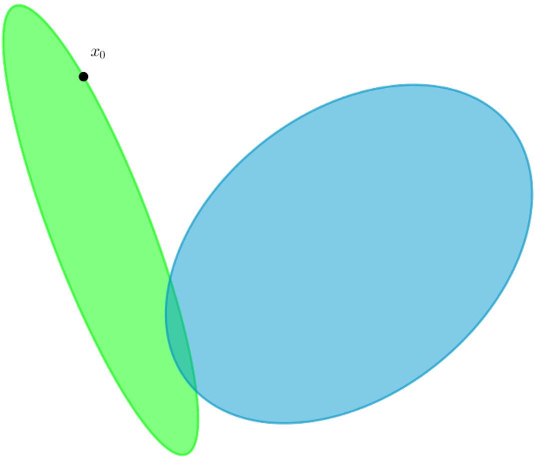
Métodos de optimización: gradiente descendente y gradiente con proyecciones



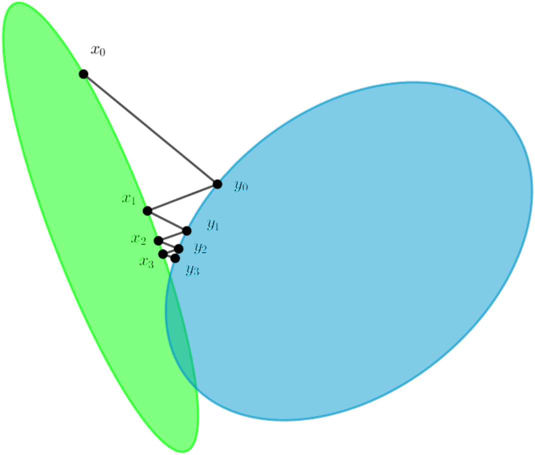
Métodos de optimización: gradiente descendente y gradiente con proyecciones



Método de las proyecciones iteradas



Método de las proyecciones iteradas



¡Gracias por su atención!

Juan Luis Suárez Díaz
jlsuarezdiaz@correo.ugr.es