UTEC
UNIVERSIDAD DE INGENIERÍA Y TECNOLOGÍA

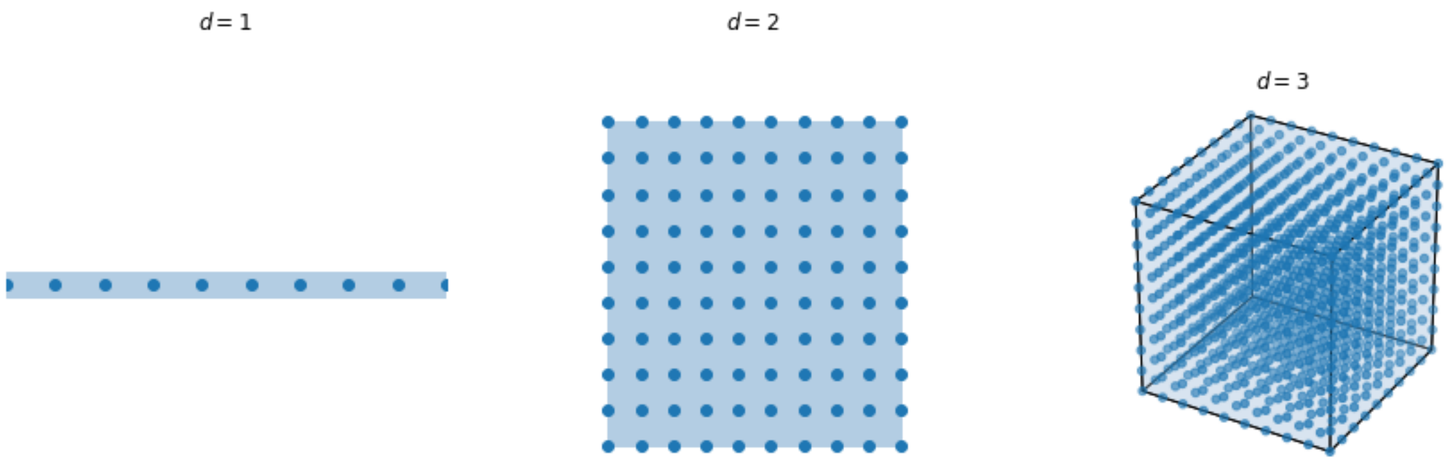
Sesión 10.1: Multiverse of Madness

CS3102 EDA

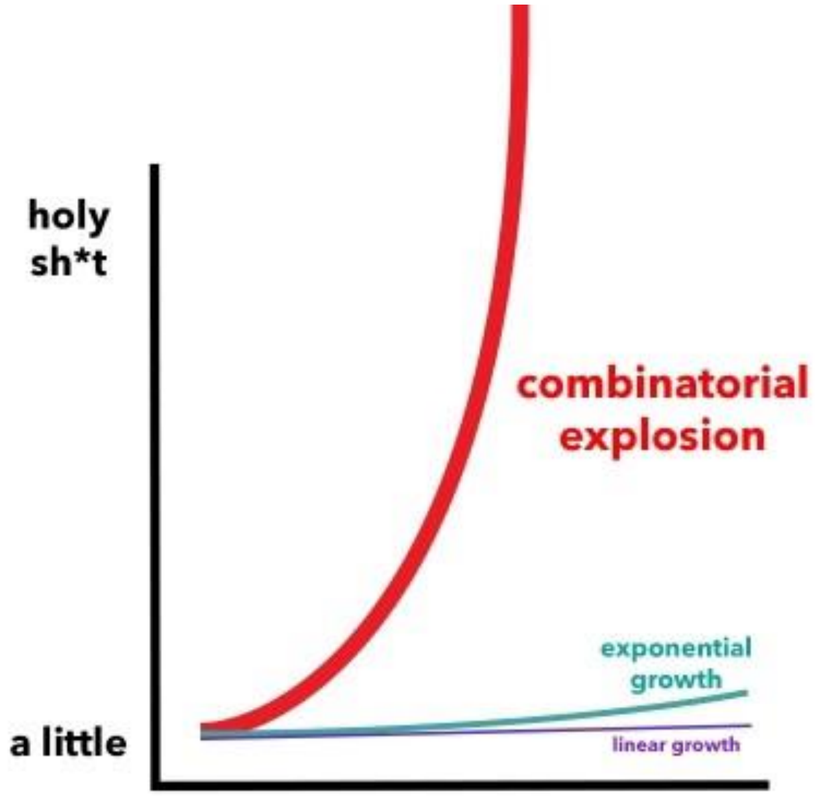
Índice

1. Maldición de la Dimensionalidad

Combinatorial *explosion*



Combinatorial *explosion*



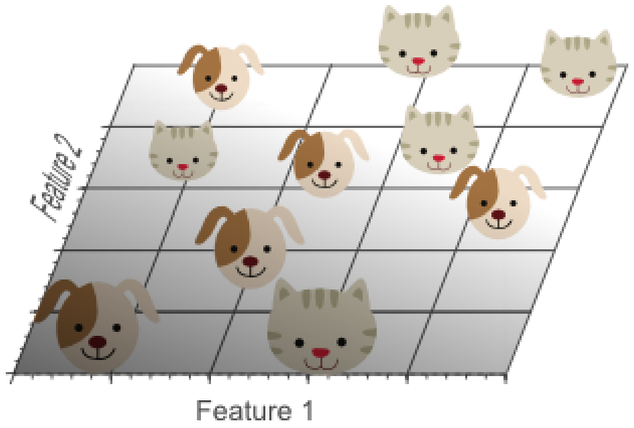
Muestreo

Una dimensión

Feature 1

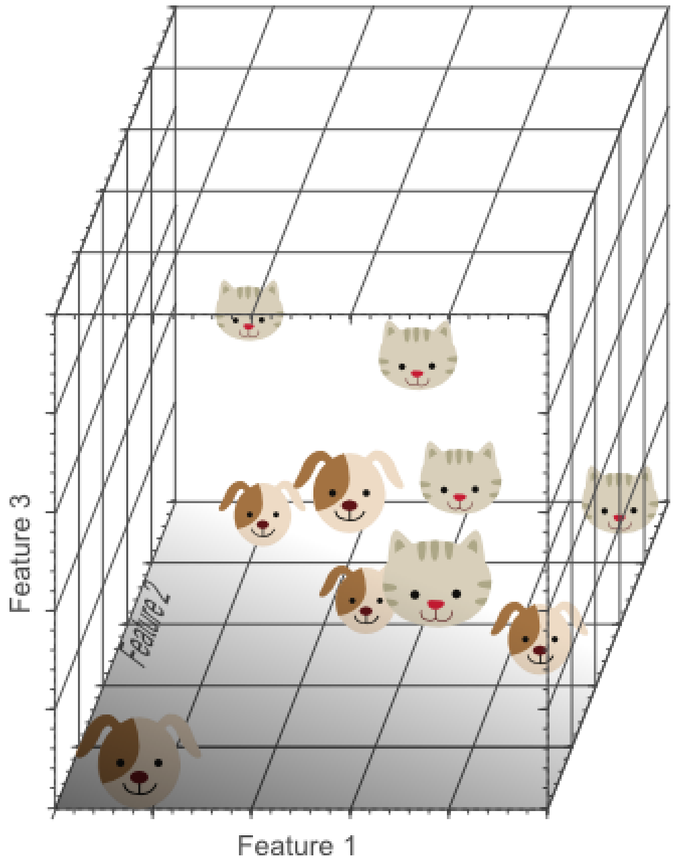
Muestreo

Dos dimensiones

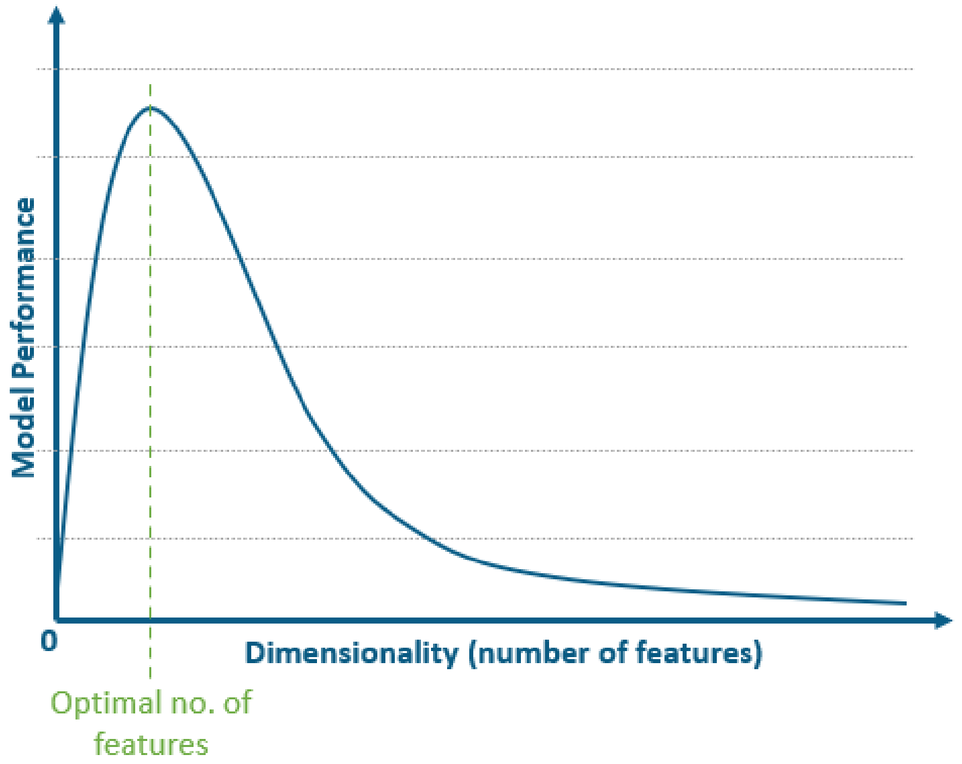


Muestreo

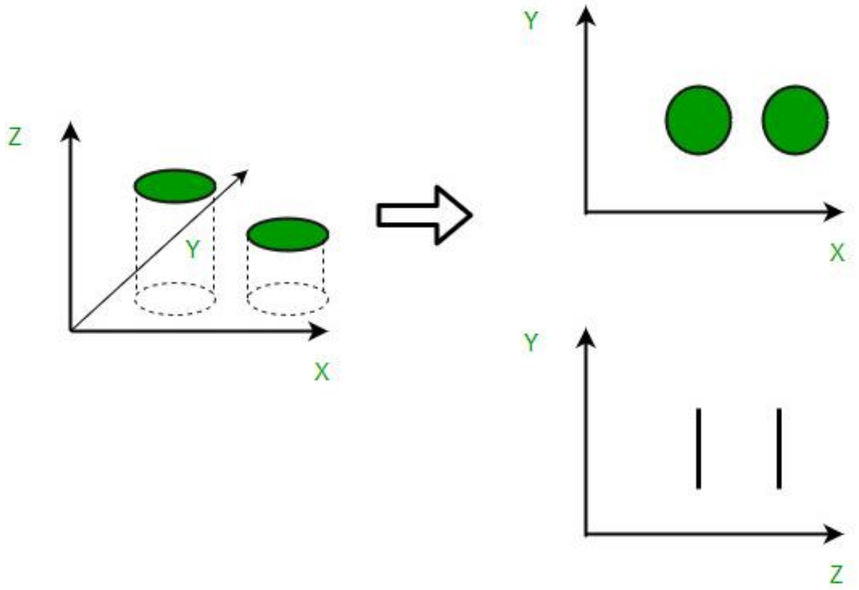
Tres dimensiones

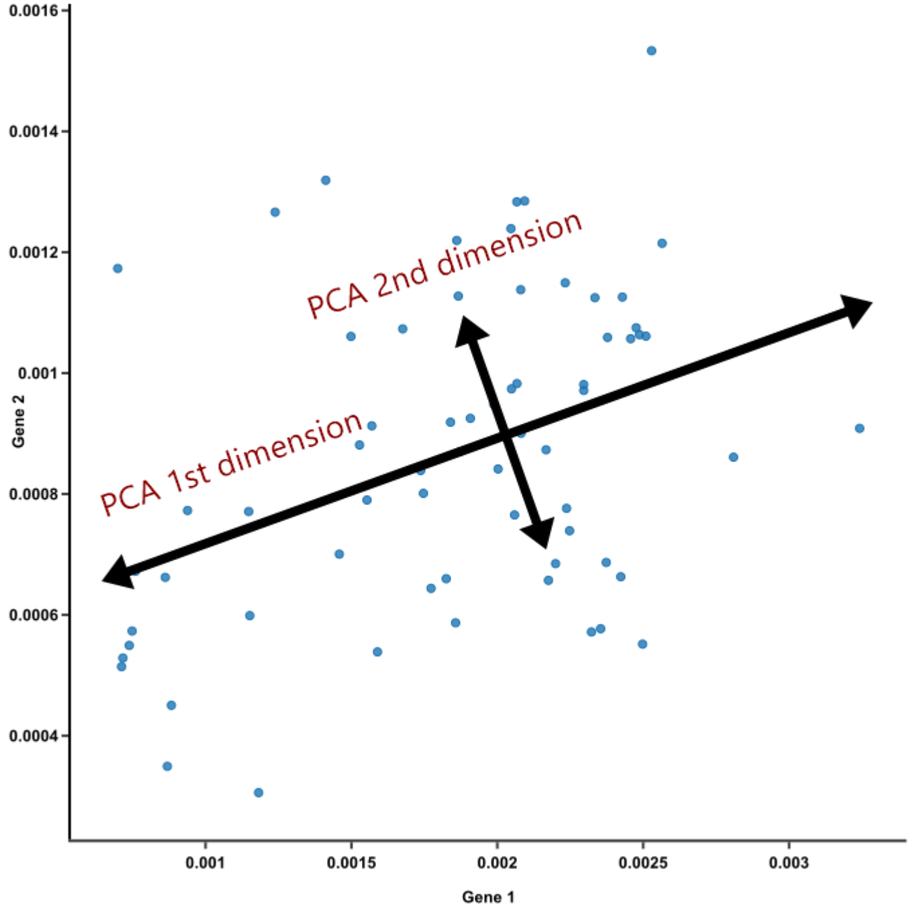


Hughes *Phenomenon*

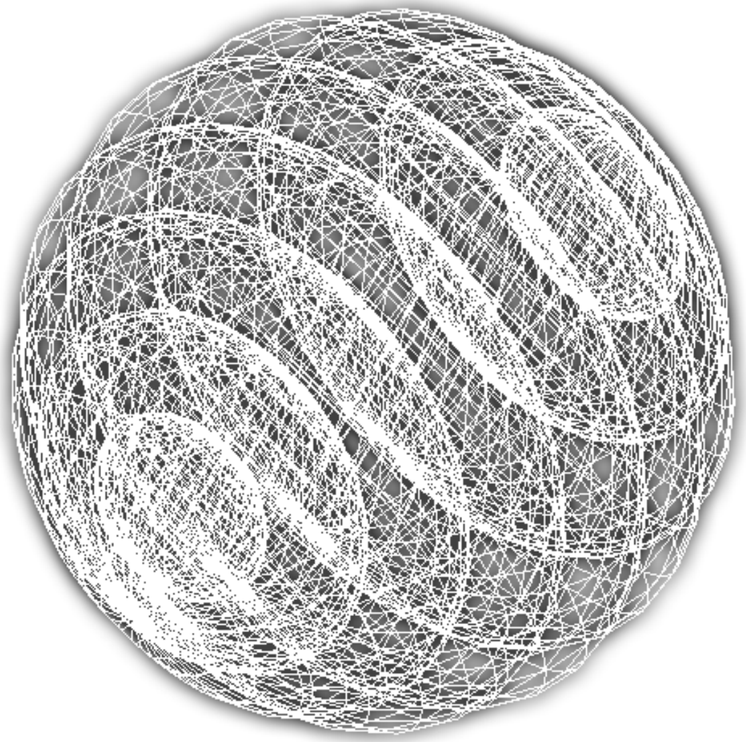


Reducción de *Dimensionalidad*

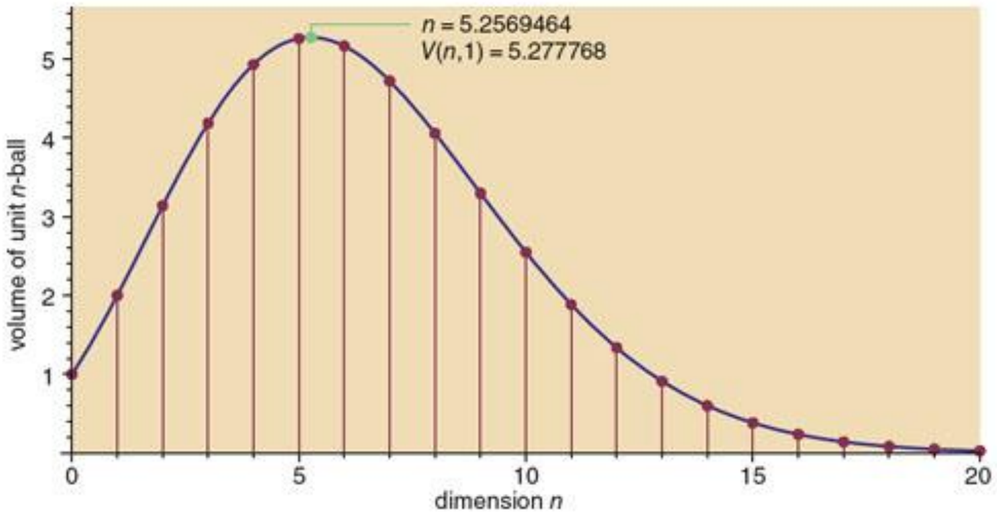




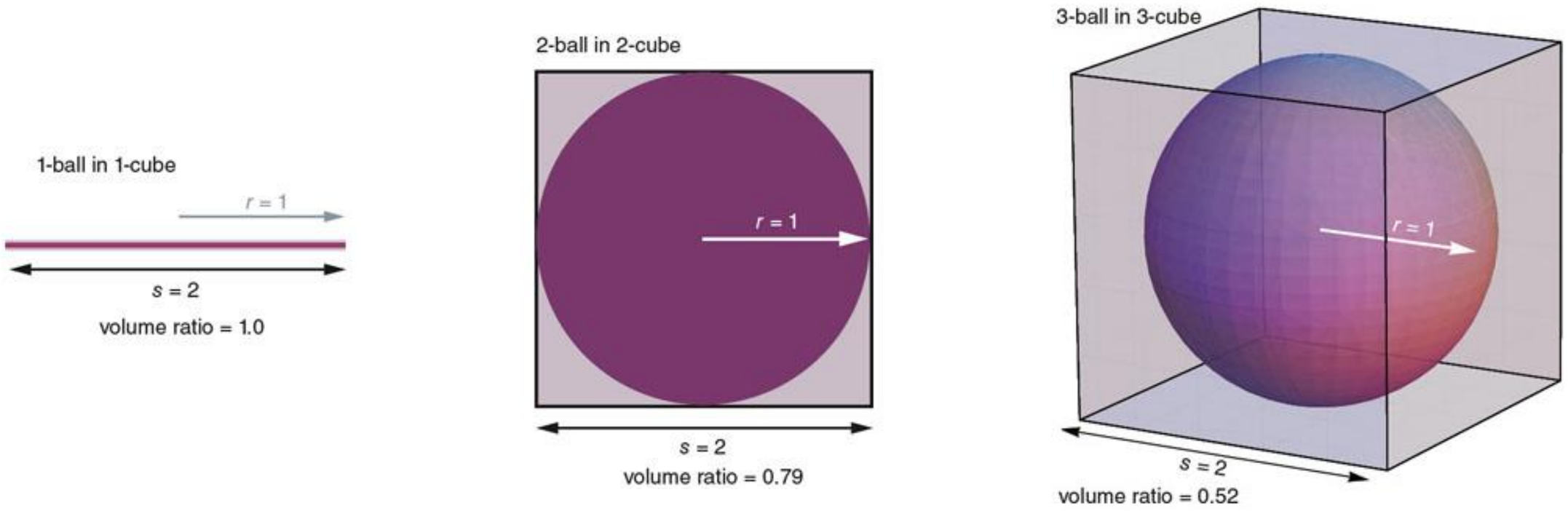
Maldición de la *Dimensionalidad*



$$V_n = \frac{\pi^{n/2} R^n}{\Gamma\left(1 + \frac{n}{2}\right)} = \frac{\pi^{n/2} R^n}{\left(\frac{n}{2}\right)!}$$



Maldición de la *Dimensionalidad*



Maldición de la *Dimensionalidad*

2. Métodos de Acceso Métrico

Espacio *métrico*

$$M = (D, d) \qquad d: M \times M \to \mathbb{R}$$

Dominio de los datos

Función distancia

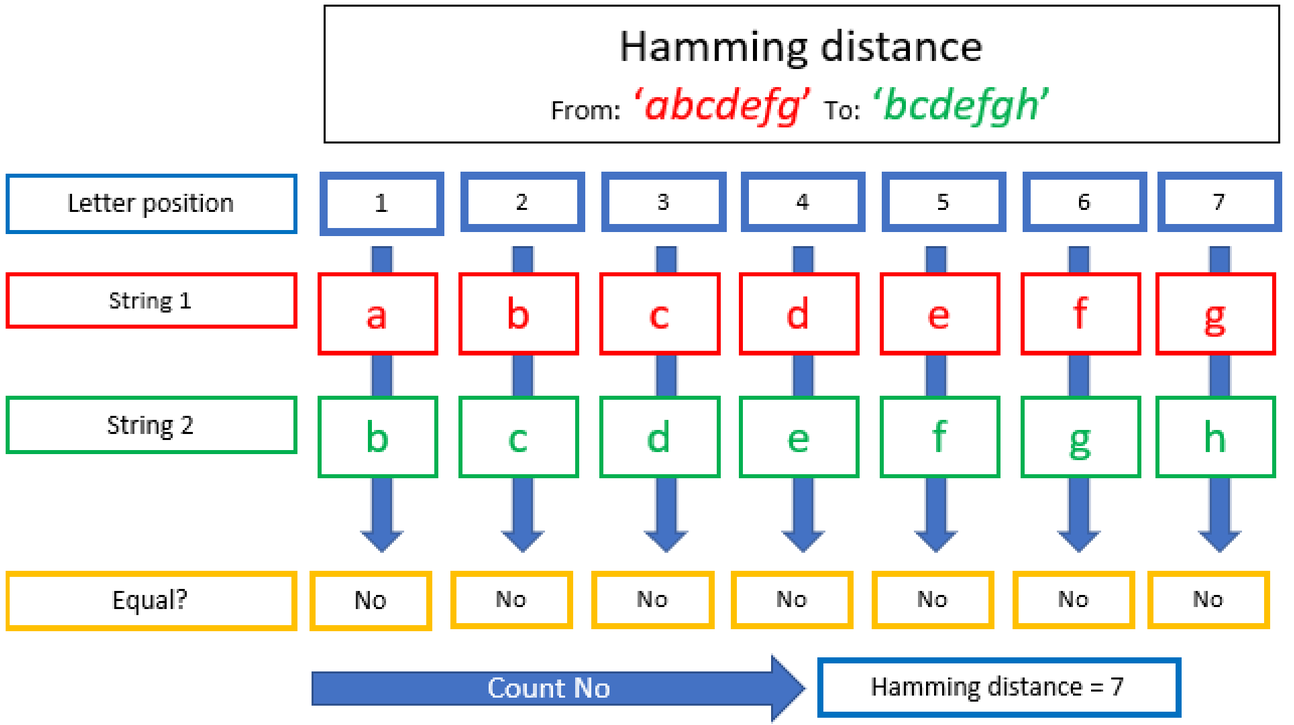
Simetría: $d(x, y) = d(y, x)$

Identidad: $x = y \leftrightarrow d(x, y) = 0$

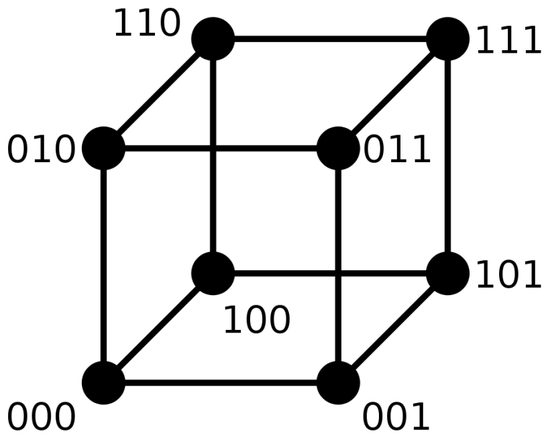
Desigualdad triangular: $d(x, z) \leq d(x, y) + d(y, z)$

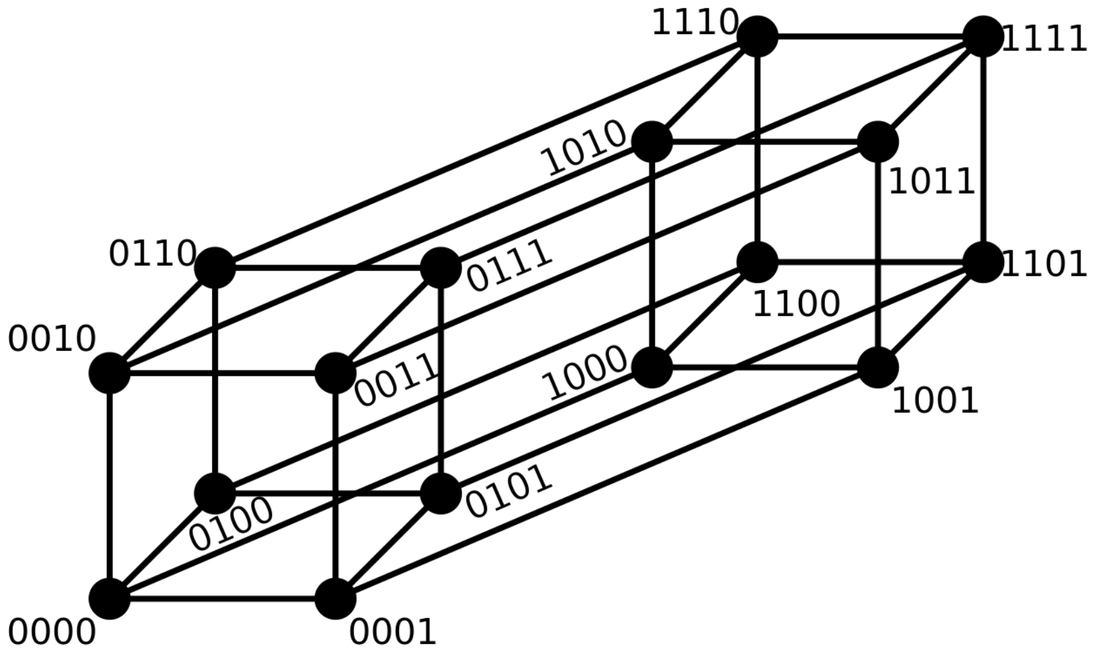
No negatividad: $d(x, y) \geq 0$

Distancia *Hamming*

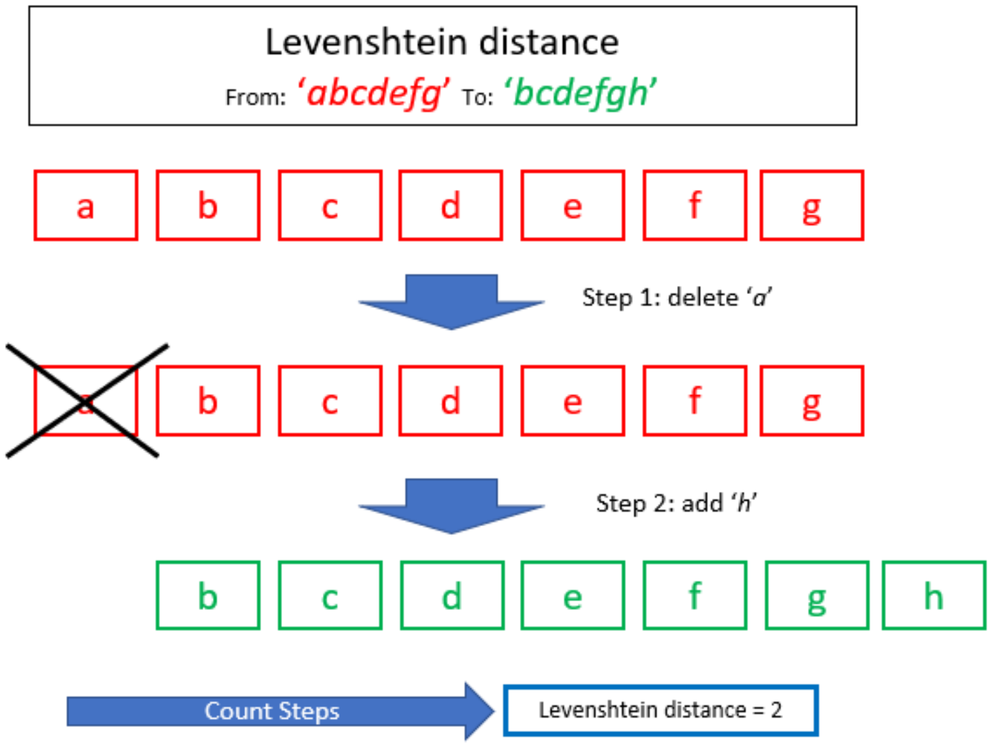


Distancia *Hamming*





Distancia *Levenshtein*

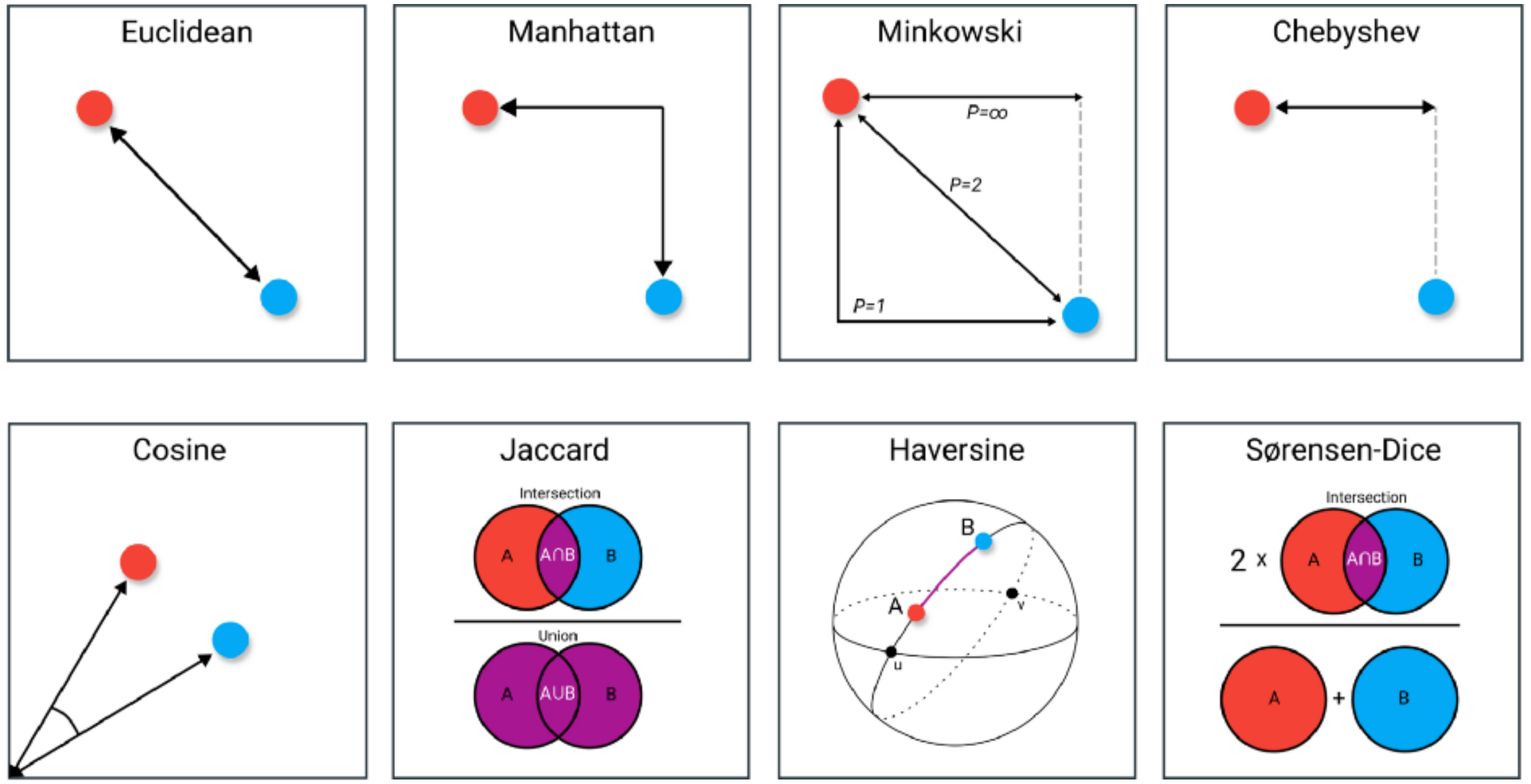


Distancia *Levenshtein*

$$\operatorname{lev}(a,b) = \begin{cases} |a| & \text{if } |b| = 0, \\ |b| & \text{if } |a| = 0, \\ \operatorname{lev}\left(\operatorname{tail}(a), \operatorname{tail}(b)\right) & \text{if } a[0] = b[0], \\ 1 + \min \begin{cases} \operatorname{lev}\left(\operatorname{tail}(a), b\right) \\ \operatorname{lev}\left(a, \operatorname{tail}(b)\right) \\ \operatorname{lev}\left(\operatorname{tail}(a), \operatorname{tail}(b)\right) \end{cases} & \text{otherwise}, \end{cases}$$

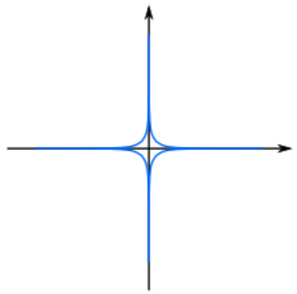
		h	e	l	l	o
	0	1	2	3	4	5
k	1	1	2	3	4	5
e	2	2	1	2	3	4
l	3	3	2	1	2	2
m	4	4	3	2	2	3

Métricas *continuas*

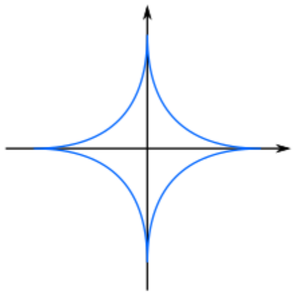


Distancias *Minkowski*

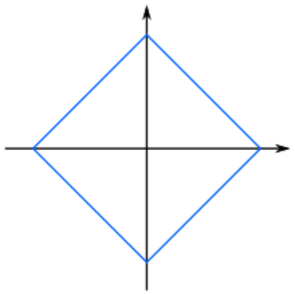
$$D(X,Y) = \left(\sum_{i=1}^{n} |x_i - y_i|^p \right)^{\frac{1}{p}}.$$



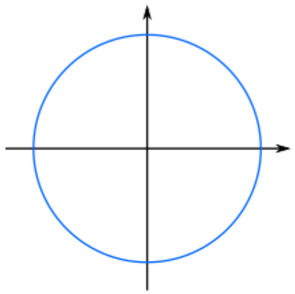
$p = 2^{-2}$
$= 0.25$



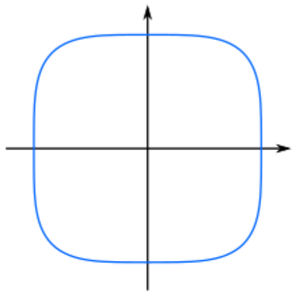
$p = 2^{-1}$
$= 0.5$



$p = 2^{0}$
$= 1$

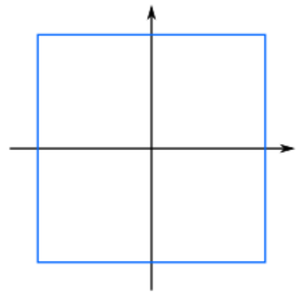


$p = 2^{1}$
$= 2$



$p = 2^{2}$
$= 4$

. . .

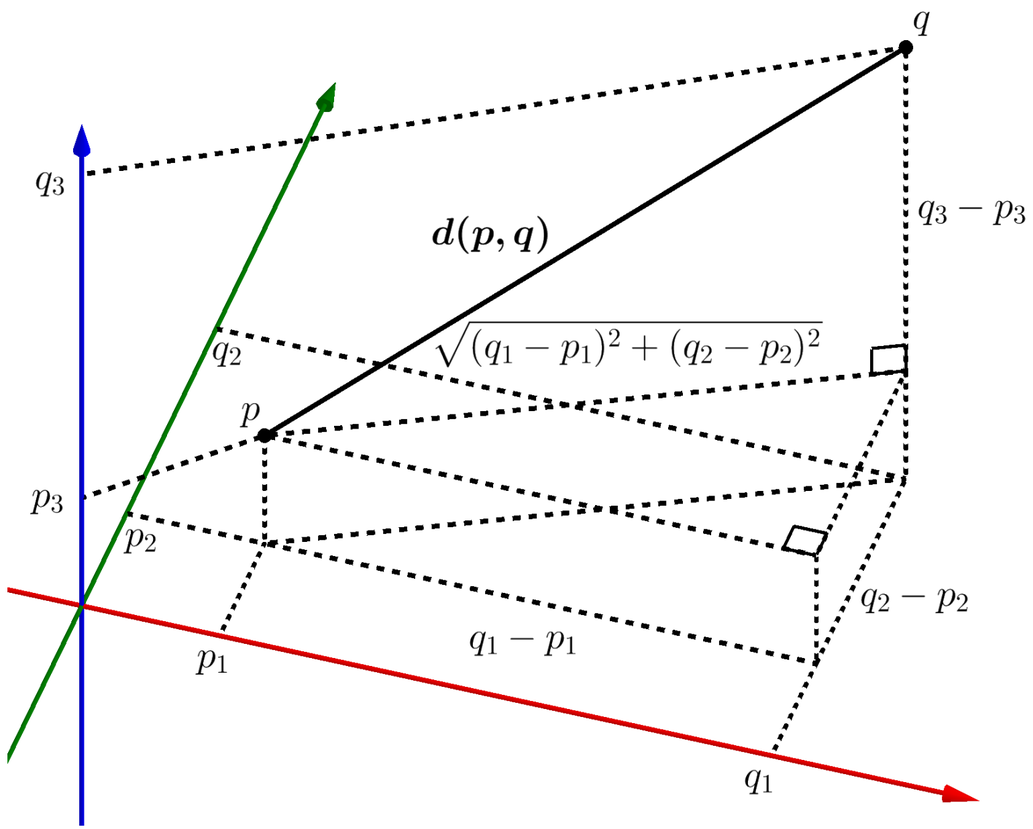


$p = 2^{\infty}$
$= \infty$

Distancia *Manhattan*

$$d_{\mathrm{T}}(\mathbf{p}, \mathbf{q}) = \|\mathbf{p} - \mathbf{q}\|_{\mathrm{T}} = \sum_{i=1}^{n} |p_i - q_i|$$

Distancia *Euclidiana*



$$d(\mathbf{p}, \mathbf{q}) = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}$$

Distancia *Chebyshev*

$$d_\infty(x,y) = \lim_{p\to\infty}\left(\sum_{i=1}^{n}|x_i - y_i|^p\right)^{1/p}$$

$$d_\infty(x,y) = \max_i(|x_i - y_i|)$$

$$d_\infty(x,y) = \max(|x_2 - x_1|, |y_2 - y_1|)$$

UTEC

UNIVERSIDAD DE INGENIERÍA
Y TECNOLOGÍA