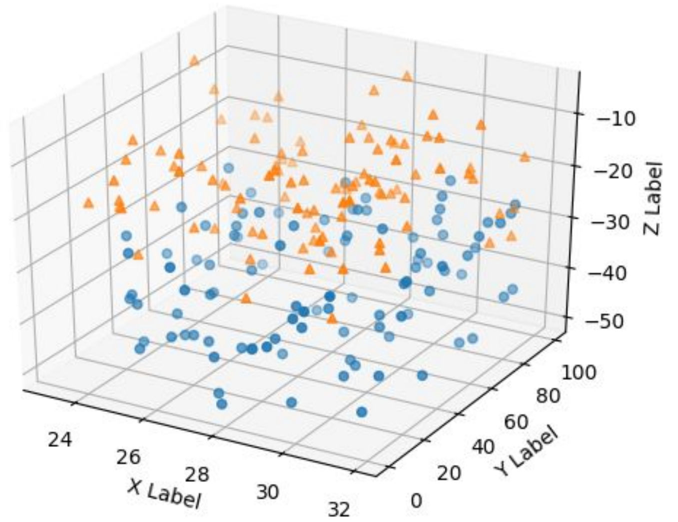
t-SNE

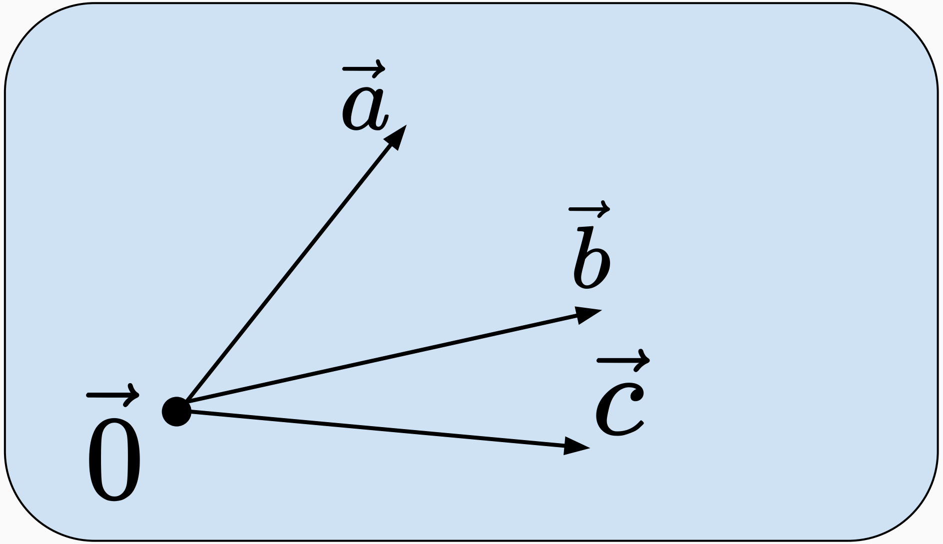
t-Distributed Stochastic Neighbor Embedding

Motivación

$\mathbb{R}^3$

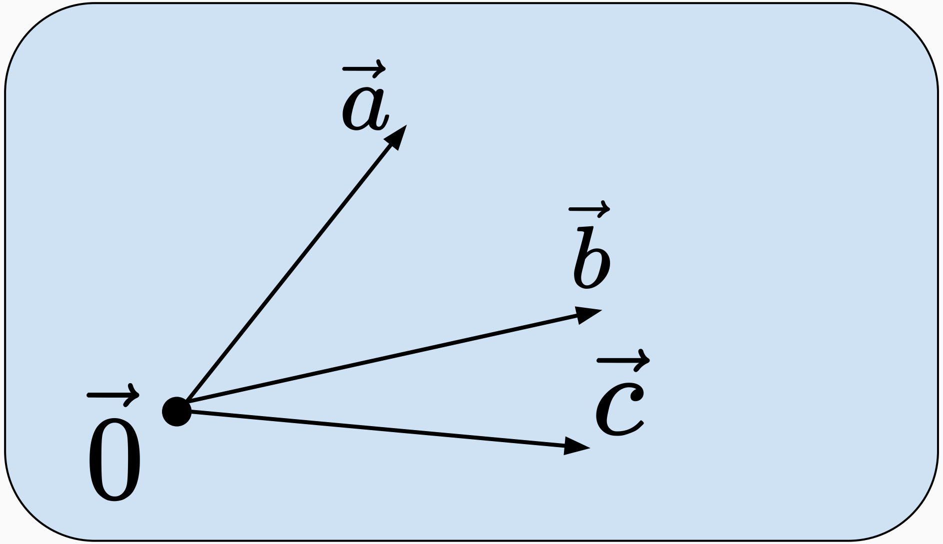


$\mathbb{R}^n$

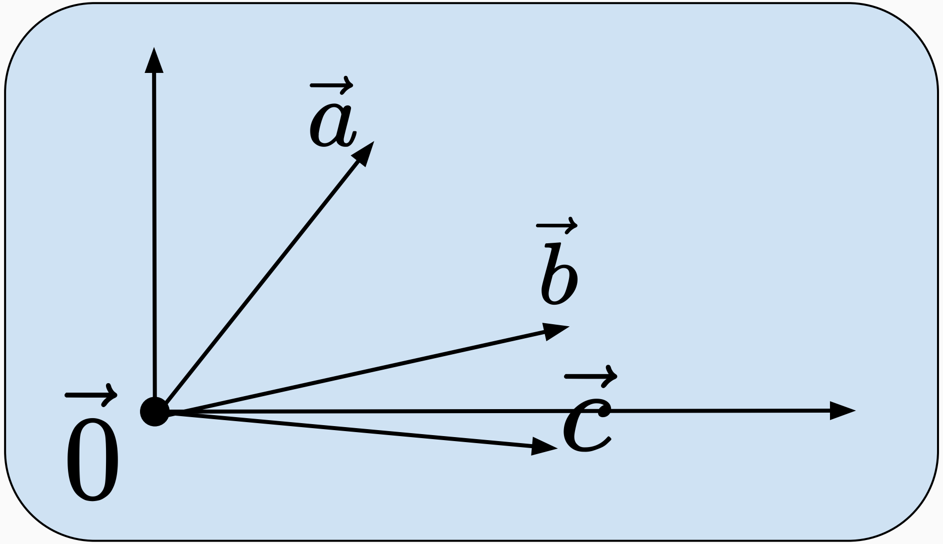


Motivación

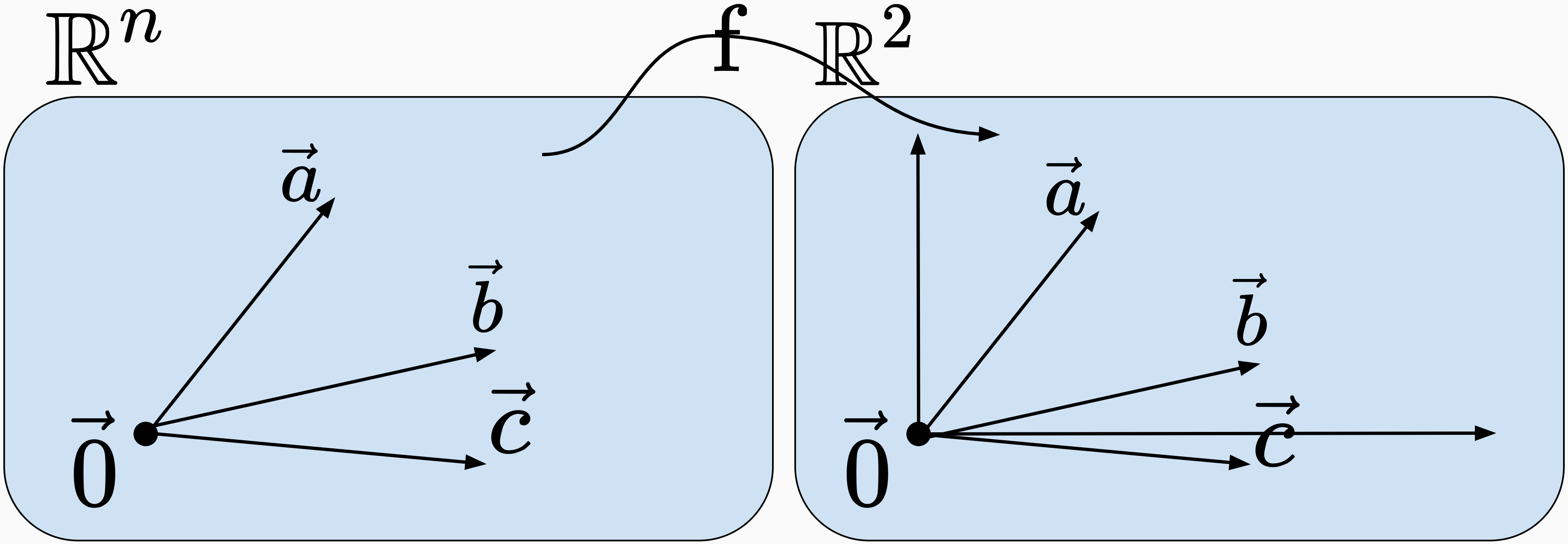
$\mathbb{R}^n$



$\mathbb{R}^2$

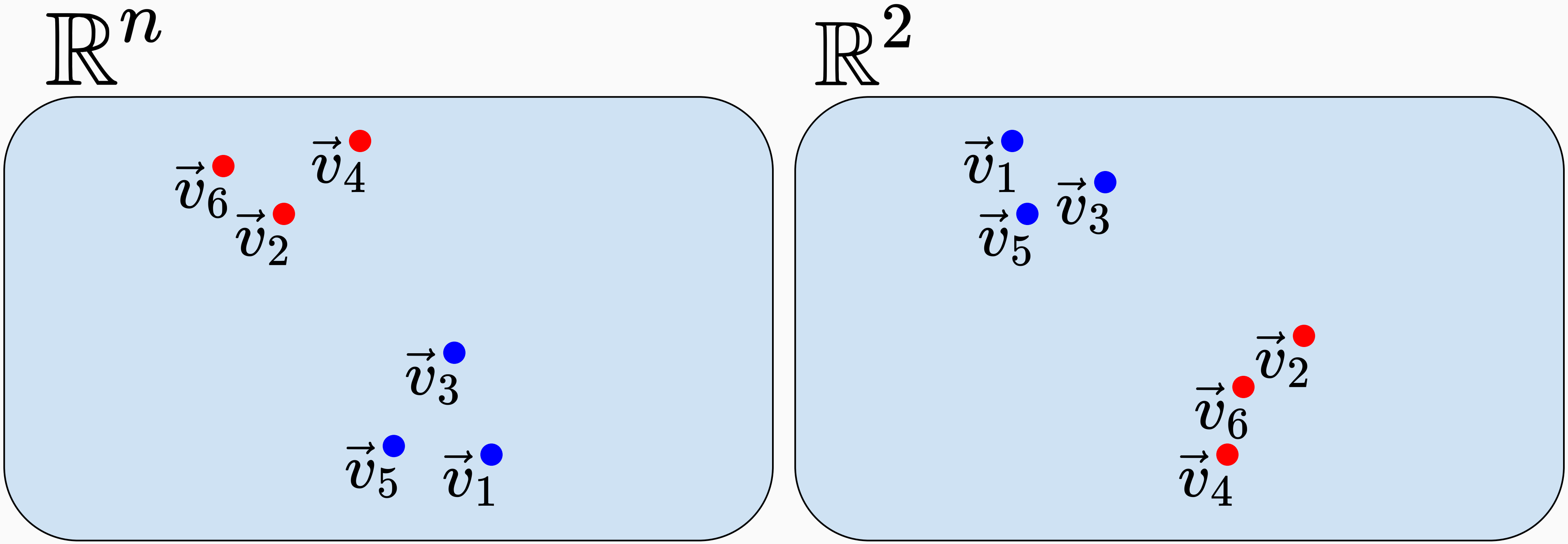


Motivación



$\mathbb{R}^n$
$\mathbf{f}$
$\mathbb{R}^2$
$\vec{a}$
$\vec{b}$
$\vec{c}$
$\vec{0}$
$\vec{a}$
$\vec{b}$
$\vec{c}$
$\vec{0}$

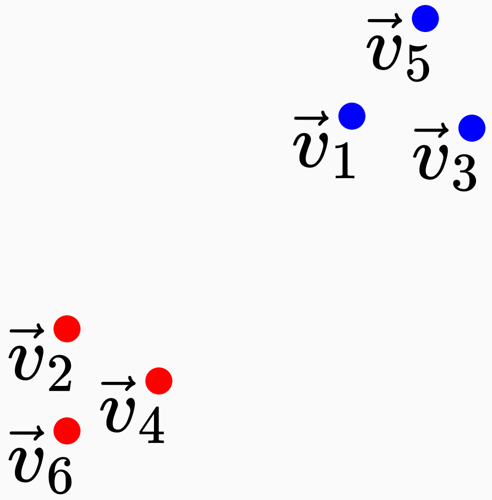
Motivación


$\mathbb{R}^n$
$\vec{v}_4$
$\vec{v}_6$
$\vec{v}_2$
$\vec{v}_3$
$\vec{v}_5$
$\vec{v}_1$
$\mathbb{R}^2$
$\vec{v}_1$
$\vec{v}_3$
$\vec{v}_5$
$\vec{v}_2$
$\vec{v}_6$
$\vec{v}_4$

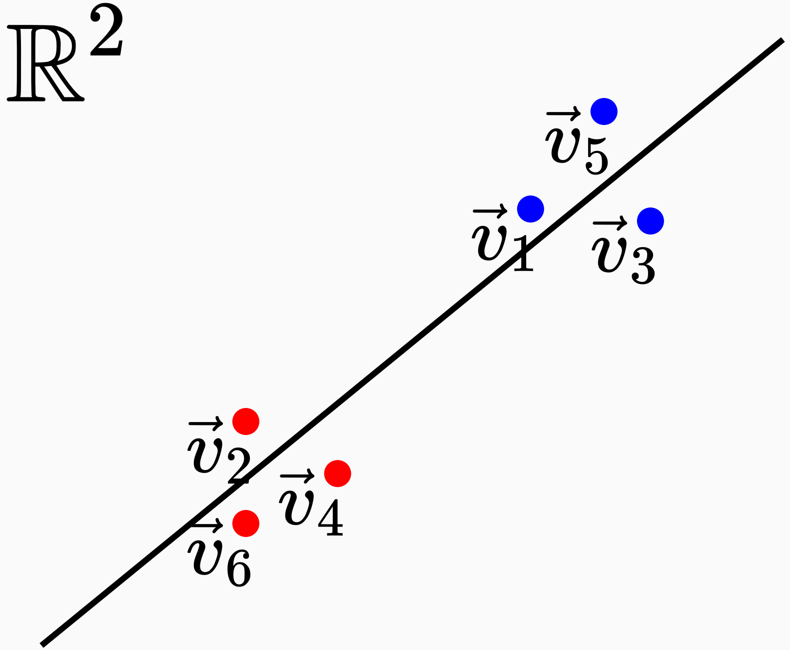
¿Qué tal PCA?

$\mathbb{R}^2$

$\mathbb{R}^1$

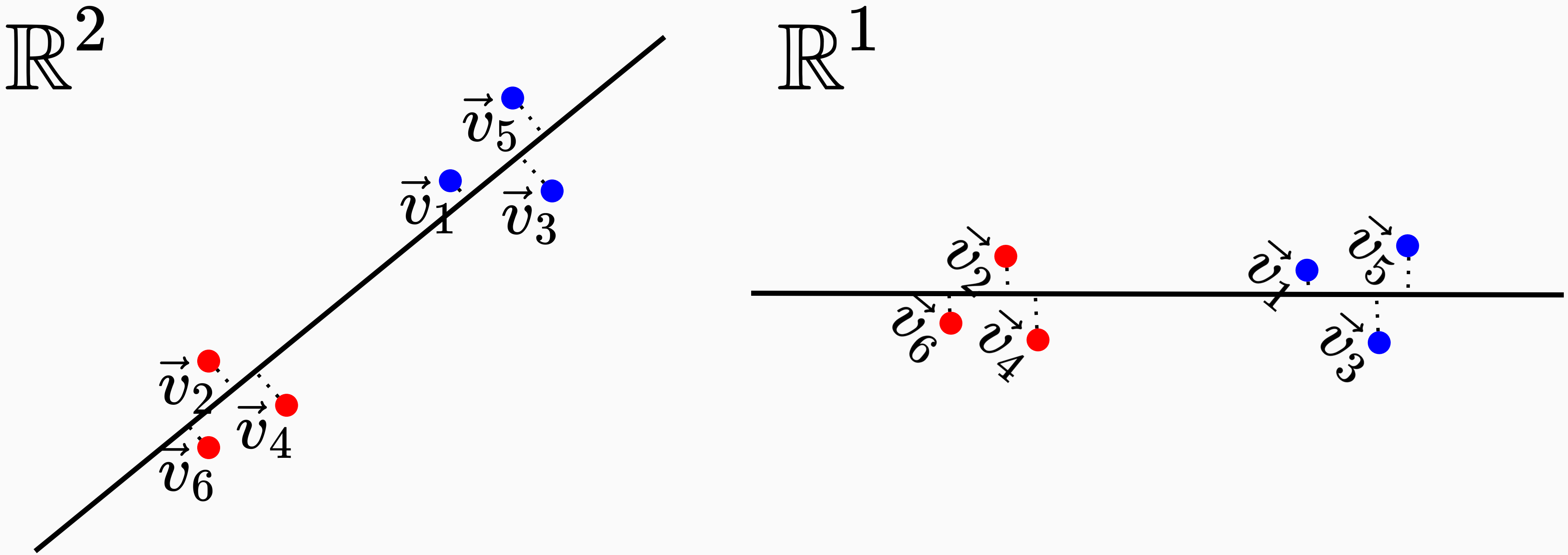


¿Qué tal PCA?



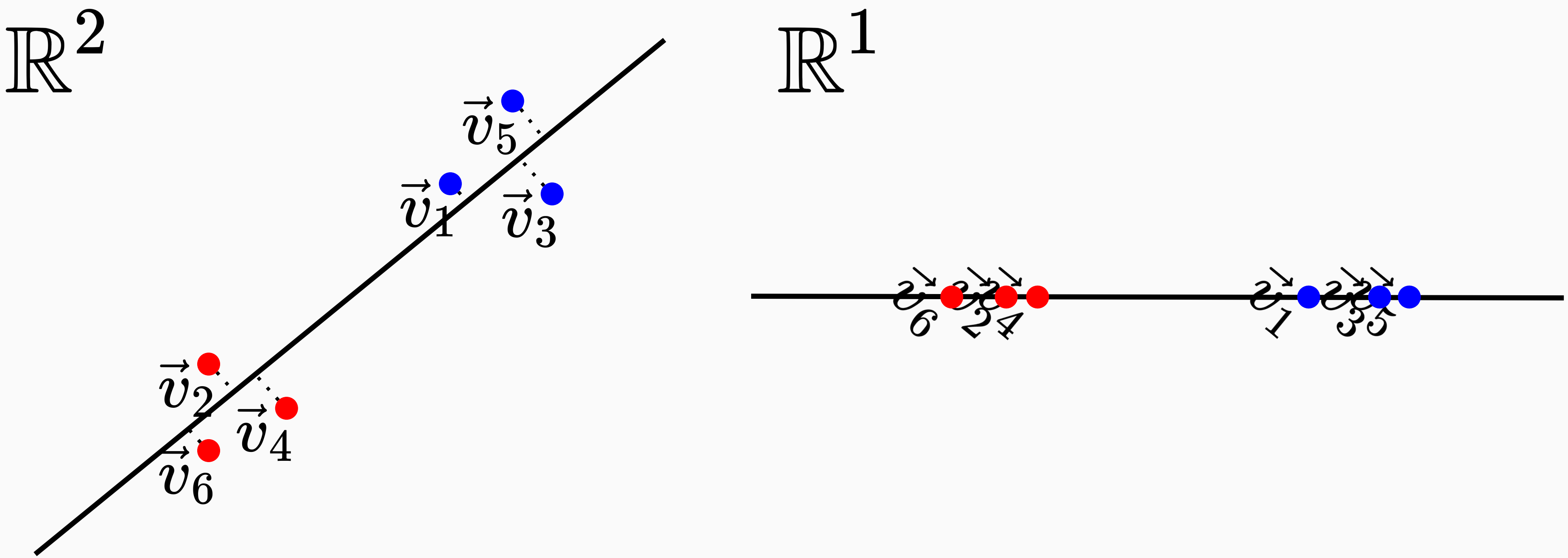
$\mathbb{R}^1$

¿Qué tal PCA?

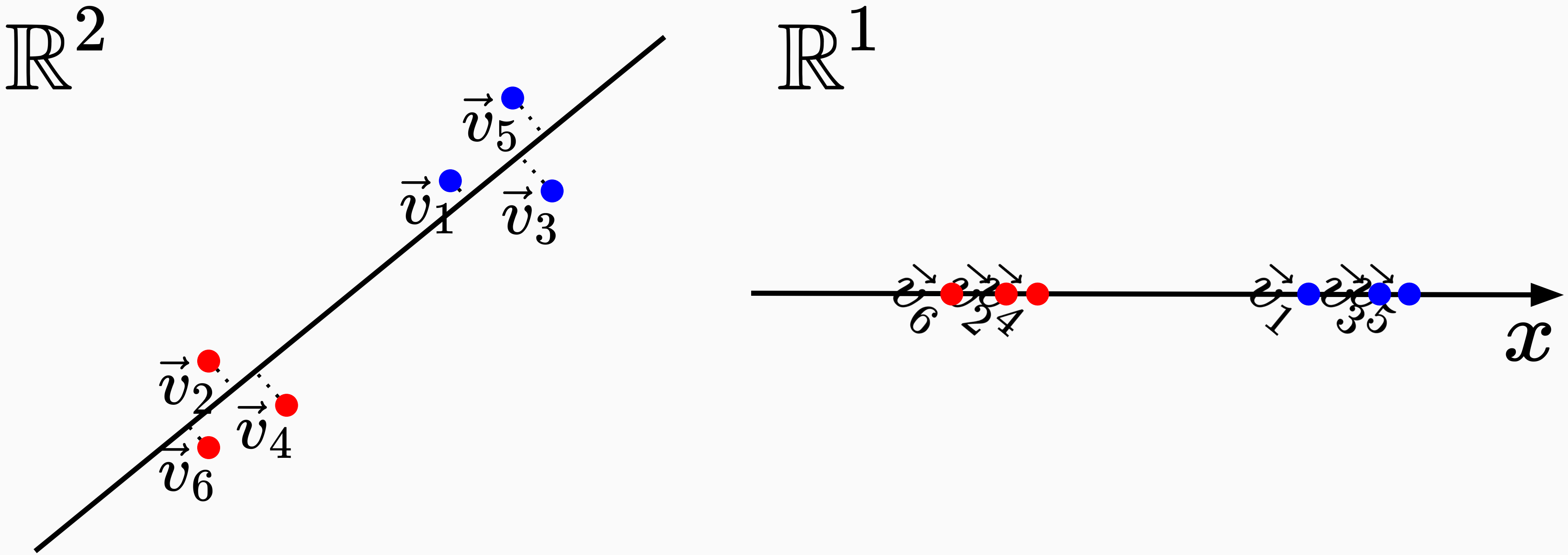


$\mathbb{R}^2$
$\vec{v}_5$
$\vec{v}_1$
$\vec{v}_3$
$\vec{v}_2$
$\vec{v}_4$
$\vec{v}_6$
$\mathbb{R}^1$
$\vec{v}_2$
$\vec{v}_6$
$\vec{v}_4$
$\vec{v}_1$
$\vec{v}_5$
$\vec{v}_3$

¿Qué tal PCA?

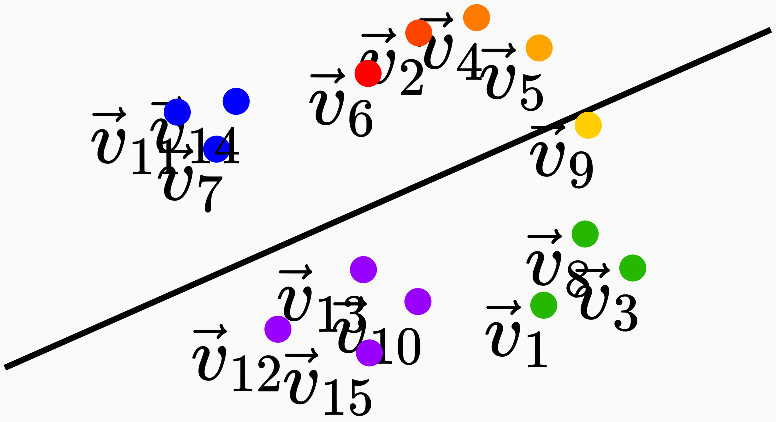


¿Qué tal PCA?



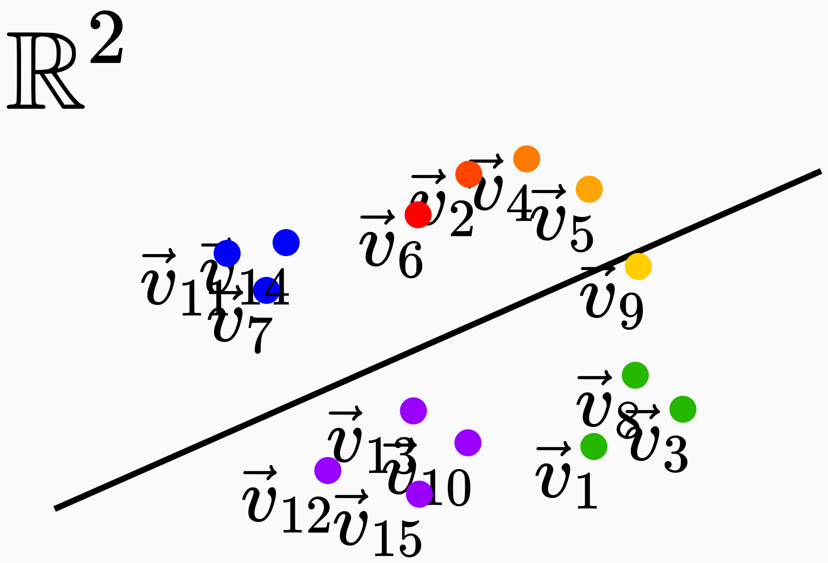
¿Qué tal PCA?

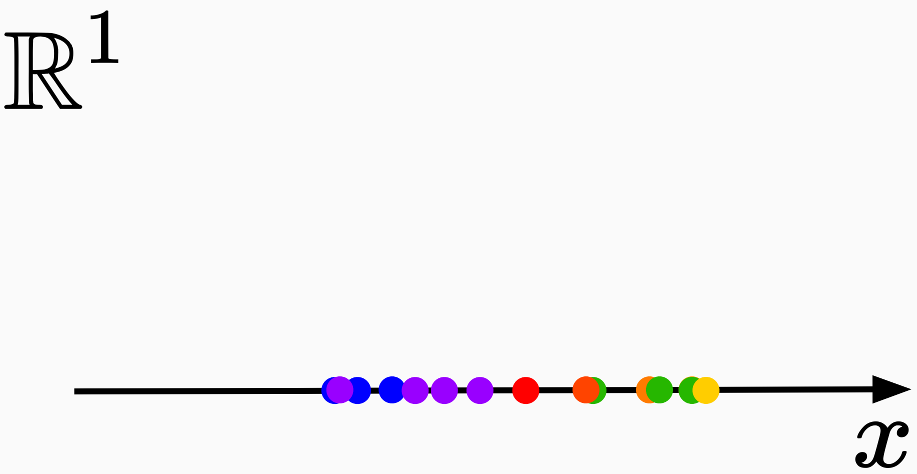
$\mathbb{R}^2$



$\mathbb{R}^1$

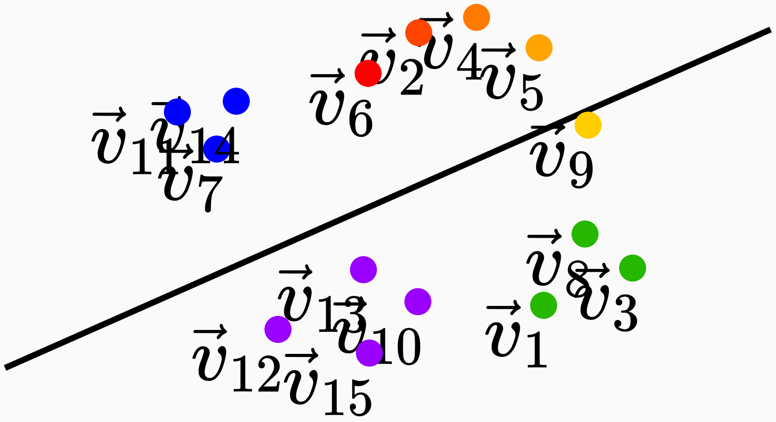
¿Qué tal PCA?


$\mathbb{R}^2$
$\vec{v}_2$
$\vec{v}_4$
$\vec{v}_5$
$\vec{v}_6$
$\vec{v}_{11}$
$\vec{v}_{14}$
$\vec{v}_7$
$\vec{v}_9$
$\vec{v}_8$
$\vec{v}_3$
$\vec{v}_{13}$
$\vec{v}_{10}$
$\vec{v}_1$
$\vec{v}_{12}$
$\vec{v}_{15}$


$\mathbb{R}^1$
x

¿Qué tal PCA?

$\mathbb{R}^2$



$\mathbb{R}^1$

SNE: Stochastic Neighbor Embedding

Stochastic Neighbor Embedding

Geoffrey Hinton and Sam Roweis
Department of Computer Science, University of Toronto
10 King's College Road, Toronto, M5S 3G5 Canada
{*hinton,roweis*}@*cs.toronto.edu*

Softmax

$$y(x)_i = \frac{exp(x_i)}{\sum_j exp(x_j)}$$

Softmax

$$x = \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix} \xrightarrow{\;y(x)_i = \frac{exp(x_i)}{\sum_j exp(x_j)}\;} y = \begin{pmatrix} 0.09 \\ 0.67 \\ 0.24 \end{pmatrix}$$

Softmax

$$x = \begin{pmatrix} 20 \\ 5 \\ 7 \end{pmatrix} \xrightarrow{\quad y(x)_i = \frac{exp(x_i)}{\sum_j exp(x_j)} \quad} y = \begin{pmatrix} 0.99 \\ 3e\text{-}7 \\ 2e\text{-}6 \end{pmatrix}$$

Softmax

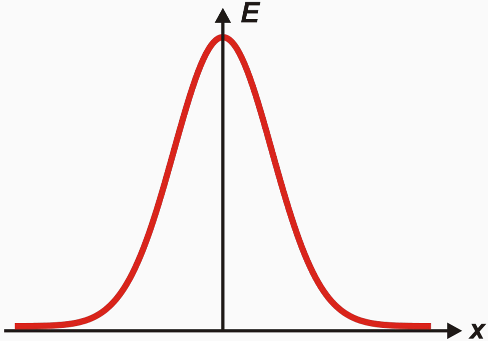
$$y(x)_i = \frac{exp(x_i)}{\sum_j exp(x_j)}$$

$$\sum_i y(x)_i = 1$$

Softmax

$$y(x)_i = \frac{exp(x_i)}{\sum_j exp(x_j)}$$

$$\sum_i y(x)_i = 1$$



Divergencia de Kullback-Leibler

$$D_{KL}(P||Q) = \sum_{s} P(s) \ln(\frac{P(s)}{Q(s)})$$

Divergencia de Kullback-Leibler

Buenos Aires

$$p(\text{Soleado}) = 13/21$$

$$p(\text{Nublado}) = 7/21$$

$$p(\text{Tormenta}) = 1/21$$

Londres

$$p(\text{Soleado}) = 2/18$$

$$p(\text{Nublado}) = 8/18$$

$$p(\text{Tormenta}) = 8/18$$

Divergencia de Kullback-Leibler

Buenos Aires

$$p(\text{Soleado}) = 13/21$$

$$p(\text{Nublado}) = 7/21$$

$$p(\text{Tormenta}) = 1/21$$

Londres

$$p(\text{Soleado}) = 2/18$$

$$p(\text{Nublado}) = 8/18$$

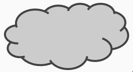
$$p(\text{Tormenta}) = 8/18$$

$$D_{KL}(P_{BsAs} || P_{Lon}) = 0.8610$$

Divergencia de Kullback-Leibler

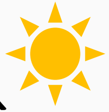
Buenos Aires

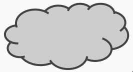
$p\Big($ Soleado  $\Big) = 13/21$

$p\Big($ Nublado  $\Big) = 7/21$

$p\Big($ Tormenta  $\Big) = 1/21$

Londres

$p\Big($ Soleado  $\Big) = 2/18$

$p\Big($ Nublado  $\Big) = 8/18$

$p\Big($ Tormenta  $\Big) = 8/18$

$$D_{KL}(P_{Lon}||P_{Lon}) = 0$$

Divergencia de Kullback-Leibler

Buenos Aires

$$p(\text{Soleado}) = 13/21$$

$$p(\text{Nublado}) = 7/21$$

$$p(\text{Tormenta}) = 1/21$$

Londres

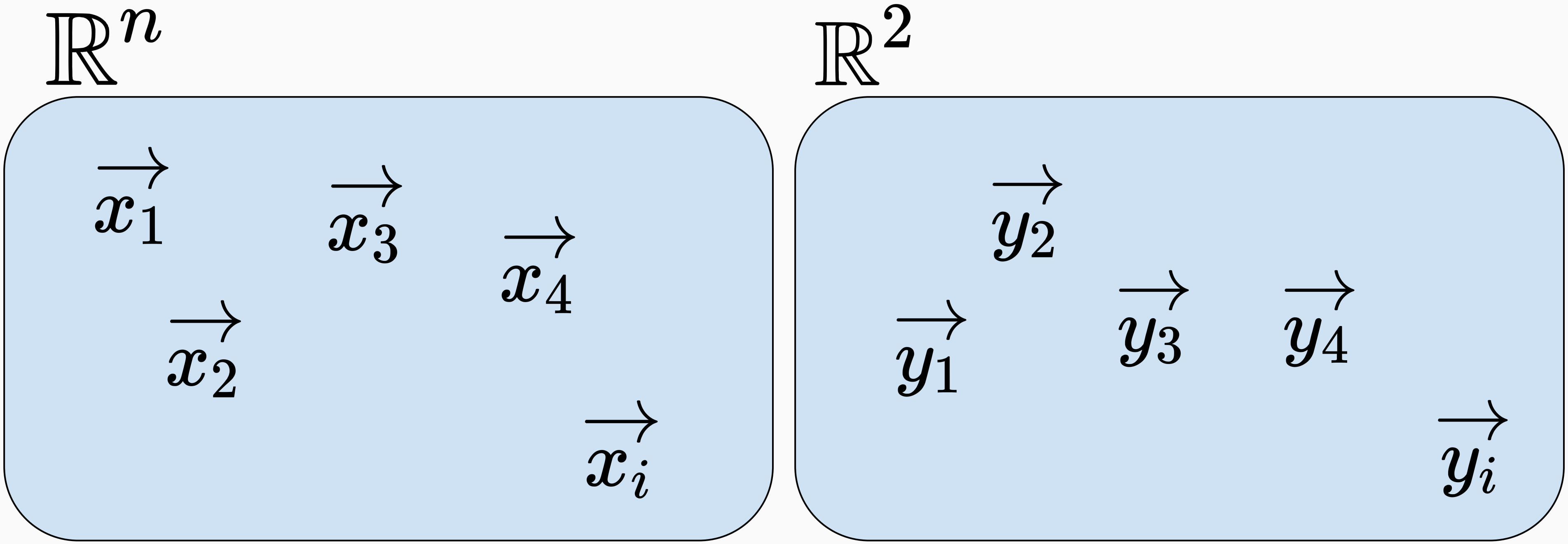
$$p(\text{Soleado}) = 2/18$$

$$p(\text{Nublado}) = 8/18$$

$$p(\text{Tormenta}) = 8/18$$

$$D_{KL}(P_{Lon} || P_{BsAs}) = 0.9297$$

SNE: Stochastic Neighbor Embedding



Disimilaridad

$$d_{ij}^2 = \frac{||x_i - x_j||^2}{2\sigma_i^2}$$

Disimilaridad

$$d_{ij}^2 = \frac{||x_i - x_j||^2}{2\sigma_i^2}$$

Disimilaridad

$$d_{ij}^2 = \frac{||x_i - x_j||^2}{2\sigma_i^2}$$

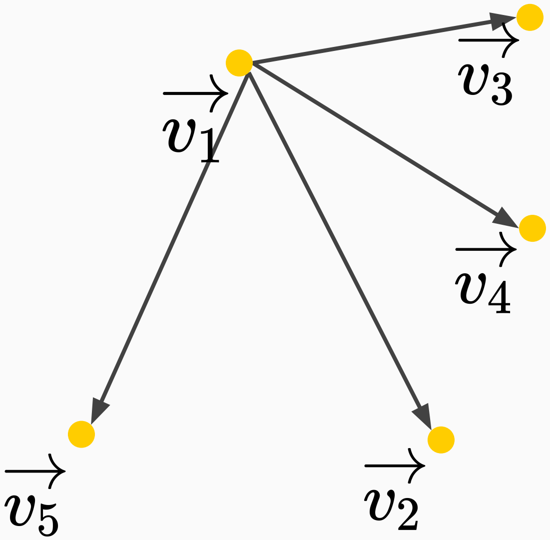
¿ ? ?

Probabilidad para vecinos

Probabilidad de que i sea vecino de j

$$p_{ij} = \frac{exp(-d_{ij}^2)}{\sum_k exp(-d_{ik}^2)}$$

Probabilidad para vecinos



$$d_{12}^2 = 5$$

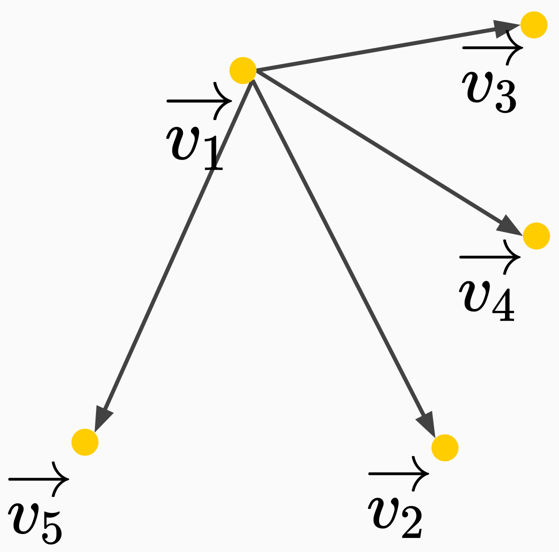
$$d_{13}^2 = 1.5$$

$$d_{14}^2 = 3$$

$$d_{15}^2 = 5$$

$$p_{ij} = \frac{exp(-d_{ij}^2)}{\sum_k exp(-d_{ik}^2)}$$

Probabilidad para vecinos



$$d_{12}^2 = 5$$

$$d_{13}^2 = 1.5$$

$$d_{14}^2 = 3$$

$$d_{15}^2 = 5$$

$$p_{ij} = \frac{exp(-d_{ij}^2)}{\sum_k exp(-d_{ik}^2)}$$

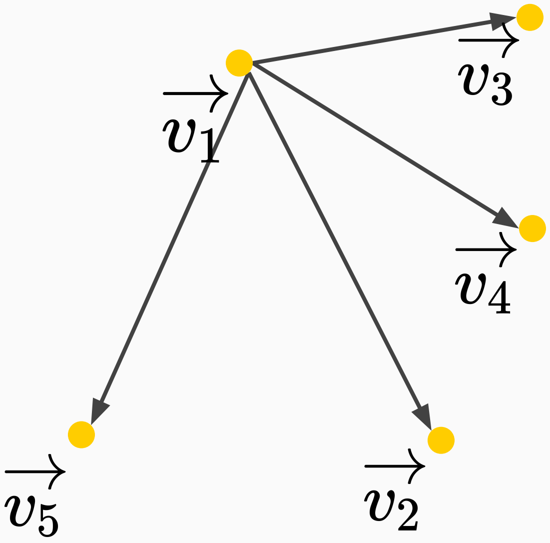
$$p_{12} = 0.02353$$

$$p_{13} = 0.77910$$

$$p_{14} = 0.17384$$

$$p_{15} = 0.02353$$

Probabilidad para vecinos



$$d_{12}^2 = 2.5$$

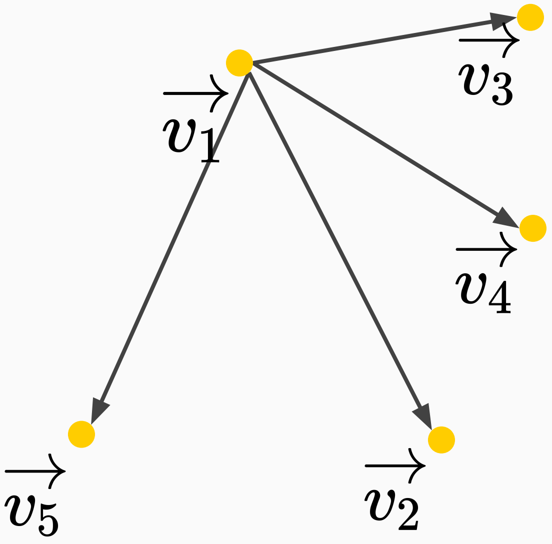
$$d_{13}^2 = 0.75$$

$$d_{14}^2 = 1.5$$

$$d_{15}^2 = 2.5$$

$$p_{ij} = \frac{exp(-d_{ij}^2)}{\sum_k exp(-d_{ik}^2)}$$

Probabilidad para vecinos



$$d_{12}^2 = 2.5$$

$$d_{13}^2 = 0.75$$

$$d_{14}^2 = 1.5$$

$$d_{15}^2 = 2.5$$

$$p_{ij} = \frac{exp(-d_{ij}^2)}{\sum_k exp(-d_{ik}^2)}$$

$$p_{12} = 0.09548$$

$$p_{13} = 0.54948$$

$$p_{14} = 0.25955$$

$$p_{15} = 0.09548$$

Probabilidad para vecinos

$$d_{12}^2 = 5 \quad d_{14}^2 = 3$$
$$d_{13}^2 = 1.5 \quad d_{15}^2 = 5$$

$$d_{12}^2 = 2.5 \quad d_{14}^2 = 1.5$$
$$d_{13}^2 = 0.75 \quad d_{15}^2 = 2.5$$

$$p_{12} = 0.02353$$
$$p_{13} = 0.77910$$
$$p_{14} = 0.17384$$
$$p_{15} = 0.02353$$

$$p_{12} = 0.09548$$
$$p_{13} = 0.54948$$
$$p_{14} = 0.25955$$
$$p_{15} = 0.09548$$

Probabilidad para vecinos

$$d_{12}^2 = 5 \quad d_{14}^2 = 3$$
$$d_{13}^2 = 1.5 \quad d_{15}^2 = 5$$

$$d_{12}^2 = 2.5 \quad d_{14}^2 = 1.5$$
$$d_{13}^2 = 0.75 \quad d_{15}^2 = 2.5$$

$$p_{12} = 0.02353$$
$$p_{13} = 0.77910$$
$$p_{14} = 0.17384$$
$$p_{15} = 0.02353$$

Mientras más escalemos hacia abajo las disimilaridades, más se reparte la probabilidad.

$$p_{12} = 0.09548$$
$$p_{13} = 0.54948$$
$$p_{14} = 0.25955$$
$$p_{15} = 0.09548$$

Probabilidad para vecinos

$d_{12}^2 = 5 \quad d_{14}^2 = 3$

$d_{13}^2 = 1.5 \quad d_{15}^2 = 5$

$d_{12}^2 = 2.5 \quad d_{14}^2 = 1.5$

$d_{13}^2 = 0.75 \quad d_{15}^2 = 2.5$

$p_{12} = 0.02353$

$p_{13} = 0.77910$

$p_{14} = 0.17384$

$p_{15} = 0.02353$

$$d_{ij}^2 = \frac{||x_i - x_j||^2}{2\sigma_i^2}$$

$p_{12} = 0.09548$

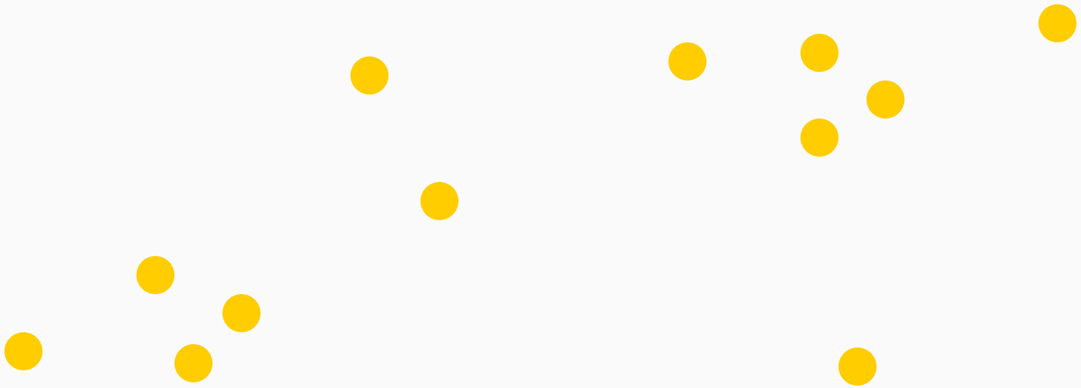
$p_{13} = 0.54948$

$p_{14} = 0.25955$

$p_{15} = 0.09548$

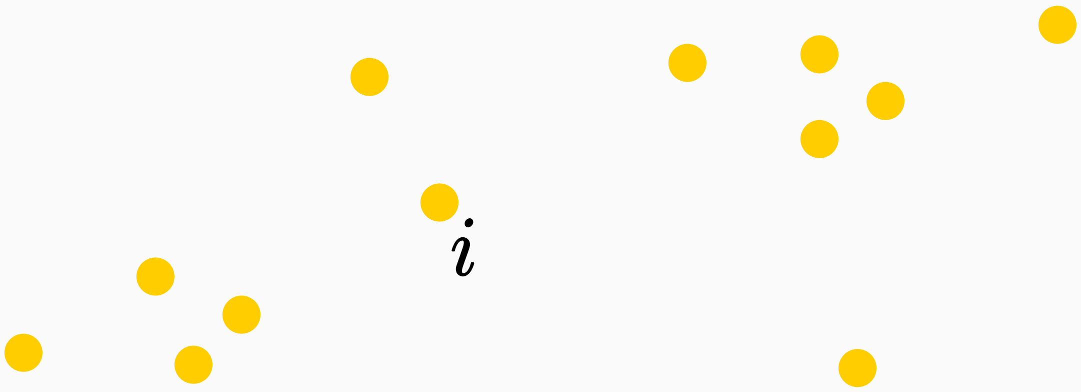
Perplexity

$$d_{ij}^2 = \frac{||x_i - x_j||^2}{2\sigma_i^2}$$



Perplexity

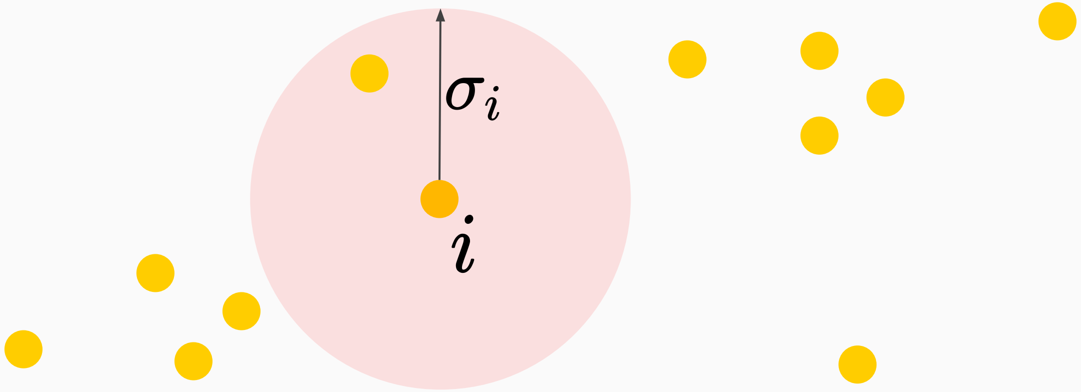
$$d_{ij}^2 = \frac{||x_i - x_j||^2}{2\sigma_i^2}$$



Perplexity

$$d_{ij}^2 = \frac{||x_i - x_j||^2}{2\sigma_i^2}$$

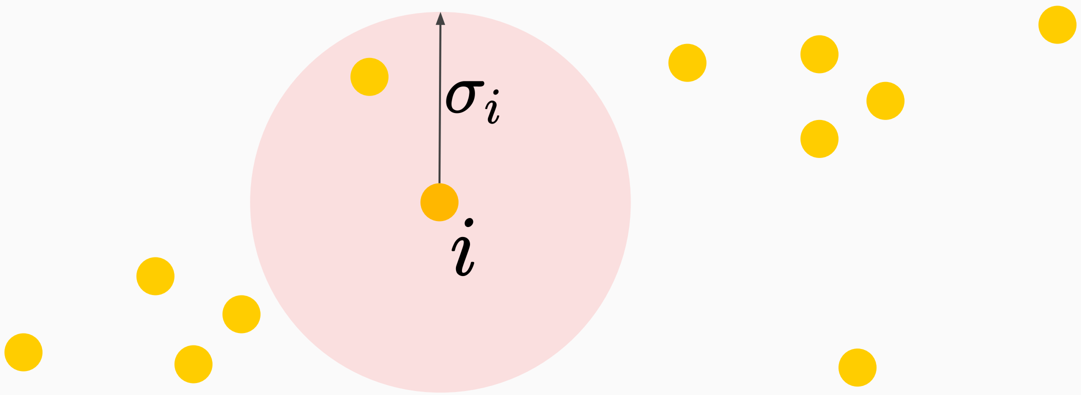
A mayor sigma mayor es la probabilidad que reservamos para aquellos que están lejos.



Perplexity

$$d_{ij}^2 = \frac{||x_i - x_j||^2}{2\sigma_i^2}$$

Queremos ser justos con las distintas densidades en todas las regiones. Hacemos que sigma para cada punto sea tal que *k* vecinos entren en sigma.

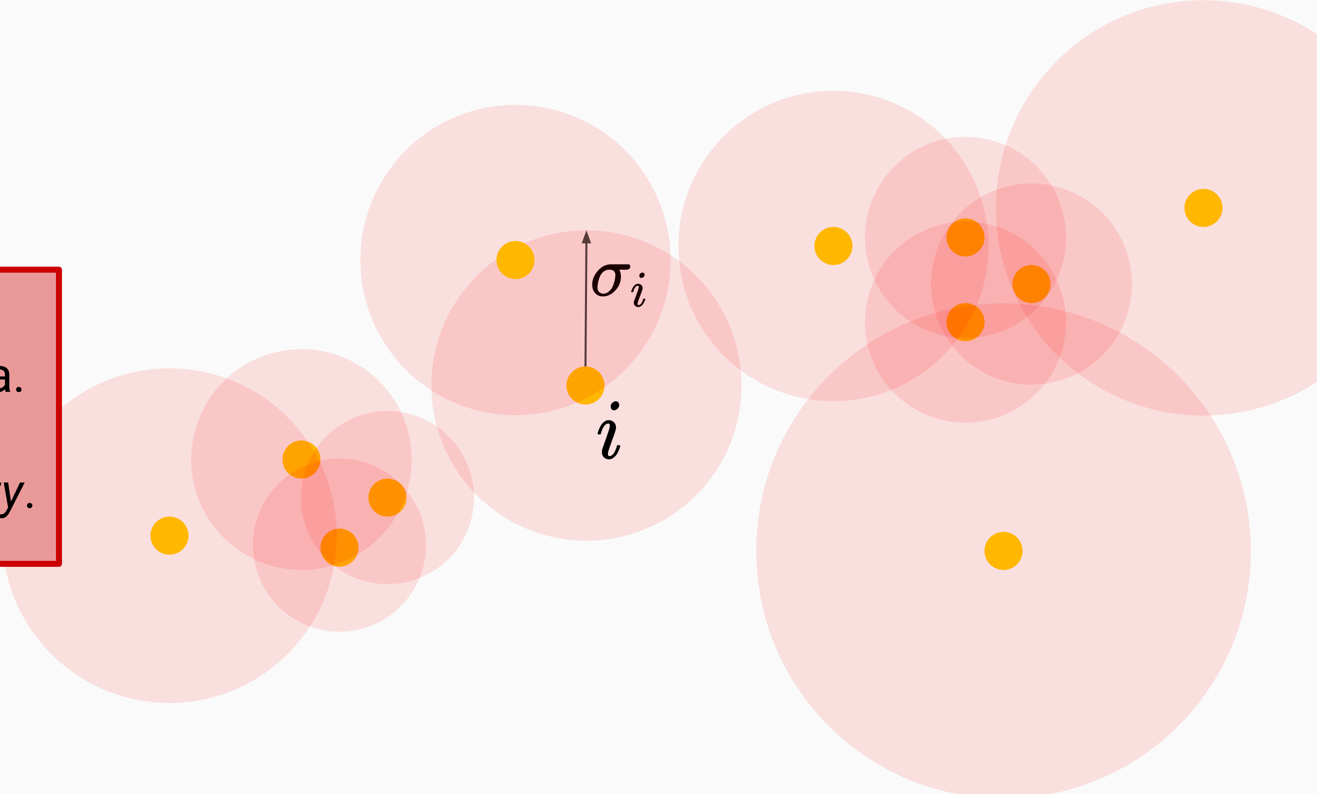


Perplexity

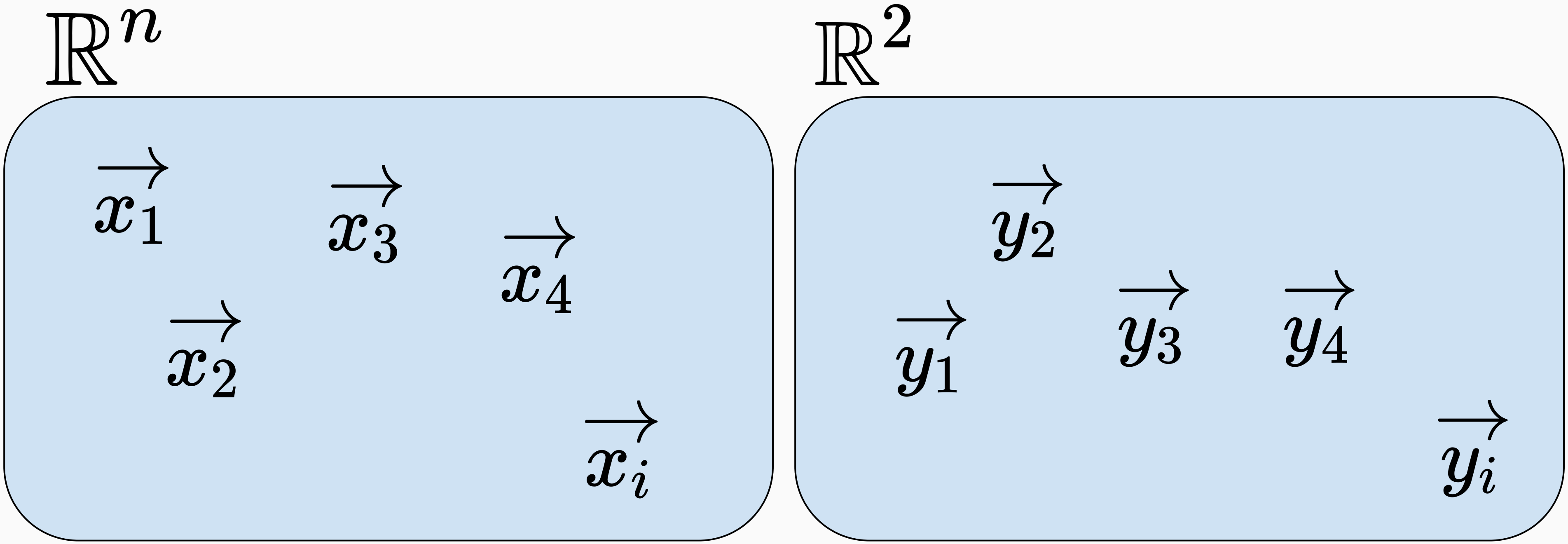
$$d_{ij}^2 = \frac{||x_i - x_j||^2}{2\sigma_i^2}$$

Para k=1 los sigmas se ven de la siguiente forma.

Llamaremos a k *perplexity*.

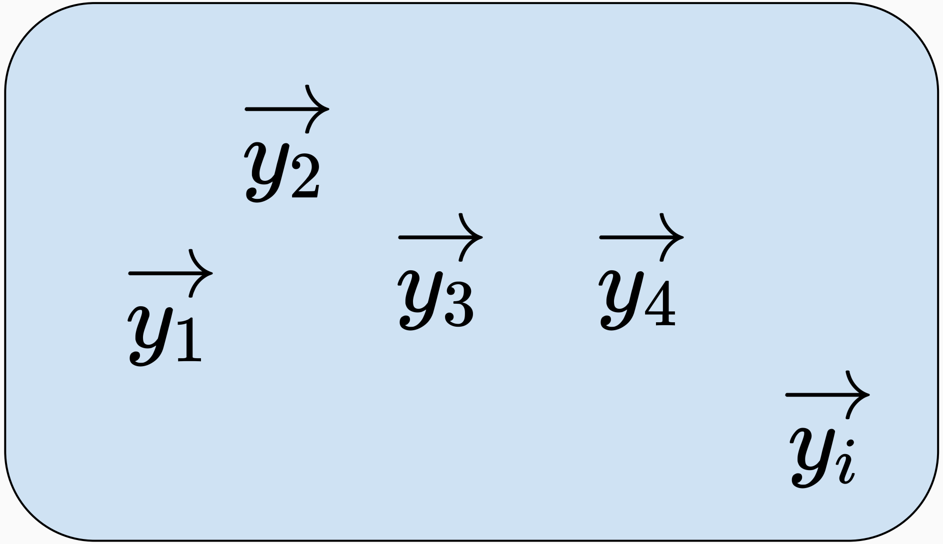


Construyendo Y

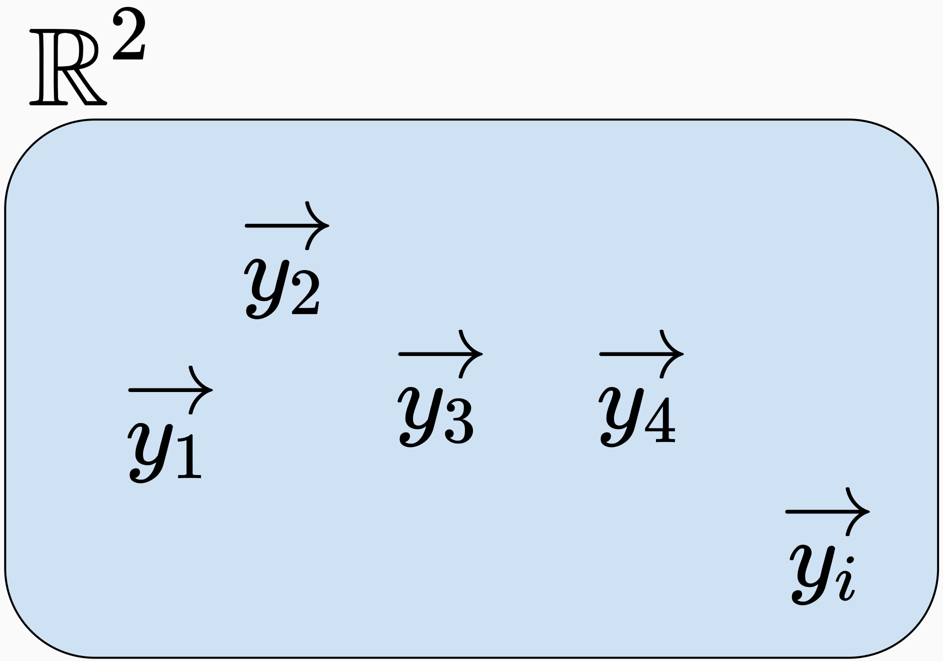


Construyendo Y

$\mathbb{R}^2$



Construyendo Y



Vamos a iniciar todos al azar 🤪

Construyendo Y

$$q_{ij} = \frac{exp(-||y_i - y_j||^2)}{\sum_k exp(-||y_i - y_k||^2)}$$

Construyendo Y

$$q_{ij} = \frac{exp(-||y_i - y_j||^2)}{\sum_k exp(-||y_i - y_k||^2)}$$

Cómo arrancamos al azar estos números son cualquier cosa 🤪

Construyendo Y

Tenemos:

$$p_{ij} = \frac{exp(-d_{ij}^2)}{\sum_k exp(-d_{ik}^2)}$$

$$q_{ij} = \frac{exp(-||y_i - y_j||^2)}{\sum_k exp(-||y_i - y_k||^2)}$$

Construyendo Y

Tenemos:

$$p_{ij} = \frac{exp(-d_{ij}^2)}{\sum_k exp(-d_{ik}^2)}$$

$$q_{ij} = \frac{exp(-||y_i - y_j||^2)}{\sum_k exp(-||y_i - y_k||^2)}$$

Construyendo Y

Tenemos:

$$p_{ij} = \frac{exp(-d_{ij}^2)}{\sum_k exp(-d_{ik}^2)}$$

$$q_{ij} = \frac{exp(-||y_i - y_j||^2)}{\sum_k exp(-||y_i - y_k||^2)}$$

Podemos usar la divergencia de Kullback-Leibler entre p y q para conocer masomenos que tan distintas son 🤓

Construyendo Y

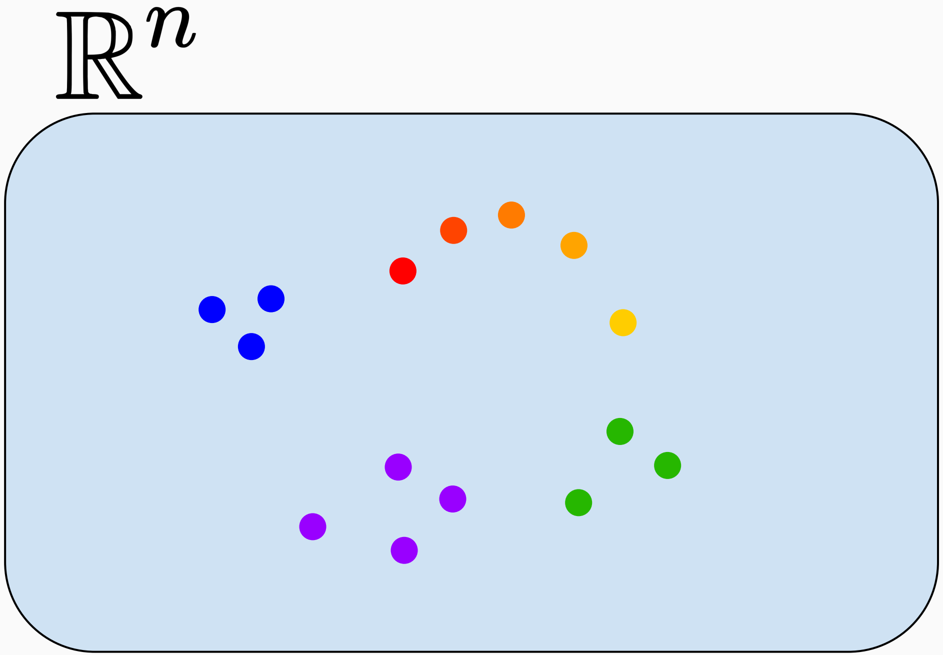
Tenemos:

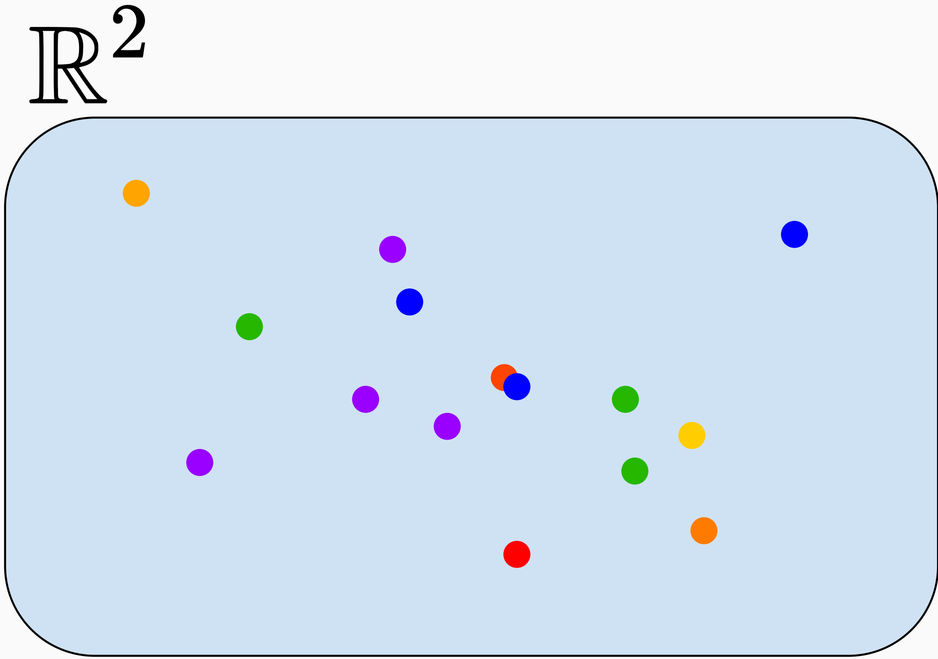
$$p_{ij} = \frac{exp(-d_{ij}^2)}{\sum_k exp(-d_{ik}^2)}$$

$$q_{ij} = \frac{exp(-||y_i - y_j||^2)}{\sum_k exp(-||y_i - y_k||^2)}$$

```
q = calcular_qs()
while(not satisfechos):
    with KL(p,q) as error:
        y = mejorar(y, error)
    q = calcular_qs()
```

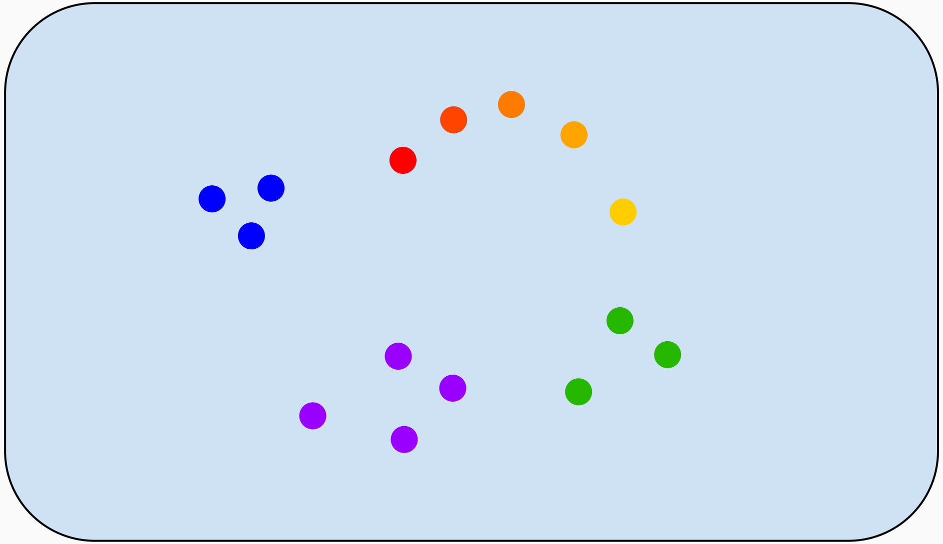
¿Y cómo se ve?



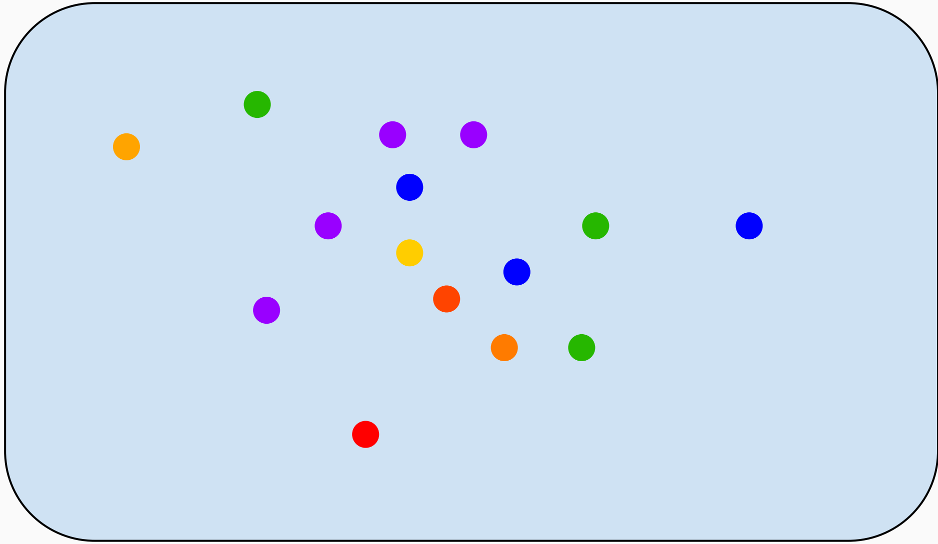


¿Y cómo se ve?

$\mathbb{R}^n$

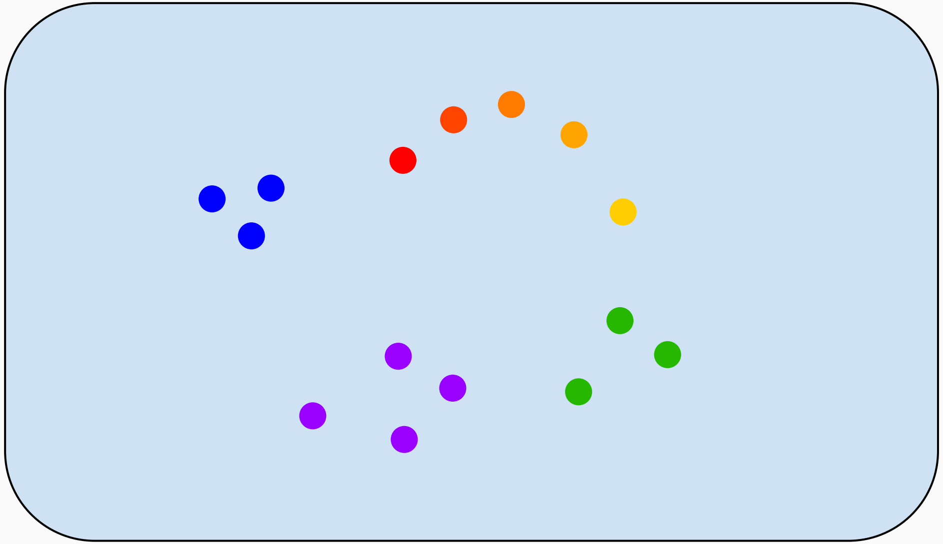


$\mathbb{R}^2$

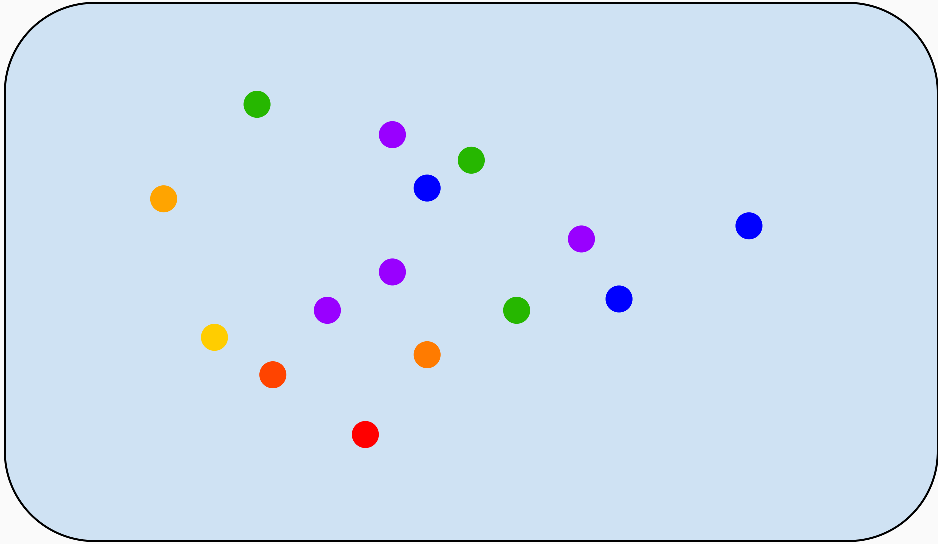


¿Y cómo se ve?

$\mathbb{R}^n$

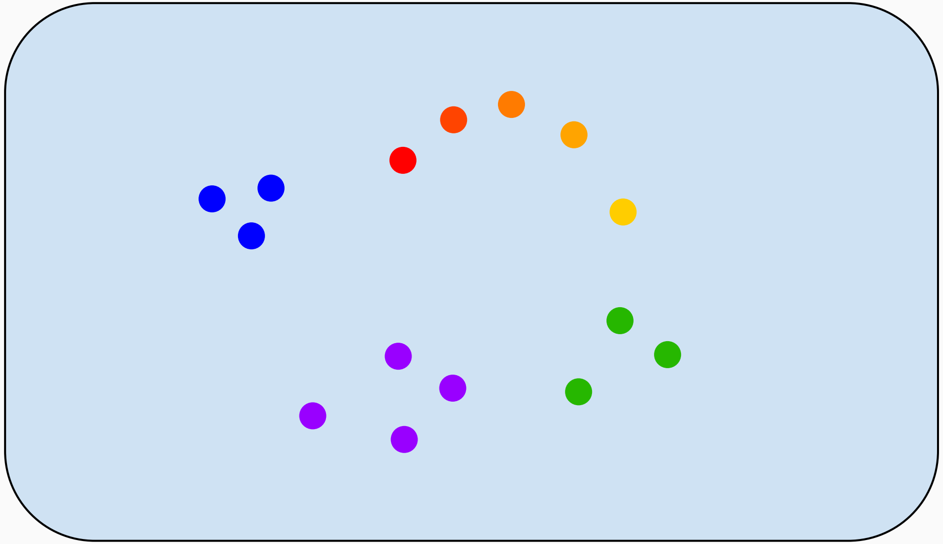


$\mathbb{R}^2$

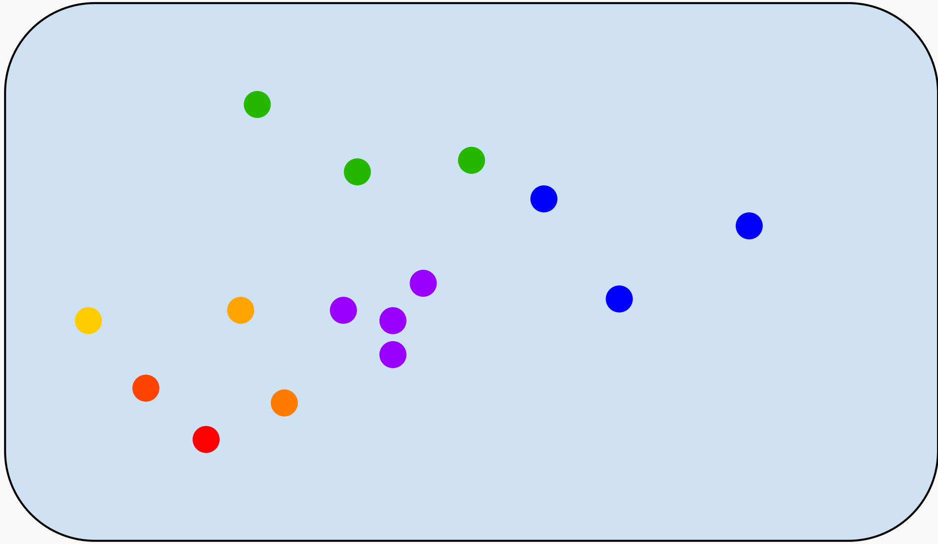


¿Y cómo se ve?

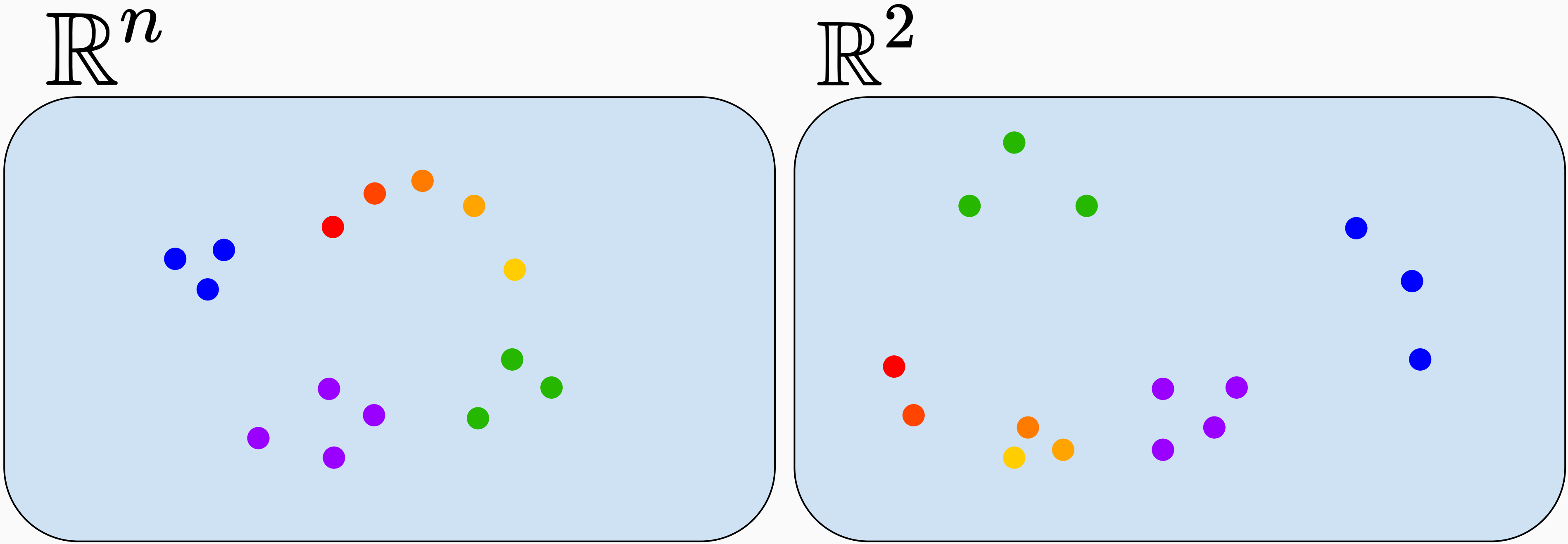
$\mathbb{R}^n$



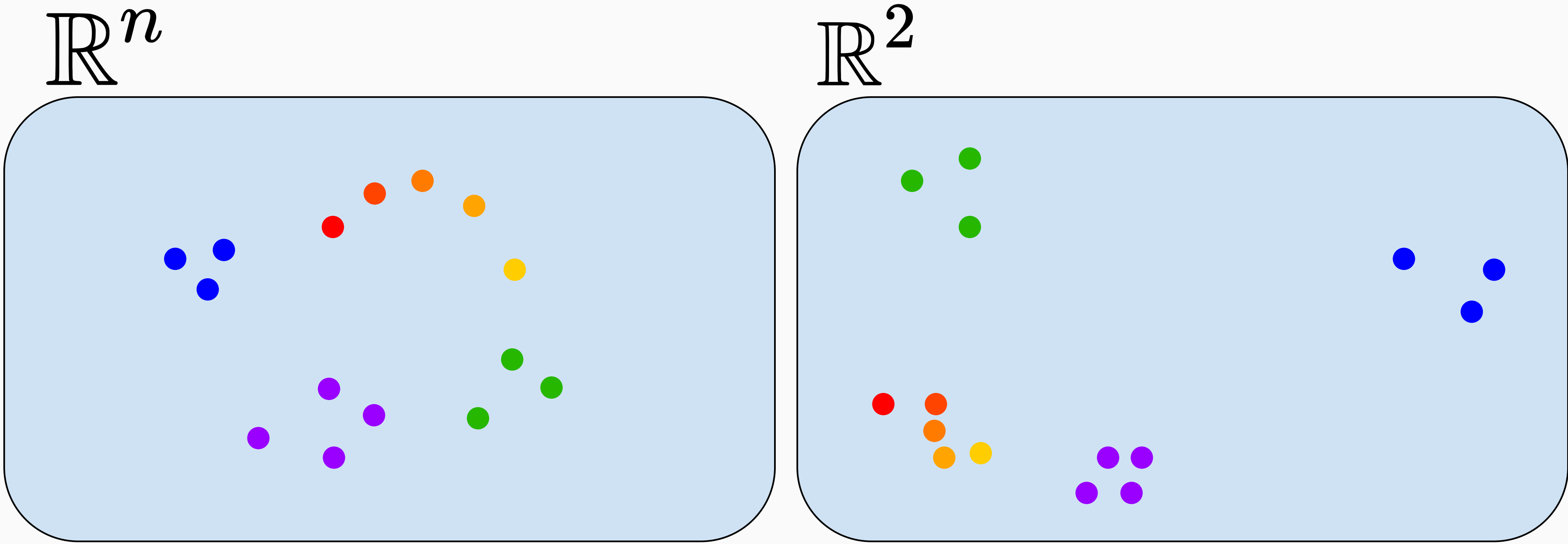
$\mathbb{R}^2$



¿Y cómo se ve?



¿Y cómo se ve?



¿Y entonces qué es t-SNE?

Visualizing Data using t-SNE

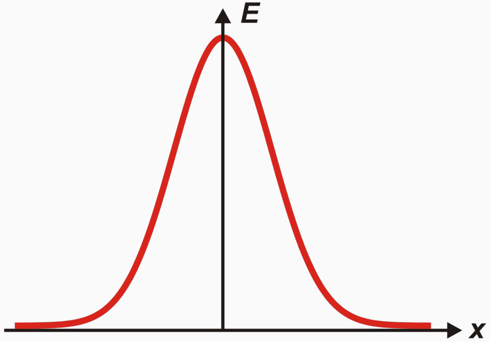
Laurens van der Maaten LVDMAATEN@GMAIL.COM
TiCC
Tilburg University
P.O. Box 90153, 5000 LE Tilburg, The Netherlands

Geoffrey Hinton HINTON@CS.TORONTO.EDU
Department of Computer Science
University of Toronto
6 King's College Road, M5S 3G4 Toronto, ON, Canada

Editor: Yoshua Bengio

¿Y entonces qué es t-SNE?

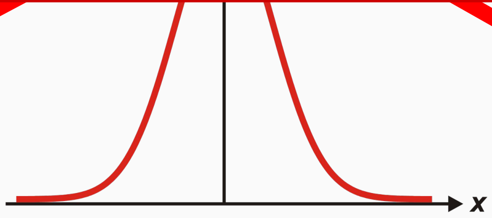
$$q_{ij} = \frac{exp(-||y_i - y_j||^2)}{\sum_k exp(-||y_i - y_k||^2)}$$



¿Y entonces qué es t-SNE?

$$q_{ij} = \cancel{\frac{exp(-||y_i - y_j||^2)}{\sum \ldots exp(\ldots ||^2)}}$$

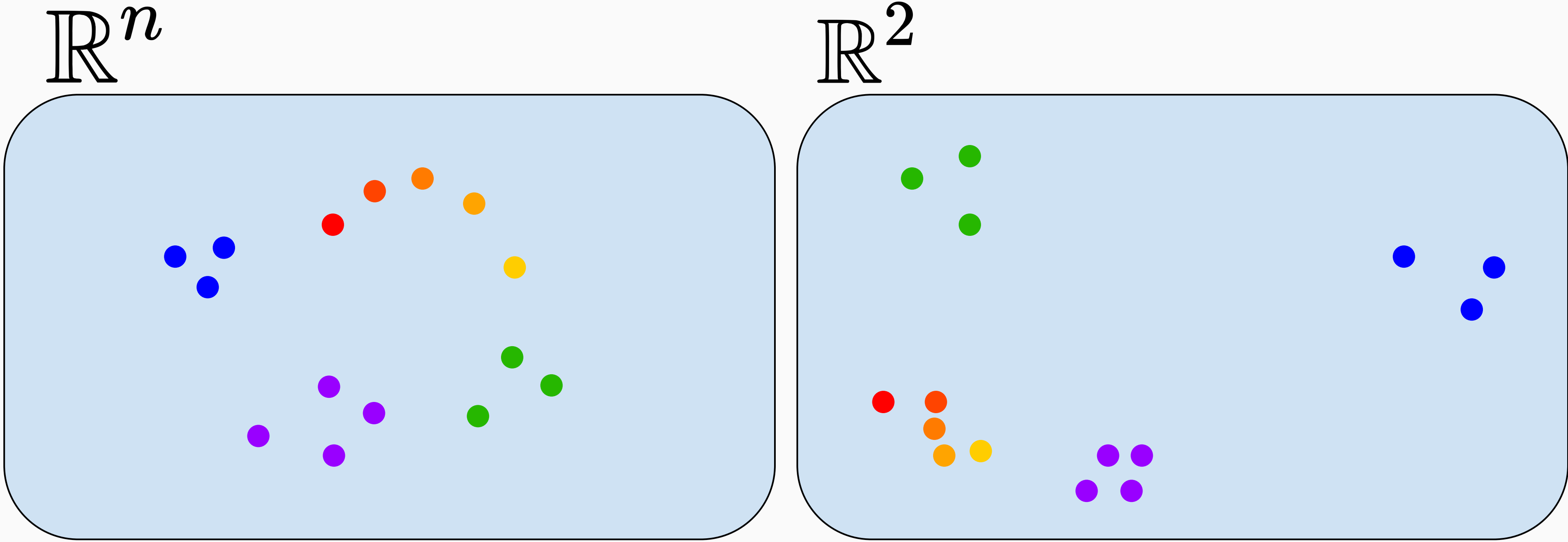
Modelamos las distancias usando algo llamado t-Student, es muy similar...



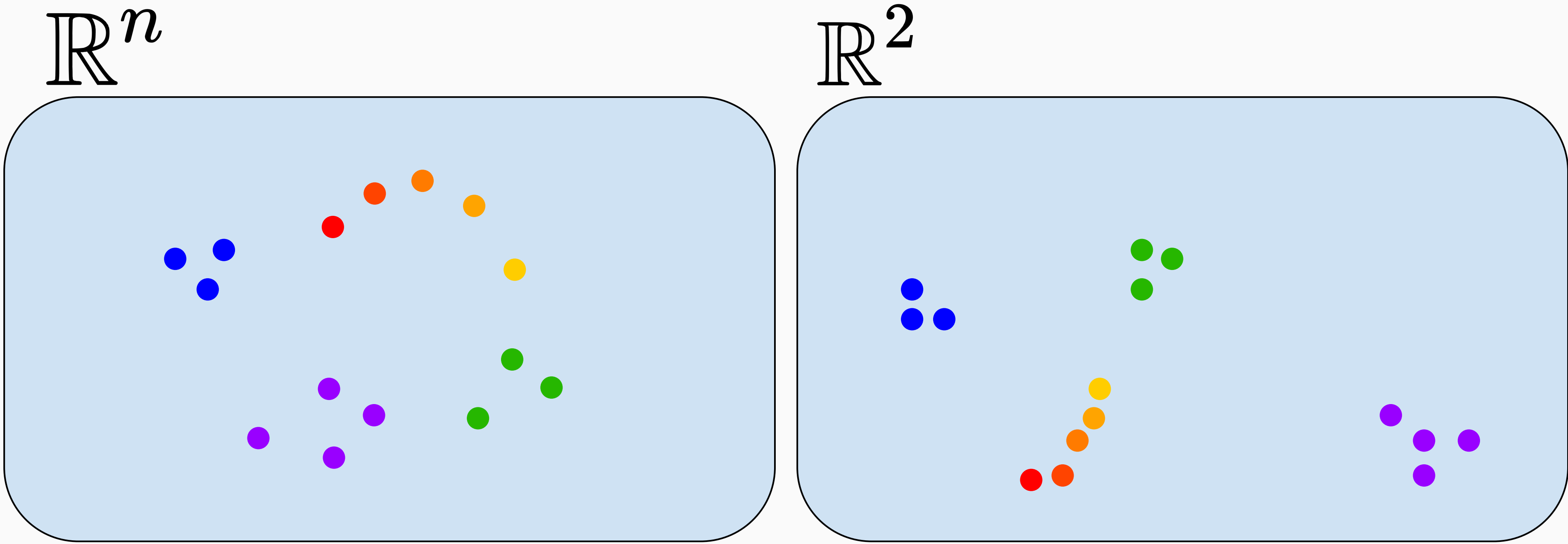
¿Y entonces qué es t-SNE?

Además hacemos que las disimilaridades sean simétricas

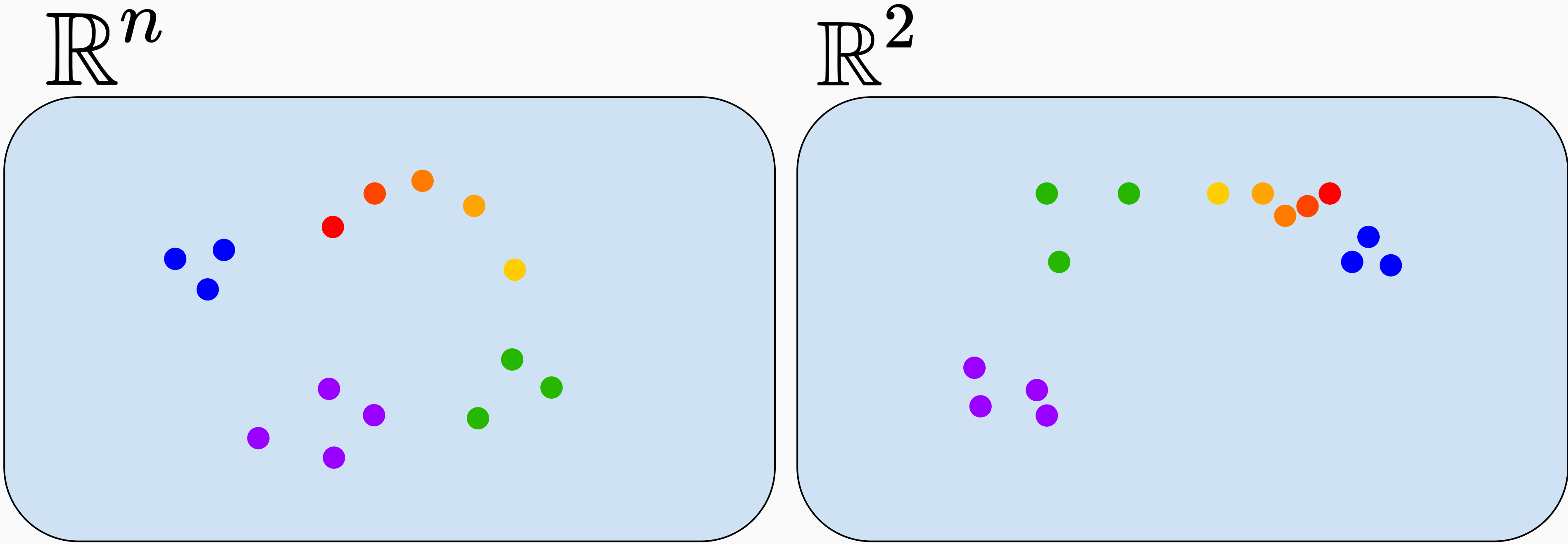
t-SNE es estocástico



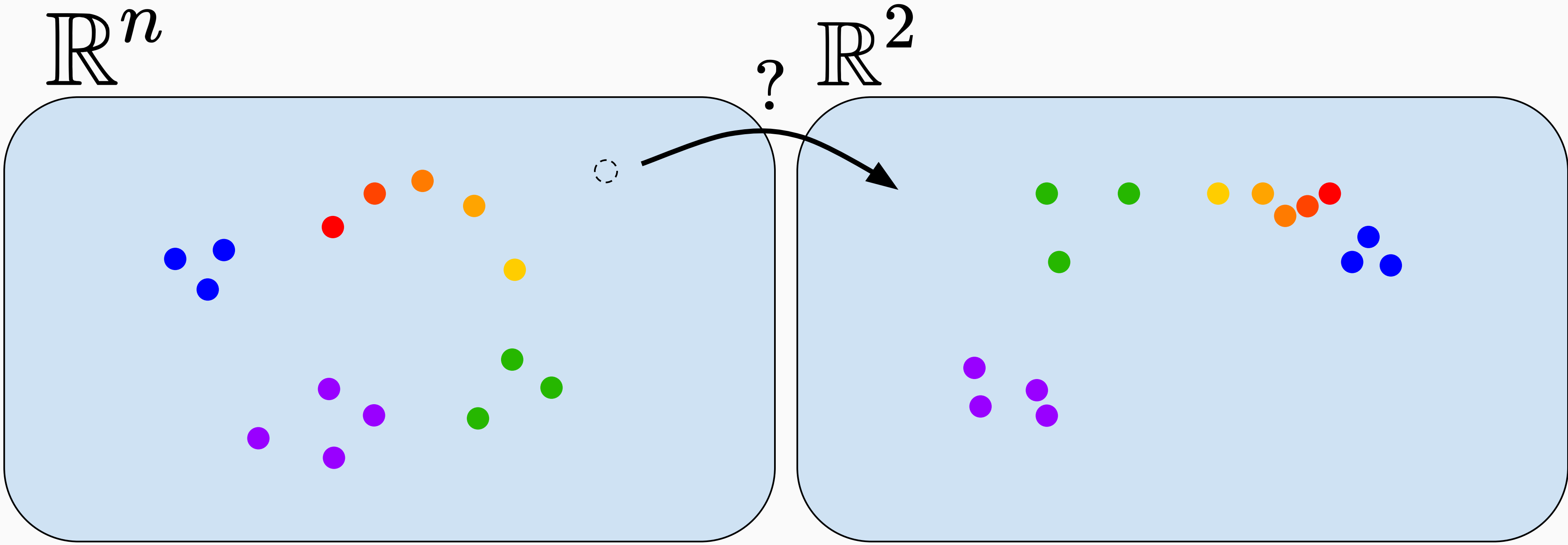
t-SNE es estocástico



t-SNE es estocástico



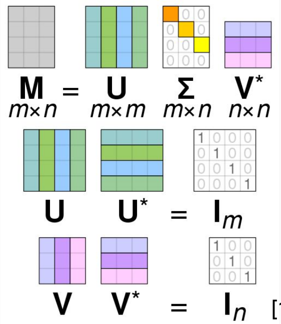
No se puede usar para nuevos puntos/predecir



Las distancias son caras

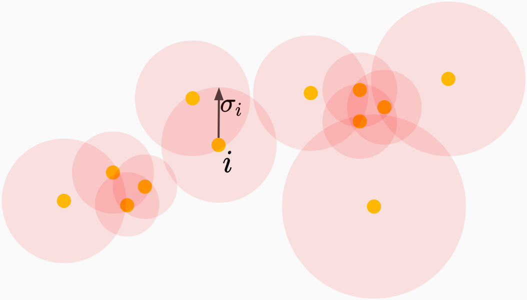
$$d_{ij}^2 = \frac{||x_i - x_j||^2}{2\sigma_i^2}$$

Las distancias son caras



[1]

$$k < n$$



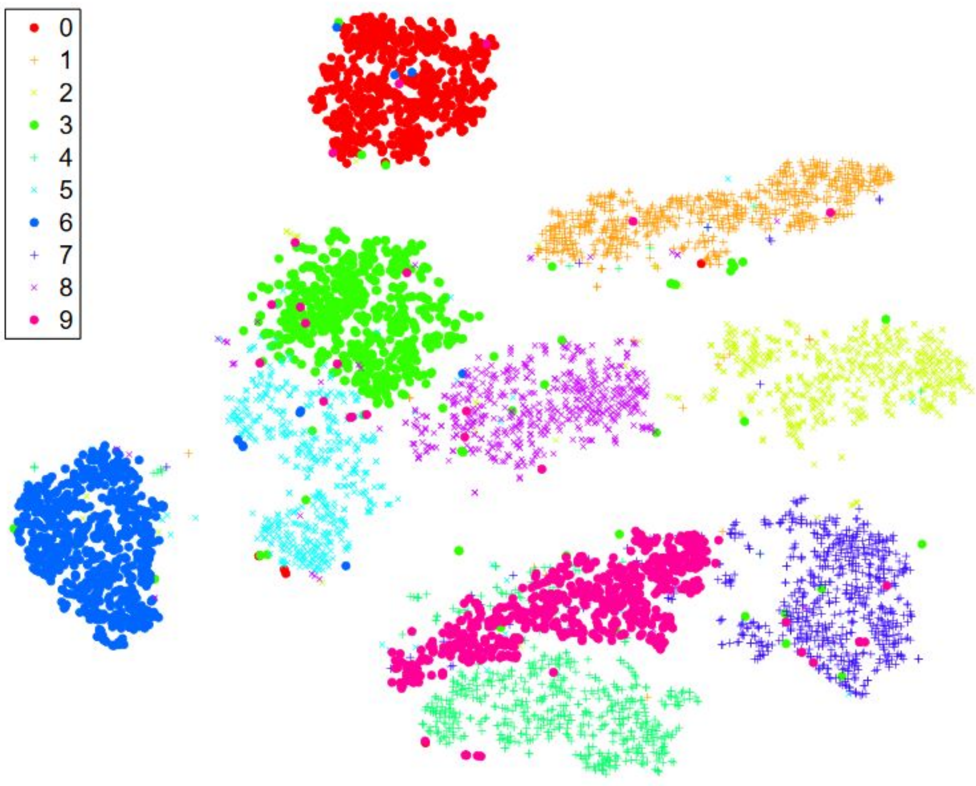
$$\mathbb{R}^n \qquad \mathbb{R}^k \qquad \mathbb{R}^2$$

$$\boxed{\overrightarrow{x_i}} \longrightarrow \boxed{\overrightarrow{x'_i}} \longrightarrow \boxed{\overrightarrow{y_i}}$$

[1] https://commons.wikimedia.org/wiki/File:Singular_value_decomposition_visualisation.svg

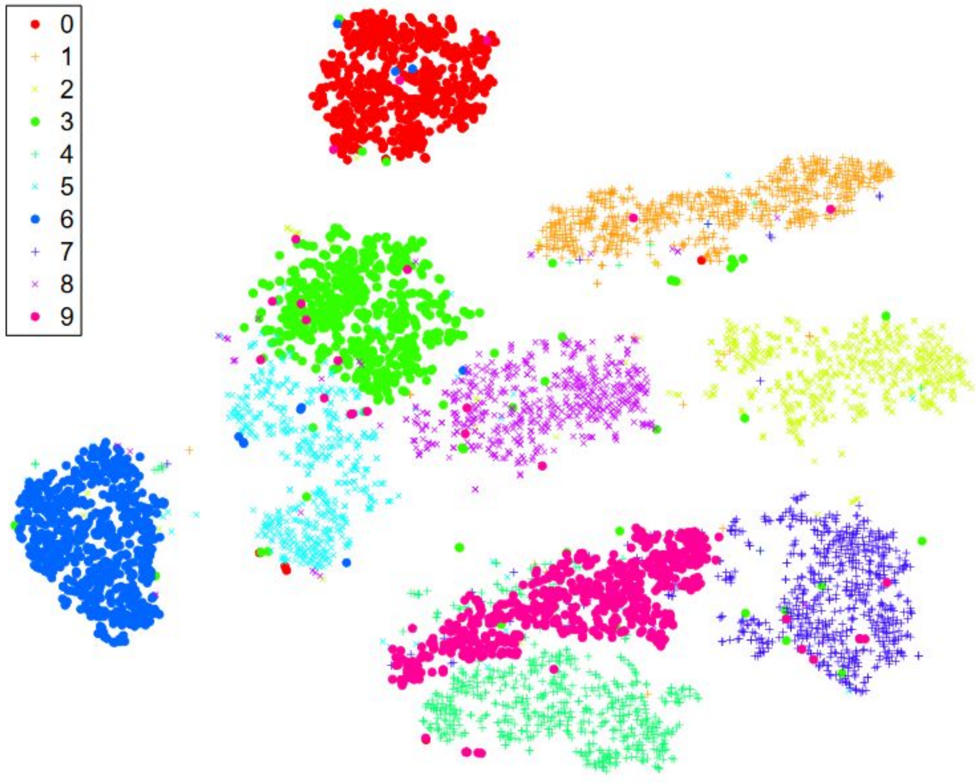
Ejemplo MNIST

Ejemplo MNIST



(a) Visualization by t-SNE.

Ejemplo MNIST



(a) Visualization by t-SNE.

Referencias

- Hinton, G. E., & Roweis, S. T. (2003). Stochastic neighbor embedding. In *Advances in neural information processing systems* (pp. 857-864).

- Maaten, L. V. D., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research*, *9*(Nov), 2579-2605.

- Sección 8.7 del apunte de la cátedra