



t-SNE: t-distributed stochastic neighbor embedding

Ing. Juan M. Rodríguez

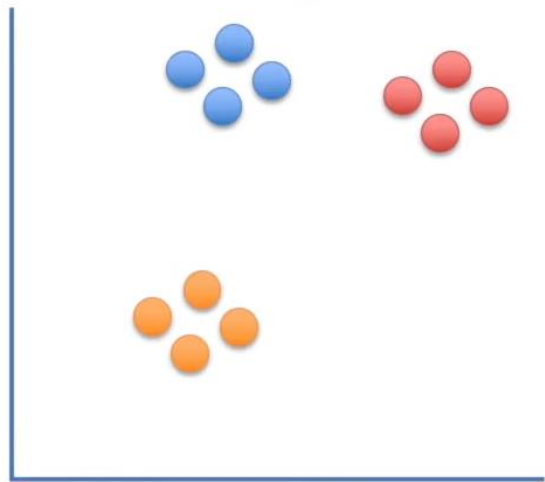
t-SNE



Se trata de otro método que al igual que PCA, toma datos de un espacio de alta dimensión y los proyecta en un espacio de menor dimensión para que puedan ser representados.

Desarrollada en 2008 por Geoffrey Hinton y Laurens Van Der Maaten

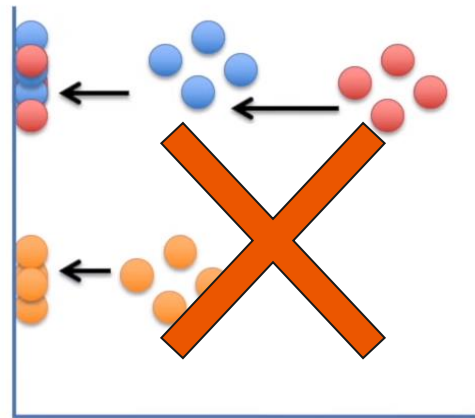
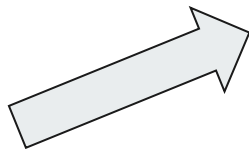
t-SNE



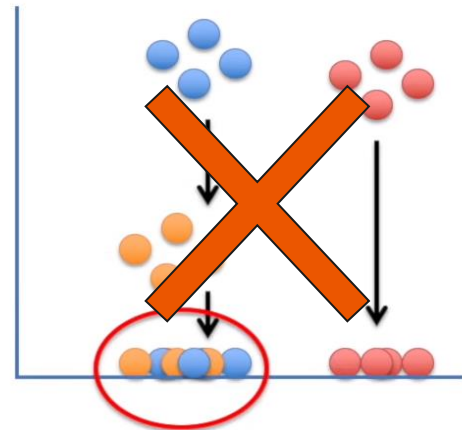
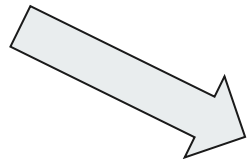
Datos en un espacio en 2D

Si solo los proyectamos en un eje....

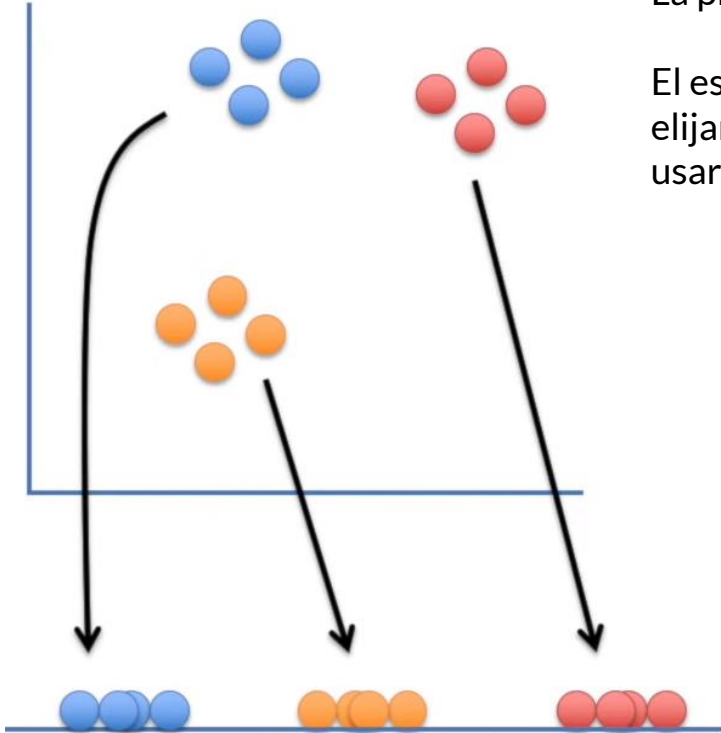
Eje Y



Eje X



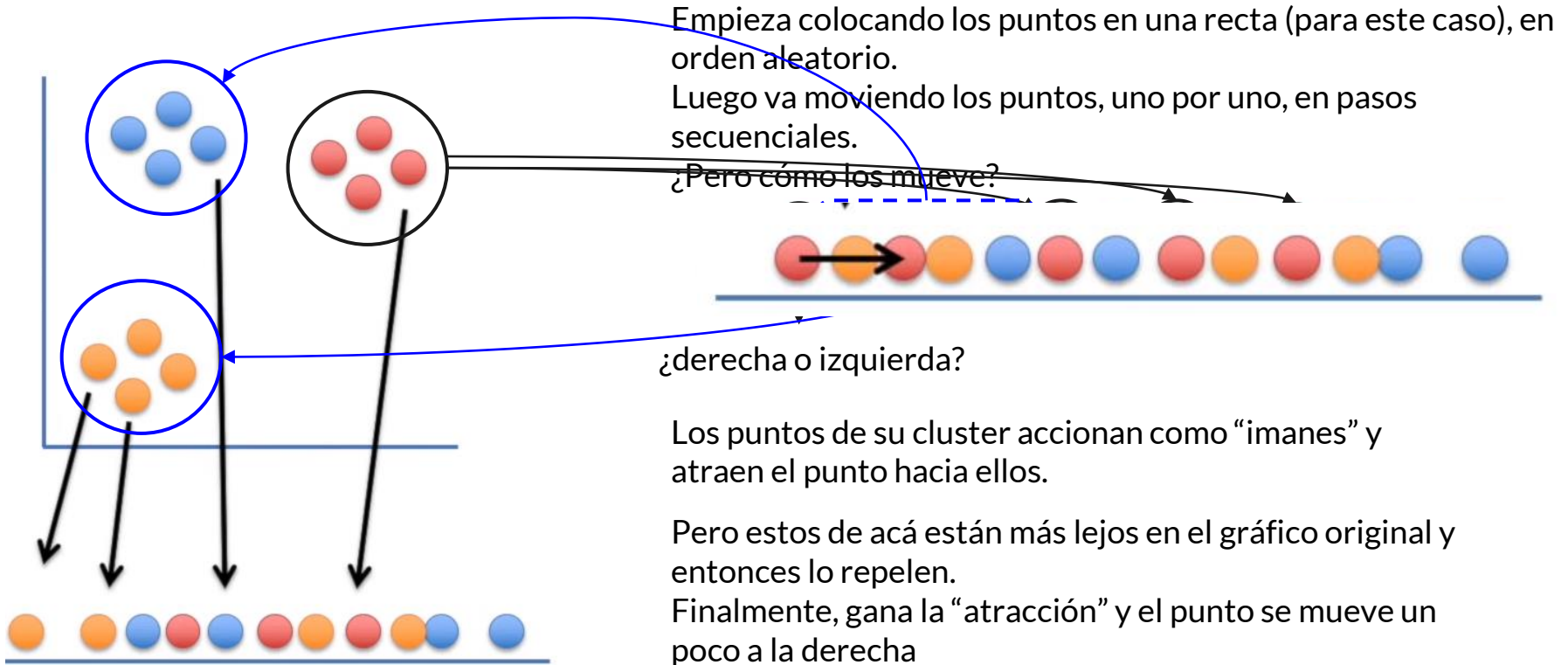
t-SNE



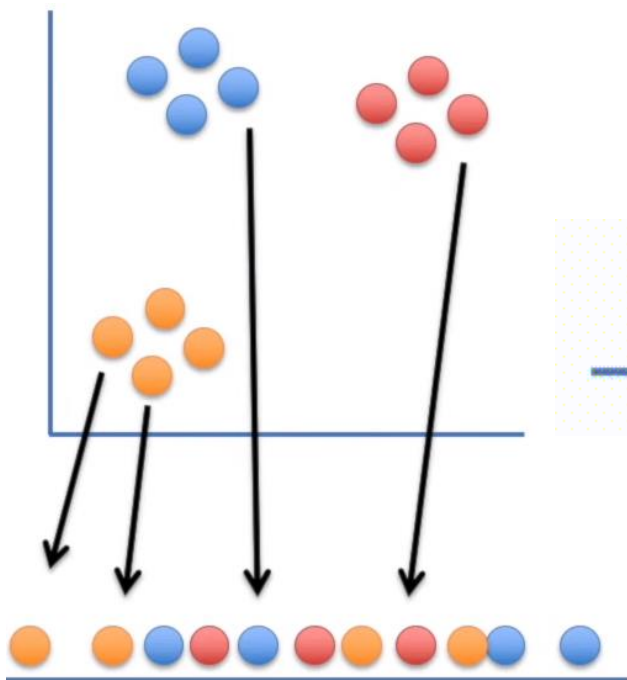
La proyección de t-SNE preserva los *clusters*

El espacio de menor dimensión puede ser cualquiera que elijamos. Habitualmente se utiliza un espacio con 2D, pero usaremos en este ejemplo un espacio unidimensional.

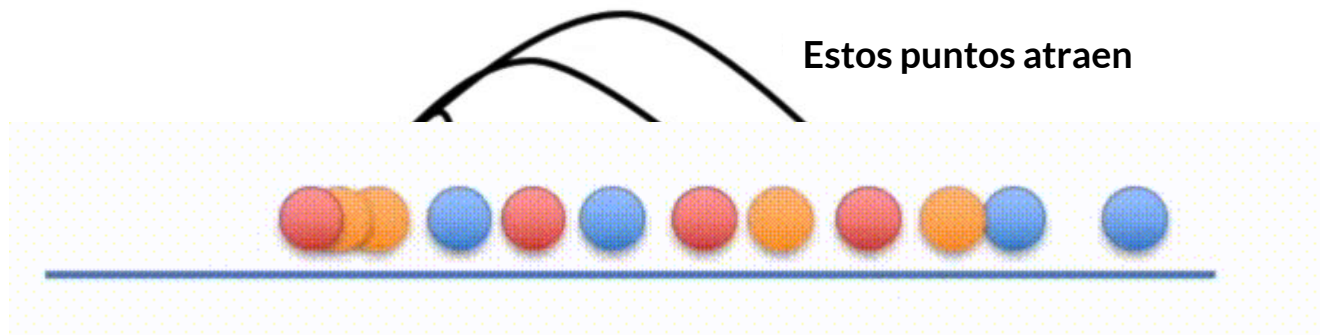
t-SNE



t-SNE



Sigue con el próximo punto en el gráfico



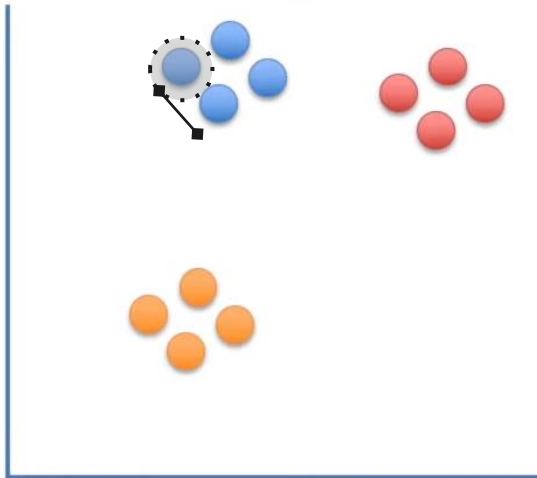
Estos puntos atraen

Y este lo repele

¿Cómo funciona internamente?

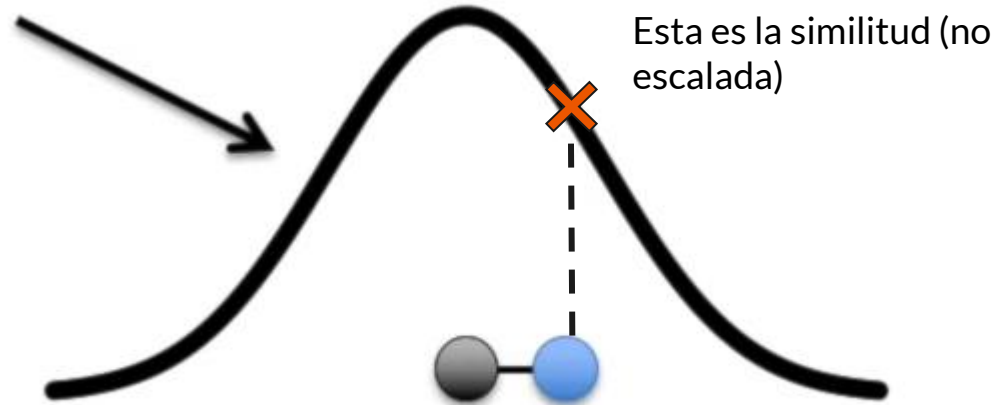
t-SNE

1. Determinar la similitud de todos los puntos el gráfico



1.a tomamos el primer punto, (punto de interés) y medimos la distancia al siguiente punto.

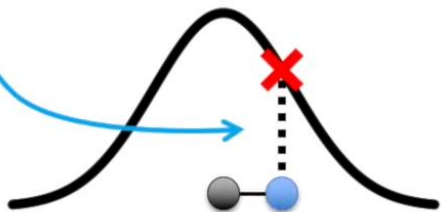
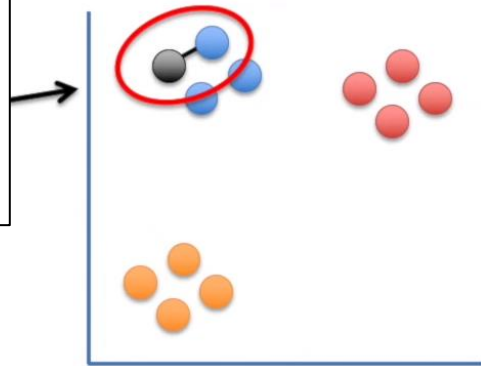
1.b luego la proyectamos sobre una curva **Normal** (campana de Gauss), centrada en el punto de interés



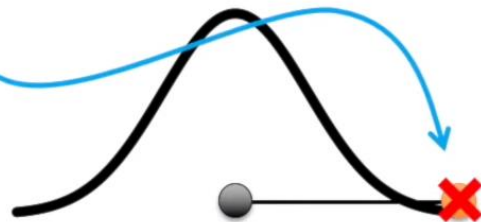
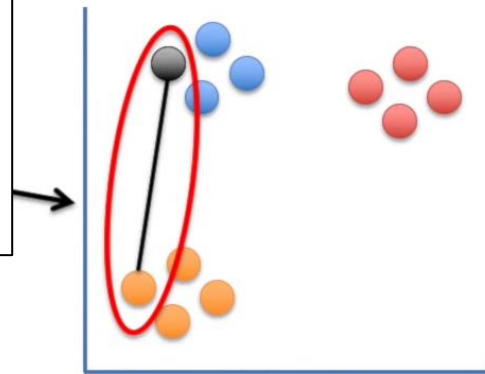
t-SNE



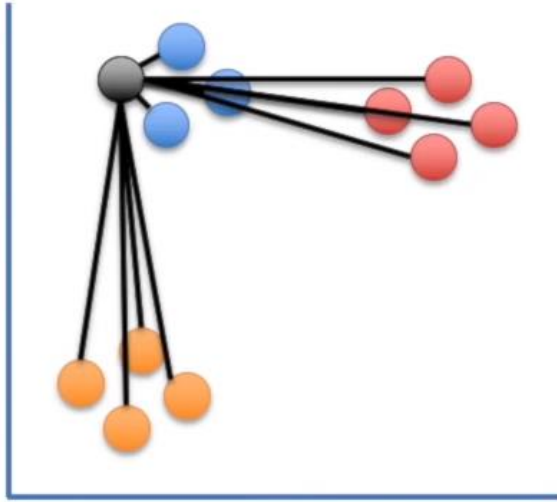
Continuamos con el siguiente punto y calculamos la similitud, de igual forma



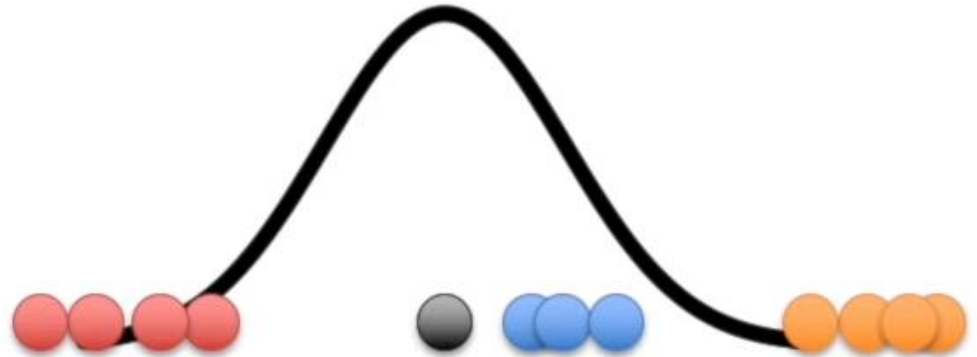
Luego con este otro punto...



t-SNE



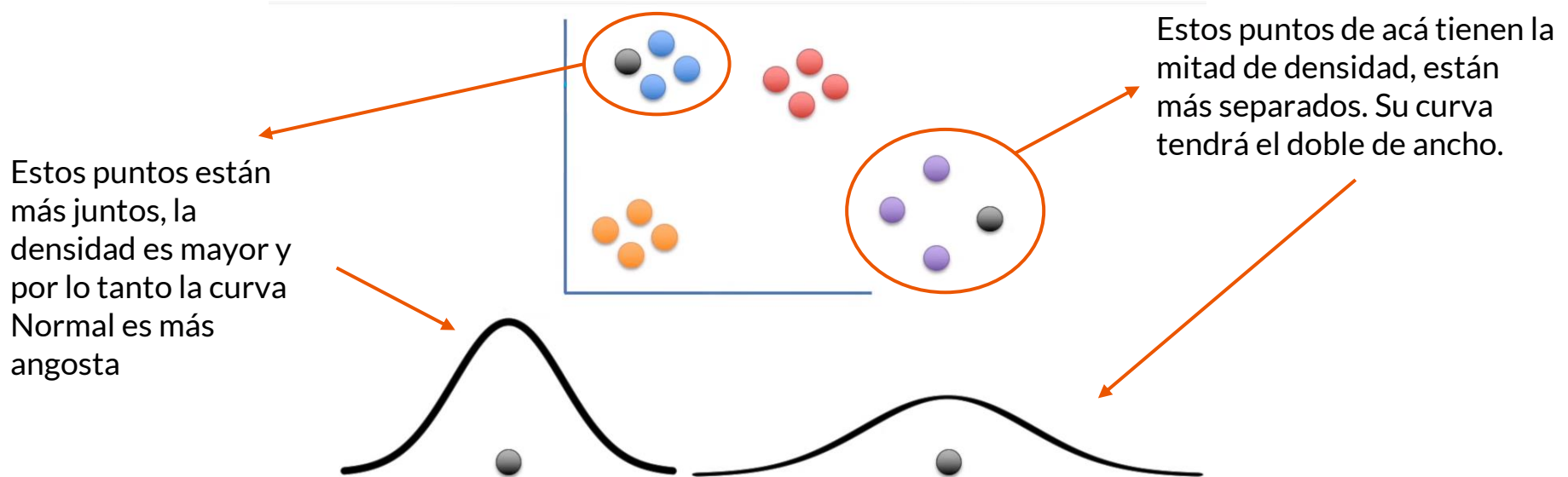
Y los proyectamos en
la misma curva.



Finalmente medimos la distancia a
todos los puntos

t-SNE

El ancho de la curva Normal, (su desvío estándar) NO es siempre igual. Depende de la densidad de los datos en la cercanía del punto de interés.



t-SNE



Cómo las curvas no son iguales, hay normalizar la similitud:

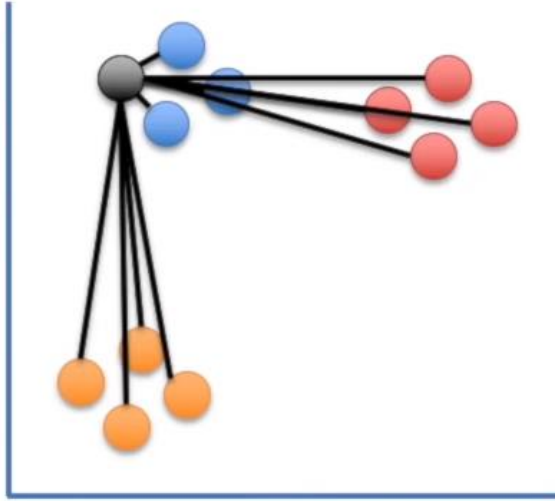
$$\frac{\text{Similitud}}{\text{Suma de todas las similitudes}} = \text{Similitud escalada} = \text{puntaje (score)}$$

En realidad, t-SNE tiene un parámetro llamado “perplejidad” (*perplexity*) que es igual a la densidad esperada para un punto.

La desviación estándar se define por este valor de **perplejidad** que corresponde al número de vecinos alrededor de cada punto. Este valor lo establece el usuario de antemano y permite estimar la desviación estándar de las distribuciones gaussianas definidas para cada punto x_i .

Cuanto mayor es la perplejidad, mayor es la variación.

t-SNE

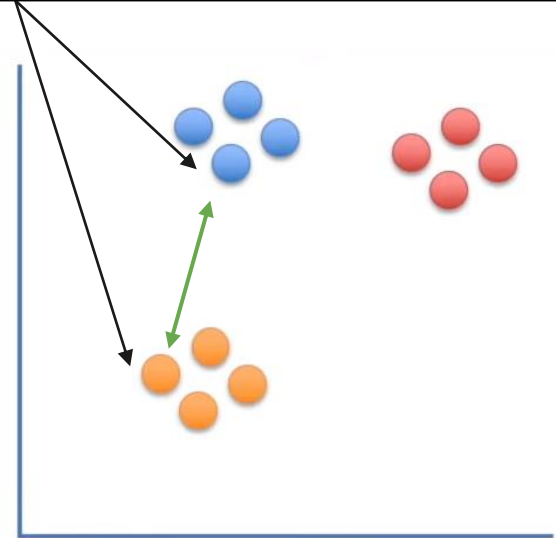


Una vez que calculamos todos los puntajes de similitud o *similarity scores* t-SNE hace un último ajuste

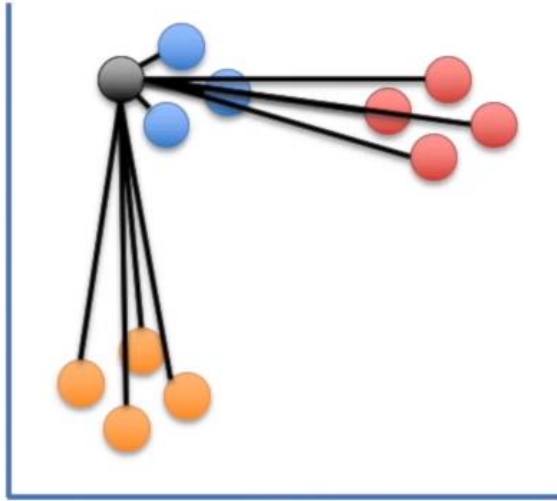
Cómo el ancho de la curva Normal usada dependen de los puntos que rodean al mismo

El puntaje de similitud del punto amarillo, respecto del azul, puede que no sea el mismo que el puntaje de similitud del punto azul respecto del amarillo

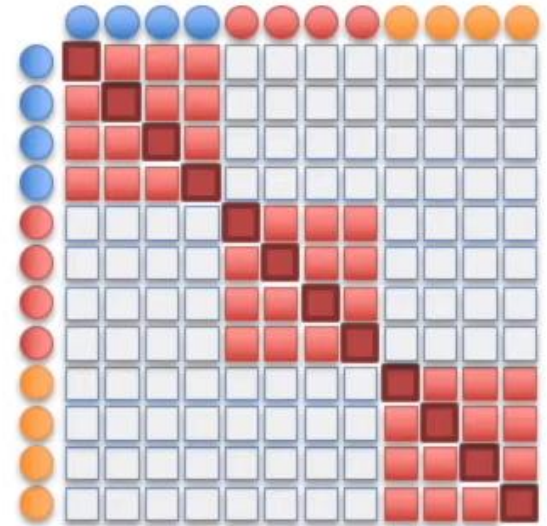
t-SNE promedia los dos puntajes de similitud, en ambas direcciones. Para unificar el valor.



t-SNE



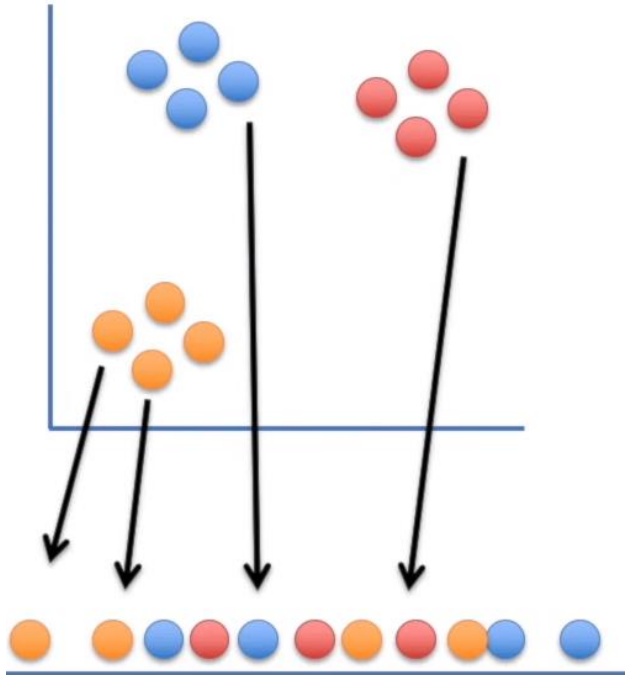
Finalmente tendremos una matriz de puntajes de similitud o *Similarity Scores*



■ = High similarity

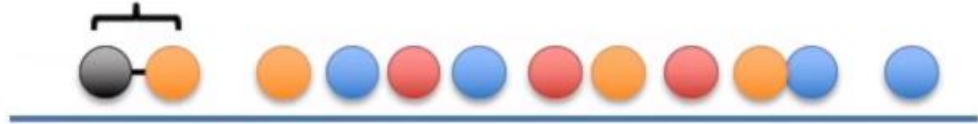
□ = Low similarity

t-SNE

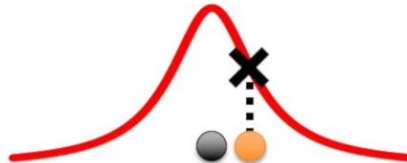


Ahora que tenemos la matriz de similitud, hacemos lo que dijimos al principio. Proyectamos los puntos del espacio en una recta de forma **aleatoria**.

El próximo paso es calcular el puntaje de similitud en el nuevo espacio.
Tomamos el primer punto, medimos la distancia al siguiente punto...

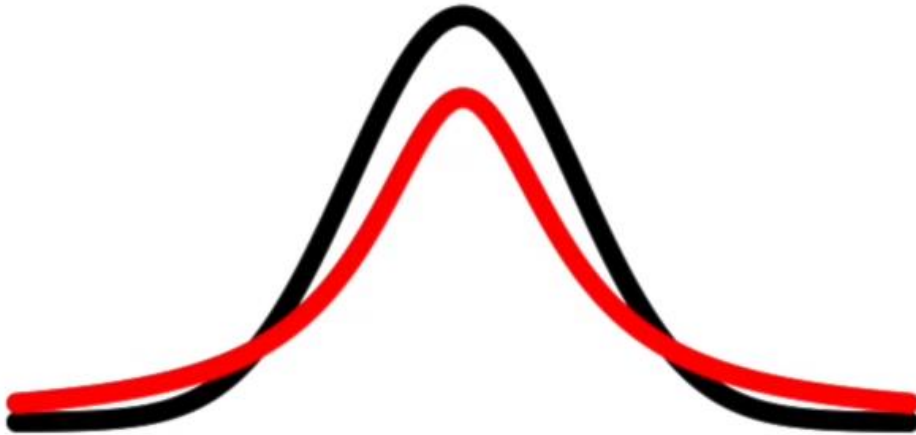


Luego lo proyectamos en una curva, centrada en el punto de interés y vemos su valor...



Pero esta NO es una curva Normal, ¡es una distribución t-Student!

t-SNE

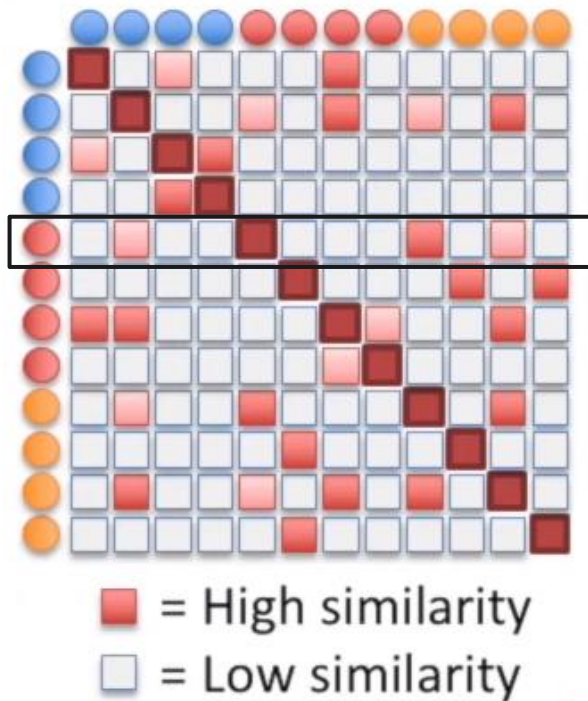


La distribución T (t-Student) (en rojo) es un poco más baja en el centro y más alta en los extremos que una distribución **Normal**.

Es la “T” en t-SNE

Se usa esta distribución en lugar de la normal ya que permite una mejor visualización de los datos, estos aparecen un poco más “dispersos”

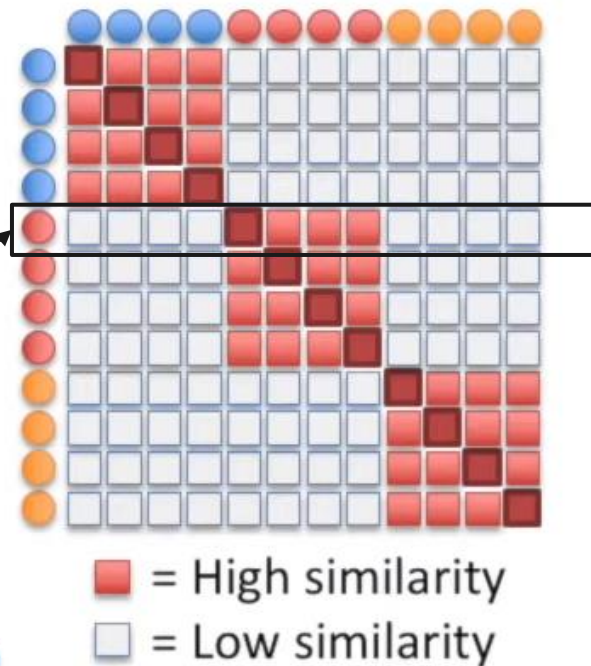
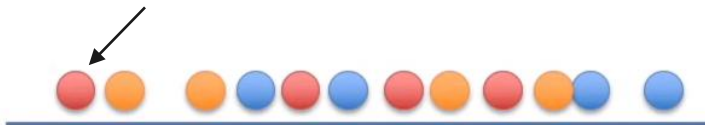
t-SNE



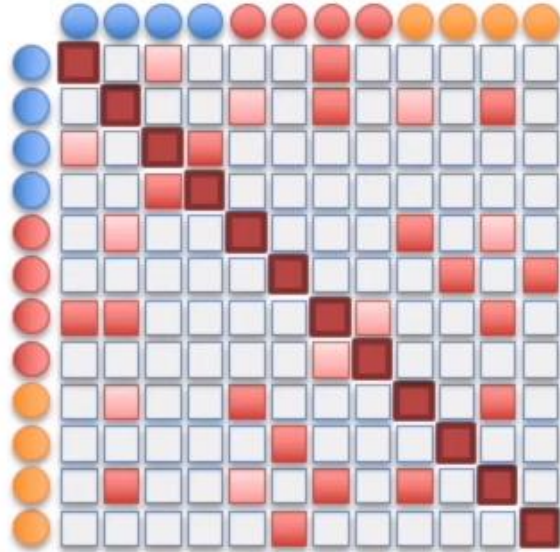
Al igual que antes, calculamos y normalizamos todos los nuevos puntajes y construimos una matriz. Solo que esta matriz no está ordenada en absoluto.

Tiene como objetivo que esta fila
Se parezca a esta otra fila

Mover este punto

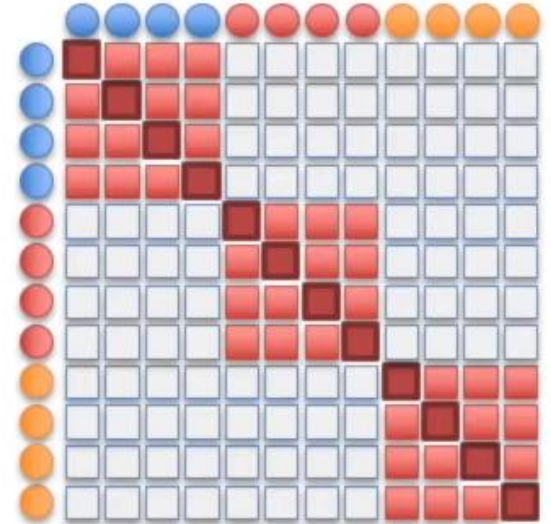


t-SNE



■ = High similarity
□ = Low similarity

No se puede calcular de una sola vez el orden de los puntos, es por ello que debe de hacerse en pequeños pasos y moviendo los puntos uno por uno, haciendo que de a poco la nueva matriz coincida con la original.

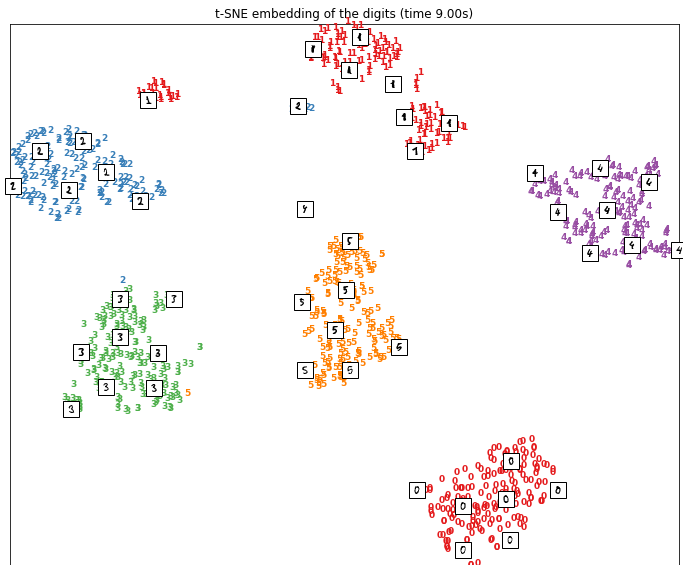


■ = High similarity
□ = Low similarity



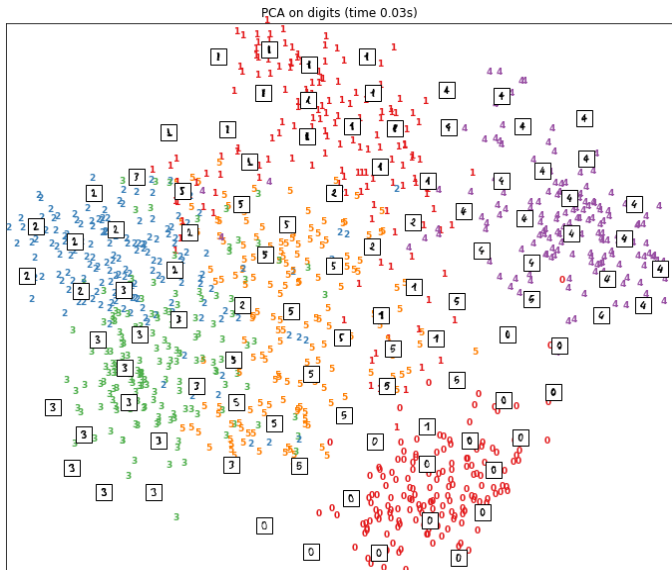
t-SNE vs PCA

t-SNE



9 segundos

PCA



0.03 segundos

t-SNE



Resumen

- **Fortalezas:**
 - De lo mejor para visualizar datos
 - Conserva estructuras no lineales globales y locales
- **Debilidades:**
 - Es estocástico (es no determinista)
 - Escala mucho en tiempo con dimensiones y puntos
 - No se puede usar para nuevos puntos