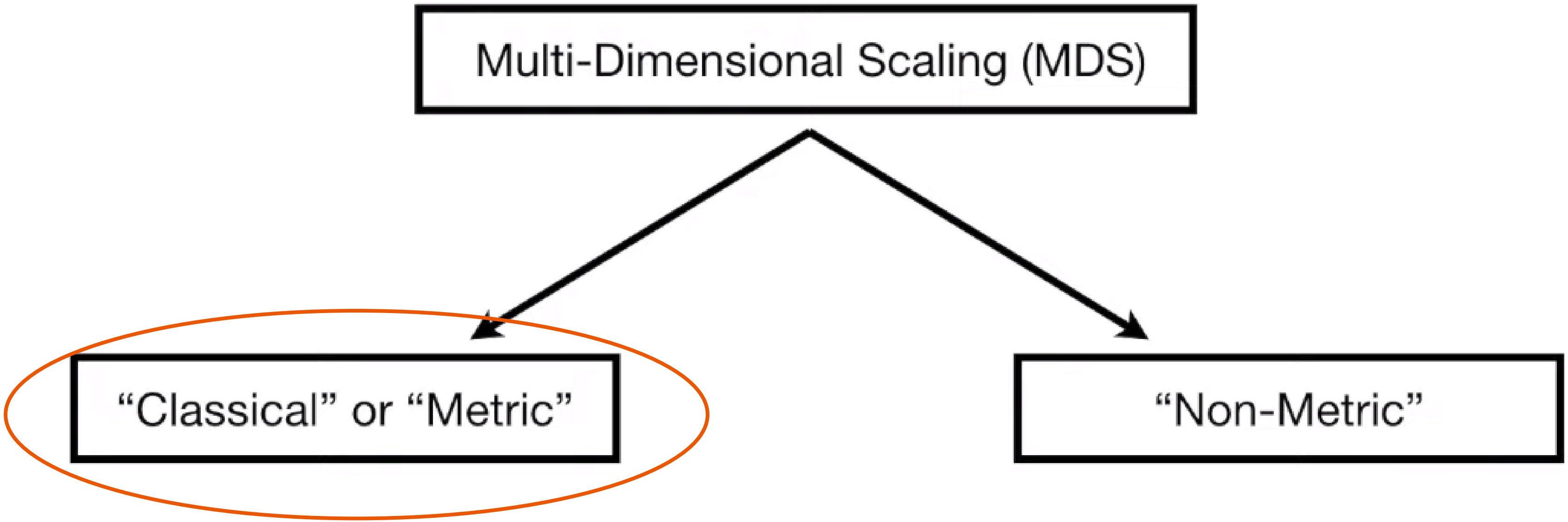
MDS: Multidimensional Scaling
y
PCoA: Principal Coordinate Analysis

Ing. Juan M. Rodríguez

MDS: Multi-Dimensional Scaling



Classical **MDS** es exactamente igual a **PCoA**: (*Principal Coordinate Analysis*)

MDS: Multi-Dimensional Scaling

- Idea: Preservar las distancias entre puntos

- Ubicar los puntos en una dimensión menor tal que las distancias se parezcan lo más posible:

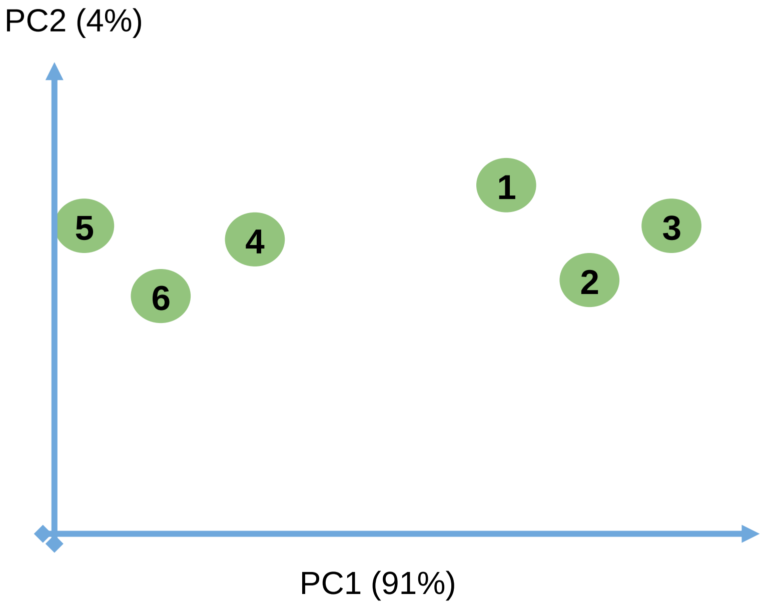
$$\min_{\mathbf{z}} \{(\sum_{i \neq j=1,\ldots,n} (D(x^{(i)}, x^{(i)}) - D(z^{(i)} - z^{(j)}))^2\}$$

MDS: Multi-Dimensional Scaling

Retomando el ejemplo de PCA

Puntaje por tema en un examen de ciencia de datos

Puntos	A 1	A 2	A 3	A 4	A 5	A 6
árboles	9	10	8	3	2	1
RNA	6	4	5	3	2.8	1
PLN	12	9	10	2.5	1.3	2
SVM	5	7	6	2	4	7



MDS: Multi-Dimensional Scaling

Puntaje por tema en un examen de ciencia de datos

Puntos	A 1	A 2	A 3	A 4	…
árboles	9	10	8	3	…
RNA	6	4	5	3	…
PLN	12	9	10	2.5	…
SVM	5	7	6	2	…

Para hacer MDS (o PCoA):

1 Calculamos la distancia entre el A1 y A2

2 Calculamos la distancia entre el A1 y A3

3 Calculamos la distancia entre el A1 y A4

4 Calculamos la distancia entre el A2 y A3

etc…

MDS: Multi-Dimensional Scaling

Puntaje por tema en un examen de ciencia de datos

Puntos	A 1	A 2	A 3	A 4	…
árboles	9	10	8	3	…
RNA	6	4	5	3	…
PLN	12	9	10	2.5	…
SVM	5	7	6	2	…

La distancia calculada puede ser cualquiera, por ejemplo la euclidiana, en ese caso la distancia entre A1 y A2 es:

$$A1 - A2 = \sqrt{(9-10)^2 + (6-4)^2 + (12-9)^2 + (5-7)^2}$$

$$A1 - A2 = 4{,}24$$

MDS: Multi-Dimensional Scaling

Se crea una matriz de distancias

	A 1	A 2	A 3	A 4	…
A1	**0**	4.24	5	3	…
A2	4.24	**0**	6	3.4	…
A3	5	6	**0**	5.6	…
A4	3	3.4	5.6	**0**	…

Ahora tengo que encontrar una nueva matriz **M** con la misma cantidad de columnas (es decir ejemplos) pero menos filas, 2 (según las dimensiones que quiera visualizar).

Puntos	A 1	A 2	A 3	A 4	…
MDS1	?	?	?	?	…
MDS2	?	?	?	?	…

Para esta nueva Matriz voy a calcular las distancias. Generando una segunda matriz de distancias, que deberá ser lo más parecida a esta:

MDS: Multi-Dimensional Scaling

Puntos	A 1	A 2	A 3	A 4	…
MDS1	x_1	x_2	x_3	x_4	…
MDS2	y_1	y_2	y_3	y_4	…

$$A1 - A2 = \sqrt{(x_1-x_2)^2 + (y_1 - y_2)^2} \quad \approxeq 4.24$$

$$A1 - A3 = \sqrt{(x_1-x_3)^2 + (y_1 - y_3)^2} \quad \approxeq 5$$

¡Queremos preservar las distancias!

	A 1	A 2	A 3	A 4	…
A1	**0**	4.24	5	3	…
A2	4.24	**0**	6	3.4	…
A3	5	6	**0**	5.6	…
A4	3	3.4	5.6	**0**	…

Tenemos que llegar a esta matriz

MDS: Multi-Dimensional Scaling

¿Cómo resolvemos este problema?

Puntos	A 1	A 2	A 3	A 4	…
MDS1	x_1	x_2	x_3	x_4	…
MDS2	y_1	y_2	y_3	y_4	…

A1 - A2 = $\sqrt{(x_1-x_2)^2 + (y_1 - y_2)^2}$ ≊ 4.24

A1 - A3 = $\sqrt{(x_1-x_3)^2 + (y_1 - y_3)^2}$ ≊ 5

¡Queremos preservar las distancias!

Llamemos:

A1 = $(x_1, y_1,) = \vec{x}_1$

A2 = $(x_2, y_2,) = \vec{x}_2$

A3 = $(x_3, y_3,) = \vec{x}_3$

etc.

MDS: Multi-Dimensional Scaling

Definimos en base a lo anterior una nueva función llamada: ***Stress***

$$Stress_D(\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_N) = \sqrt{\underbrace{\sum_{i \neq j = 1,\ldots,N} (d_{ij} - ||\vec{x}_i - \vec{x}_j||)^2}}$$

Nuestro objetivo ahora será minimizar esta función

$x_1, \ldots, x_N$
Son cada uno de los vectores, que representan a A1, A2, etc. en el nuevo espacio

d_{ij} es la distancia original calculada

$||x_i - x_j||$ es la nueva distancia que deberíamos encontrar

Si esto es 0 la distancia es la misma en ambos espacios. Por ello debemos minimizar este valor, de hecho la suma de todas estas restas.

MDS: Multi-Dimensional Scaling

Para encontrar los valores que minimicen la función de ***Stress*** hay varios algoritmos matemáticos, como:

- Método de descenso empinado de **Kruskal** (*Kruskal's steepest descent method*): un método de descenso por gradiente
- Método iterativo de mayorización de **De Leeuw** (*De Leeuw's iterative majorization method*), también llamado **SMACOF**

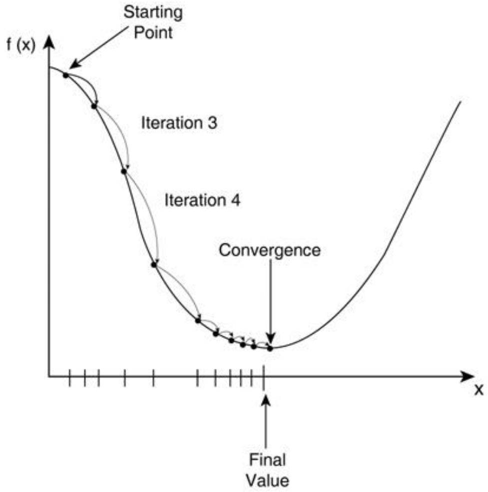
Una cosa importante a tener en cuenta es que estos métodos son enfoques iterativos, que a veces dan resultados diferentes

Descenso por gradiente

MDS: Descenso por gradiente

La técnica de **descenso por gradiente** es una técnica genérica, utilizada en muchos algoritmos, la podemos usar en regresión lineal para ajustar la mejor recta, en PCA y por supuesto es una parte fundamental del algoritmo ***backpropagation*** utilizado para el entrenamiento de redes neurales

Se trata de encontrar el mínimo de la función de error, utilizando para ello el gradiente como brújula. Ya que el gradiente apunta siempre en la dirección de máximo crecimiento, ir en la dirección opuesta nos acercará al mínimo.



MDS: Descenso por gradiente

$$\sqrt{\sum_{i \neq j=1,\ldots,N} \left(d_{ij} - ||x_i - x_j||\right)^2}$$

Veamos cómo podemos minimizar esta función

Elijamos valores aleatorios para el nuevo espacio MDS

Puntos	A 1	A 2	A 3	A 4
MDS1	10	8	4	9
MDS2	5	9	6	2

NOTA: asumimos solo 4 observaciones: A1..A4, por simplicidad.

Las distancias son las calculadas en la matriz

Idealmente debería dar cero.
Eso quiere decir que las distancias son iguales

$$\sqrt{(4.24 - ||A1-A2||)^2 + (5 - ||A1-A3||)^2 + (3 - ||A1-A4||)^2 + (6 - ||A2-A3||)^2 + (3.4 - ||A2-A4||)^2 + (5.6 - ||A3-A4||)^2} = 0$$

MDS: Descenso por gradiente

$$Stress = \sqrt{(4.24 - \| A1-A2 \|)^2 + (5 - \| A1-A3 \|)^2 + (3 - \| A1-A4 \|)^2 + (6 - \| A2-A3 \|)^2 + (3.4 - \| A2-A4 \|)^2 + (5.6 - \| A3-A4 \|)^2}$$

Puntos	A 1	A 2	A 3	A 4
MDS1	10	8	4	9
MDS2	5	9	6	2

$\| A1 - A2 \| = \sqrt{(10-8)^2 + (5-9)^2} =$ **4,47**

$\| A1 - A3 \| = \sqrt{(10-4)^2 + (5-6)^2} =$ **6,08**

$\| A1 - A4 \| = \sqrt{(10-9)^2 + (5-2)^2} =$ **3,16**

$\| A2 - A3 \| = \sqrt{(8-4)^2 + (9-6)^2} =$ **5**

$\| A2 - A4 \| = \sqrt{(8-9)^2 + (9-2)^2} =$ **7.1**

$\| A3 - A4 \| = \sqrt{(4-9)^2 + (6-2)^2} =$ **6.4**

Queríamos 0 y obtuvimos 4.05. Eso quiere decir que nuestro error es:

Error = 0 - 4.05 = - 4.05
|Error| = | 0 - 4.05| = 4.05

$$\sqrt{(4.24 - 4.47)^2 + (5 - 6.08)^2 + (3 - 3.16)^2 + (6 - 5)^2 + (3.4 - 7.1)^2 + (5.6 - 6.4)^2} = 4{,}05$$

MDS: Descenso por gradiente

Calculemos la función **Error** paramétrica. En función de cada variable independiente. Y llamemos a cada una de las componentes de A1.. A4, como sub MDS1 y sub MDS2, por ejemplo $A1 = (A1_{MDS1}, A1_{MDS2})$

$$\mathbf{Err(A1_{MDS1}, A1_{MDS2}, \ldots, A4_{MDS1}, A4_{MDS2})} = \sqrt{\begin{aligned} &(\mathbf{4.24} - \sqrt{(A1_{MDS1}-A2_{MDS1})^2 + (A1_{MDS2}-A2_{MDS2})^2})^2 + \\ &(\mathbf{5} - \sqrt{(A1_{MDS1}-A3_{MDS1})^2 + (A1_{MDS2}-A3_{MDS2})^2})^2 + \\ &(\mathbf{3} - \sqrt{(A1_{MDS1}-A4_{MDS1})^2 + (A1_{MDS2}-A4_{MDS2})^2})^2 + \\ &(\mathbf{6} - \sqrt{(A2_{MDS1}-A3_{MDS1})^2 + (A2_{MDS2}-A3_{MDS1})^2})^2 + \\ &(\mathbf{3.4} - \sqrt{(A2_{MDS1}-A4_{MDS1})^2 + (A2_{MDS2}-A4_{MDS2})^2})^2 + \\ &(\mathbf{5.6} - \sqrt{(A3_{MDS1}-A4_{MDS1})^2 + (A3_{MDS2}-A4_{MDS2})^2})^2 \end{aligned}}$$

Necesitamos calcular el gradiente de la función de error para ver hacia donde crece este y "avanzar" en dirección opuesta

$$\vec{\nabla} Err = \left(\frac{\partial Err}{\partial A1_{MDS1}}, \frac{\partial Err}{\partial A1_{MDS2}}, \frac{\partial Err}{\partial A2_{MDS1}}, \frac{\partial Err}{\partial A2_{MDS2}}, \frac{\partial Err}{\partial A3_{MDS1}}, \frac{\partial Err}{\partial A3_{MDS2}}, \frac{\partial Err}{\partial A4_{MDS1}}, \frac{\partial Err}{\partial A4_{MDS2}}\right)$$

MDS: Descenso por gradiente

Hagamos el cálculo solo para A1 y su primer componente y veamos cómo nos ayuda a mejorar el resultado

Todos los términos donde no aparece $A1_{MDS1}$ los reemplazamos por sus valores numéricos

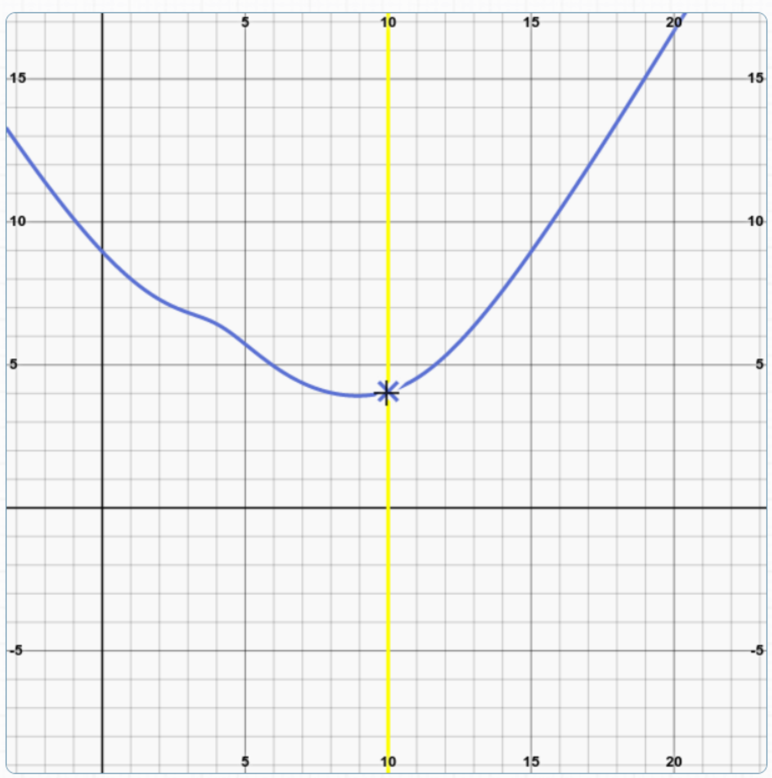
$$\frac{\partial Err}{\partial A1_{MDS1}} = \frac{\partial \sqrt{\left(\mathbf{4.24} - \sqrt{(A1_{MDS1} - 8)^2 + 16}\right)^2 + \left(\mathbf{5} - \sqrt{(A1_{MDS1} - 4)^2 + 1}\right)^2 + \left(\mathbf{3} - \sqrt{(A1_{MDS1} - 9)^2 + 9}\right)^2 + 15{,}33}}{\partial A1_{MDS1}}$$

Si llamamos a $A1_{MDS1}$, "x" por simplicidad obtendremos la siguiente fórmula para la derivada

$$\frac{\partial Err}{\partial A1_{MDS1}} = \frac{-\frac{2\left(5-\sqrt{(x-4)^2+1}\right)(x-4)}{\sqrt{(x-4)^2+1}} - \frac{2\left(\frac{106}{25}-\sqrt{(x-8)^2+16}\right)(x-8)}{\sqrt{(x-8)^2+16}} - \frac{2\left(3-\sqrt{(x-9)^2+9}\right)(x-9)}{\sqrt{(x-9)^2+9}}}{2\sqrt{\left(5-\sqrt{(x-4)^2+1}\right)^2 + \left(\frac{106}{25}-\sqrt{(x-8)^2+16}\right)^2 + \left(3-\sqrt{(x-9)^2+9}\right)^2 + \frac{1533}{100}}}$$

MDS: Descenso por gradiente

Esta es la función de error, para $A1_{MDS1}$
El valor actual es 10, y vemos que baja si nos movemos a la izquierda.



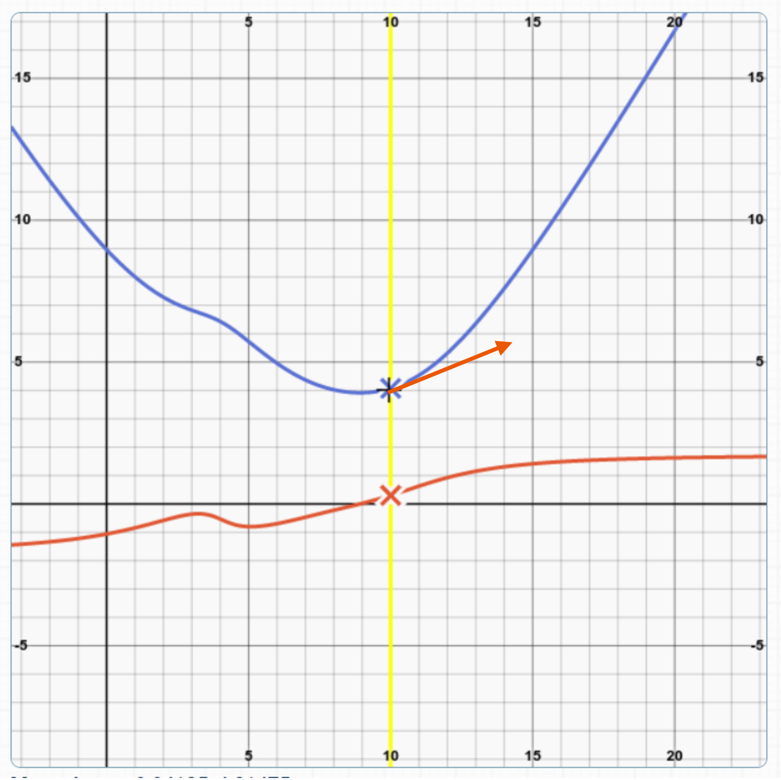
MDS: Descenso por gradiente

Esta es la función de error, para $A1_{MDS1}$
El valor actual es 10, y vemos que baja si nos movemos a la izquierda.

Veamos que nos indica su derivada.

En 10, nos da un valor de 0.3, nos indica que crece en dicha dirección. Es decir crece en dirección positiva. El 0.3, nos indica la tasa de aumento del error.

Cómo queremos que decrezca iremos en dirección opuesta, nos moveremos hacia el lado negativo de X.
¿Pero cuanto?
Usamos para ello el parámetro ***learning rate*** que puede tomar valores entre 0 y 1.



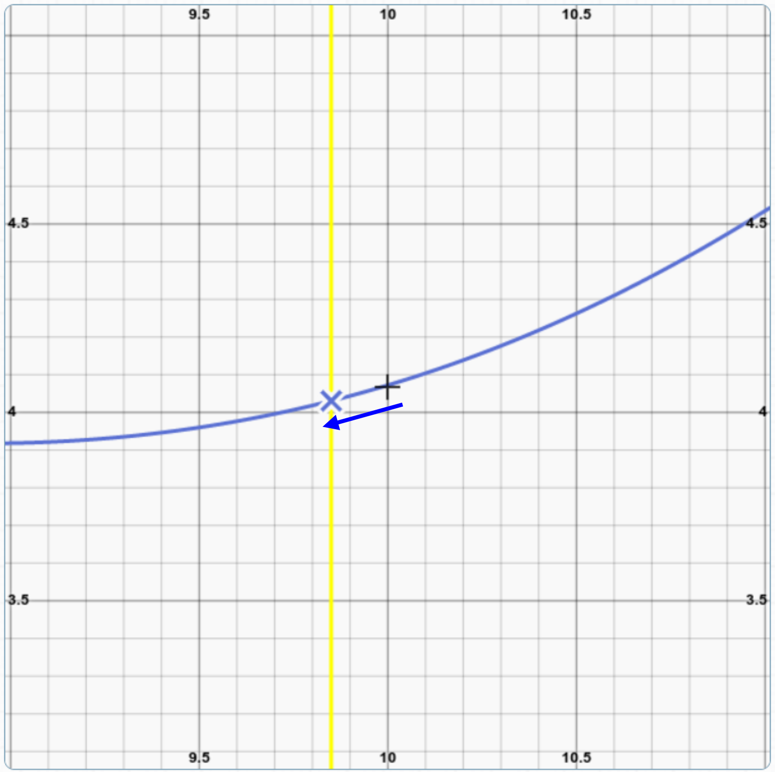
MDS: Descenso por gradiente

Si *learning rate* es, por ejemplo 0.5
Actualizaremos $A1_{MDS1}$, con la siguiente fórmula:

$A1'_{MDS1} = A1'_{MDS1} + 0.5 \times (-1) \times (0.3)$

$A1'_{MDS1} = 9,85$

Nos movemos un pequeño paso en la dirección correcta para achicar el error



MDS: Descenso por gradiente

$$\text{Stress}=\sqrt{(4.24 - \| A1\text{-}A2 \|)^2 + (5 - \| A1\text{-}A3 \|)^2 + (3 - \| A1\text{-}A4 \|)^2 + (6 - \| A2\text{-}A3 \|)^2 + (3.4 - \| A2\text{-}A4 \|)^2 + (5.6 - \| A3\text{-}A4 \|)^2}$$

Puntos	A 1	A 2	A 3	A 4
MDS1	9.85	8	4	9
MDS2	5	9	6	2

$\| A1 - A2 \| = \sqrt{(9.85-8)^2 + (5-9)^2} =$ **4,41**

$\| A1 - A3 \| = \sqrt{(9.85-4)^2 + (5-6)^2} =$ **5.93**

$\| A1 - A4 \| = \sqrt{(9.85-9)^2 + (5-2)^2} =$ **3,11**

$\| A2 - A3 \| = \sqrt{(8-4)^2 + (9-6)^2} =$ **5**

$\| A2 - A4 \| = \sqrt{(8-9)^2 + (9-2)^2} =$ **7.1**

$\| A3 - A4 \| = \sqrt{(4-9)^2 + (6-2)^2} =$ **6.4**

Redujimos el error de 4.05 a 4

$$\sqrt{(4.24 - 4.41)^2 + (5 - 5.93)^2 + (3 - 3.11)^2 + (6 - 5)^2 + (3.4 - 7.1)^2 + (5.6 - 6.4)^2} = \mathbf{4}$$

MDS: Descenso por gradiente

El cálculo que se hizo para el componente $A1_{MDS1}$, hay que repetirlo para todos los componentes y en **N** pasos iterativos ir achicando el error cada vez más, hasta llegar a cero o al menor valor posible (el mínimo de la función de error).

Cabe destacar que en general las raíces se las elimina de la ecuación para facilitar los cálculos, y lo que se intenta minimizar es el **error cuadrático** (es decir el error al cuadrado, o sin la raíz) ya que minimizar el error cuadrático es lo mismo que minimizar el error real.

En este ejemplo se calculó la derivada con la raíz cuadrada, para no perder la forma original de la función de error.

Volviendo a MDS

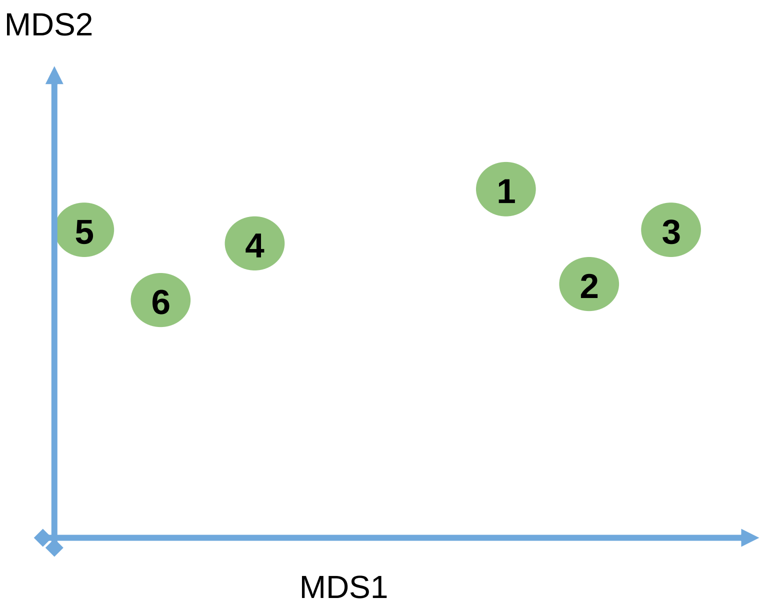
MDS: Multi-Dimensional Scaling

Eventualmente habremos obtenido toda una nueva serie de coordenadas

Puntos	A 1	A 2	A 3	A 4	A 5	A 6
MDS1	8	11	5	2	6	2
MDS2	5	4	9	2	9	2

¡Y obtenemos el mismo gráfico que obtuvimos con PCA!

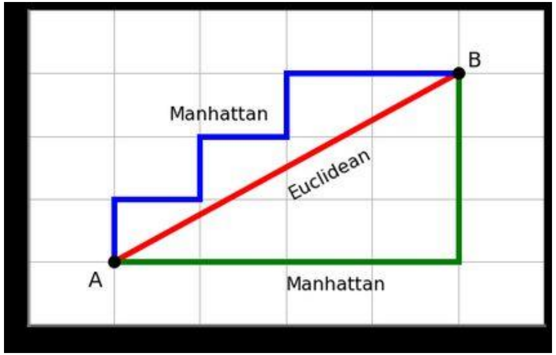
Al graficar el resultado final de dicha matriz:



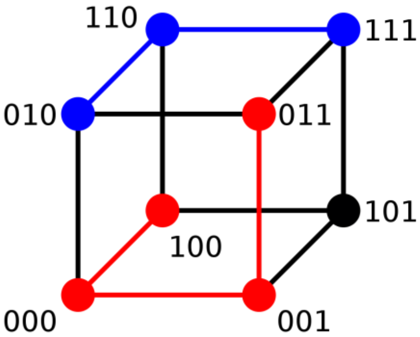
MDS: Multi-Dimensional Scaling

Tratar de formar *clusters* con la técnica de minimizar las distancias lineales entre los puntos, es exactamente lo mismo que maximizar la correlación lineal.
Pero siempre podemos utilizar otras medidas de distancias:

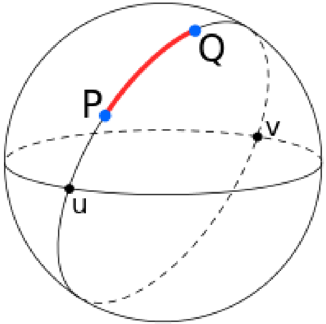
Manhattan distance



Hamming Distance



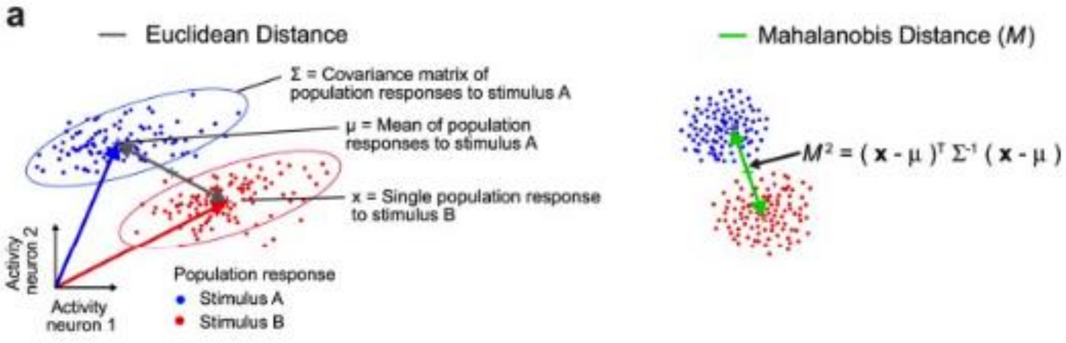
Great-circle distance



Hamming Distance : mide el número mínimo de sustituciones requeridas para cambiar una cadena por otra. **karolin**" y "**kathrin**" tienen distancia 3.

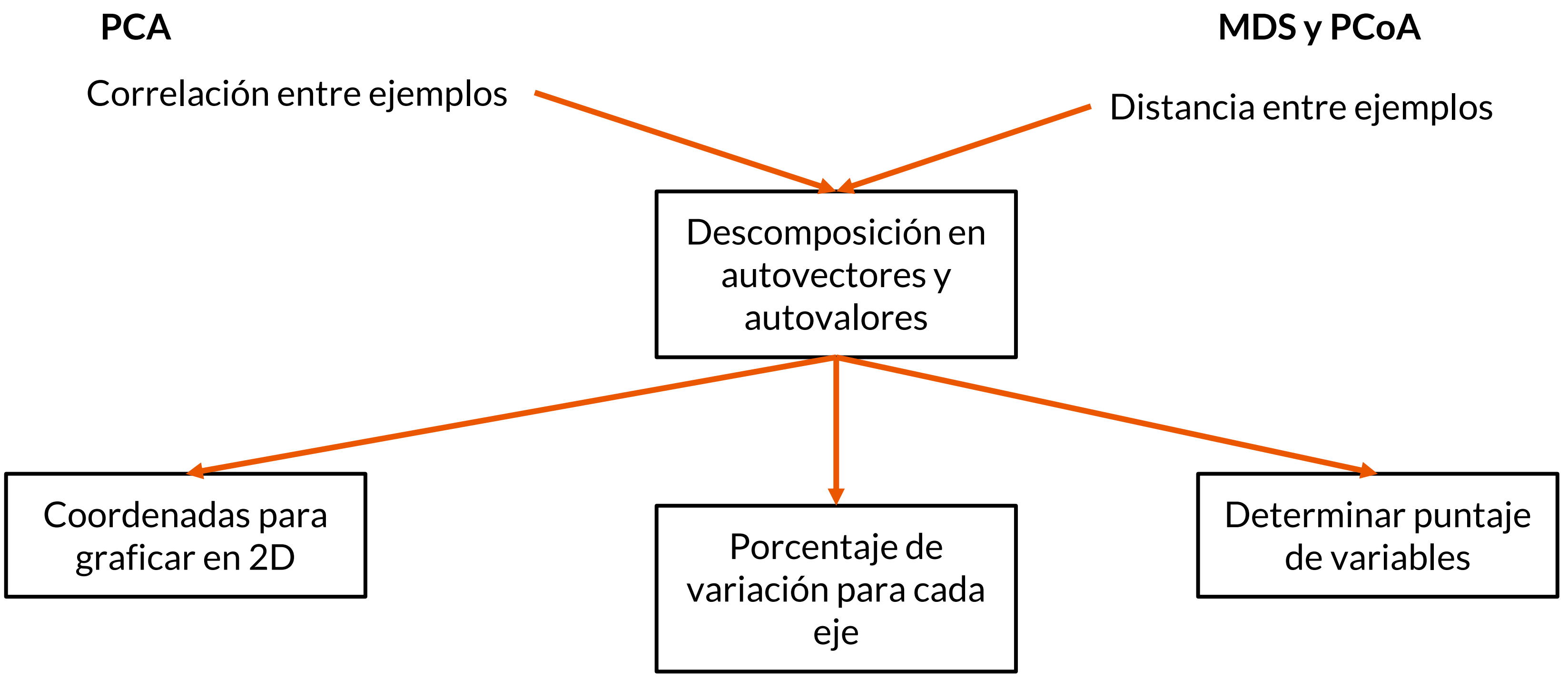
MDS: Multi-Dimensional Scaling

Mahalanobis distance



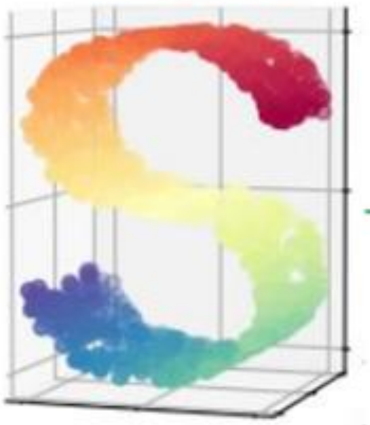
Se puede definir cualquier fórmula apropiada al problema a resolver para medir cuán “cerca” o “lejos” están dos variables

MDS: Multi-Dimensional Scaling



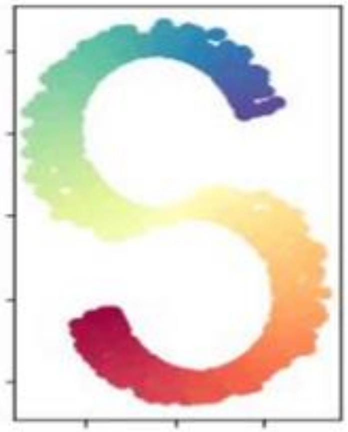
MDS: Multi-Dimensional Scaling

Ejemplo: Ambos gráficos tienen la misma matriz de distancia.



Distance Matrix

	A	B	C	D	E	F
A	0	16	47	72	77	79
B	16	0	37	57	65	66
C	47	37	0	40	30	35
D	72	57	40	0	31	23
E	77	65	30	31	0	10
F	79	66	35	23	10	0



MDS: Multi-Dimensional Scaling

Fortalezas:

- Soporta varios tipos de distancias
- Permite transformaciones no lineales

Debilidades:

- Optimización iterativa con mínimos locales
- Difícil determinar que distancia a usar es la mejor