QUEEN'S UNIVERSITY BELFAST
SCHOOL OF ELECTRONICS, ELECTRICAL ENGNIEERING AND COMPUTER SCIENCE

Data Analysis & Visualisation

CSC3062

BEng (CS & SE), MEng (CS & SE), BIT & CIT

Dr Reza Rafiee

Semester 1 2019

Visualising high dimensional data using t-SNE

Reference: http://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf

t-SNE

Visualisation is key to understand data easily

t-Distributed Stochastic Neighbour Embedding (t-SNE) is a **non-linear technique** for dimensionality reduction

Applications

- Image processing
- Natural language processing (NLP)
- Bioinformatics and Genomic data
- Speech processing
- …

Idea behind t-SNE

- Calculating **the probability of similarity of points (i.e., data points)** in **high-dimensional** space and calculating the probability of similarity of points in the corresponding **low-dimensional** space.

- The **similarity of points** is calculated as the conditional probability that a **point A would choose point B as its neighbour** if neighbours were picked in proportion to their probability density under a **Gaussian (normal distribution)** centered at A.

- To measure the minimisation of the sum of difference of conditional probability, t-SNE minimises the sum of **Kullback-Leibler divergence** of overall points

Kullback-Leiber Divergence (KL) compares two distributions.

KL is a measure of how one probability distribution is different from a second, reference probability distribution.

QUEEN'S UNIVERSITY BELFAST | SCHOOL OF ELECTRONICS, ELECTRICAL ENGNIEERING AND COMPUTER SCIENCE

Pair-wise similarity in SNE

$$p_{j|i} = \frac{\exp(-\parallel x_i - x_j \parallel^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\parallel x_i - x_k \parallel^2 / 2\sigma_i^2)}$$

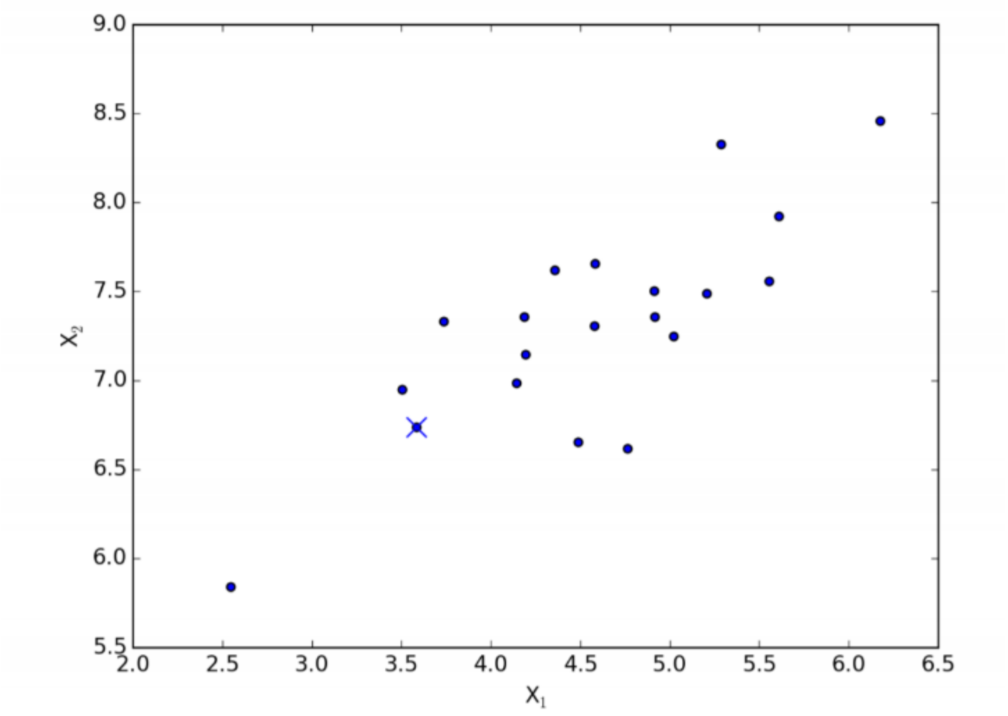
$$q_{j|i} = \frac{\exp(-\parallel y_i - y_j \parallel^2)}{\sum_{k \neq i} \exp(-\parallel y_i - y_k \parallel^2)}$$

$$p_{i|i} = 0, q_{i|i} = 0$$

SNE converts Euclidean distances to similarities, that can be interpreted as probabilities.

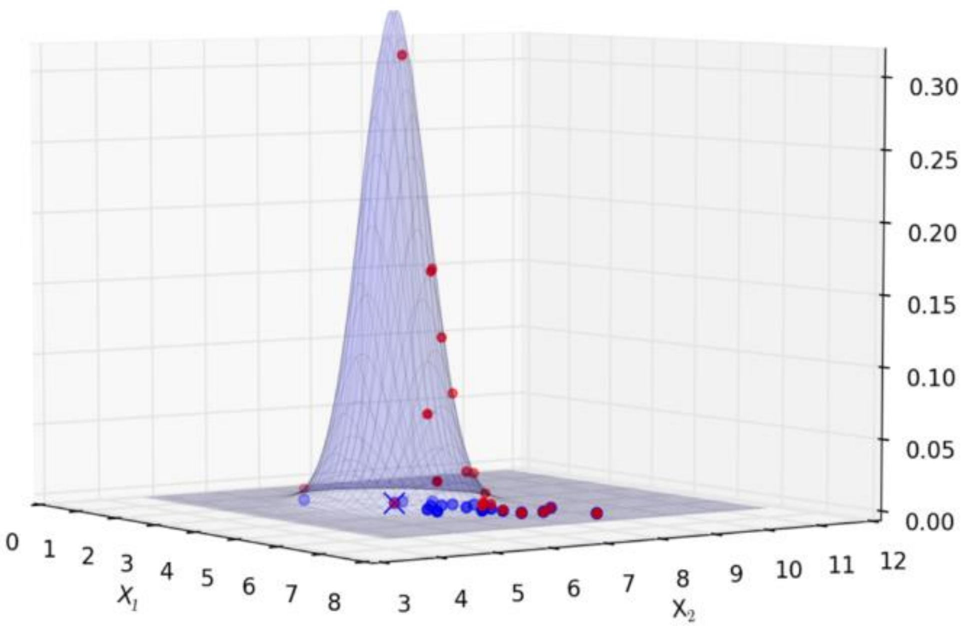
Pair-wise similarity in SNE

Data points in high-dimensional space



$p_{j|i}$

Similarity in high dimension



This is why t-SNE can be interpreted as topology-based dimensionality reduction technique

SNE converts Euclidean distances to similarities, that can be interpreted as probabilities.

SNE, symmetric SNE & t-SNE

SNE ⇒

Modelisation:

$$p_{j|i} = \frac{\exp(-\|x_i - x_j\|^2/2\sigma_i^2)}{\sum_{k\neq i}\exp(-\|x_i - x_k\|^2/2\sigma_i^2)}$$

$$q_{j|i} = \frac{\exp(-\|y_i - y_j\|^2)}{\sum_{k\neq i}\exp(-\|y_i - y_k\|^2)}$$

Cost Function:

$$C = \sum_i KL(P_i||Q_i)$$

Derivatives:

$$\frac{dC}{dy_i} = 2\sum_j (p_{j|i} - q_{j|i} + p_{i|j} - q_{i|j})(y_i - y_j)$$

Symmetric SNE ⇒

Modelisation:

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n}$$

$$q_{ij} = \frac{\exp(-\|y_i - y_j\|^2)}{\sum_{k\neq l}\exp(-\|y_k - y_l\|^2)}$$

Cost Function:

$$C = KL(P||Q)$$

Derivatives:

$$\frac{dC}{dy_i} = 4\sum_j (p_{ij} - q_{ij})(y_i - y_j)$$

- **Faster Computa-tion**

t-SNE

Modelisation:

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n}$$

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k\neq l}(1 + \|y_k - y_l\|^2)^{-1}}$$

Cost Function:

$$C = KL(P||Q)$$

Derivatives:

$$\frac{dC}{dy_i} = 4\sum_j (p_{ij} - q_{ij})(y_i - y_j)(1 + \| y_i - y_j \|^2)^{-1}$$

- **Even Faster Computation**
- **Better Behaviour**

Perplexity (cost function parameter)

The perplexity is defined as

$$Perp(P_i) = 2^{H(P_i)},$$

where $H(P_i)$ is the Shannon entropy of P_i measured in bits

$$H(P_i) = -\sum_j p_{j|i} \log_2 p_{j|i}.$$

The perplexity can be interpreted as a smooth measure of the effective number of neighbors. The performance of SNE is fairly robust to changes in the perplexity, and typical values are between 5 and 50.

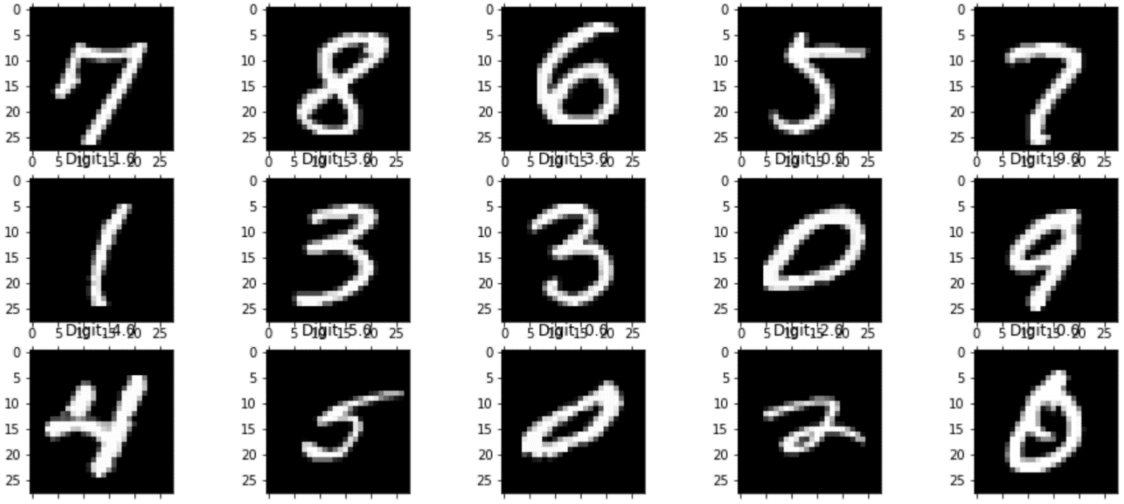
Algorithm

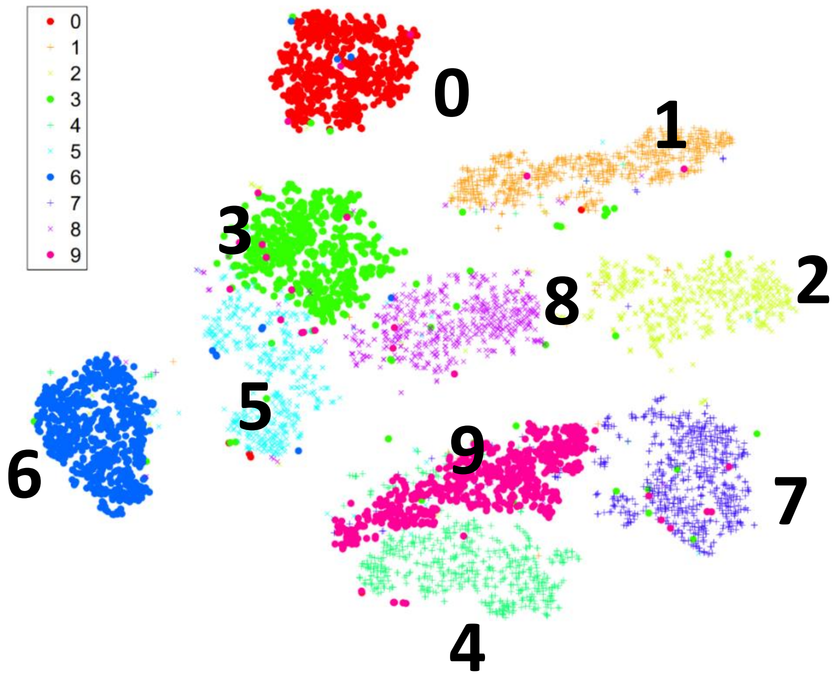
Algorithm 1: Simple version of t-Distributed Stochastic Neighbor Embedding.

Data: data set $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$,
cost function parameters: perplexity $Perp$,
optimization parameters: number of iterations T, learning rate η, momentum $\alpha(t)$.
Result: low-dimensional data representation $\mathcal{Y}^{(T)} = \{y_1, y_2, \ldots, y_n\}$.
begin
compute pairwise affinities $p_{j|i}$ with perplexity $Perp$ (using Equation 1)
set $p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n}$
sample initial solution $\mathcal{Y}^{(0)} = \{y_1, y_2, \ldots, y_n\}$ from $\mathcal{N}(0, 10^{-4} I)$
for t=*1* **to** T **do**
compute low-dimensional affinities q_{ij} (using Equation 4)
compute gradient $\frac{\delta C}{\delta \mathcal{Y}}$ (using Equation 5)
set $\mathcal{Y}^{(t)} = \mathcal{Y}^{(t-1)} + \eta \frac{\delta C}{\delta \mathcal{Y}} + \alpha(t) \left(\mathcal{Y}^{(t-1)} - \mathcal{Y}^{(t-2)} \right)$
end
end

Applying t-SNE on datasets

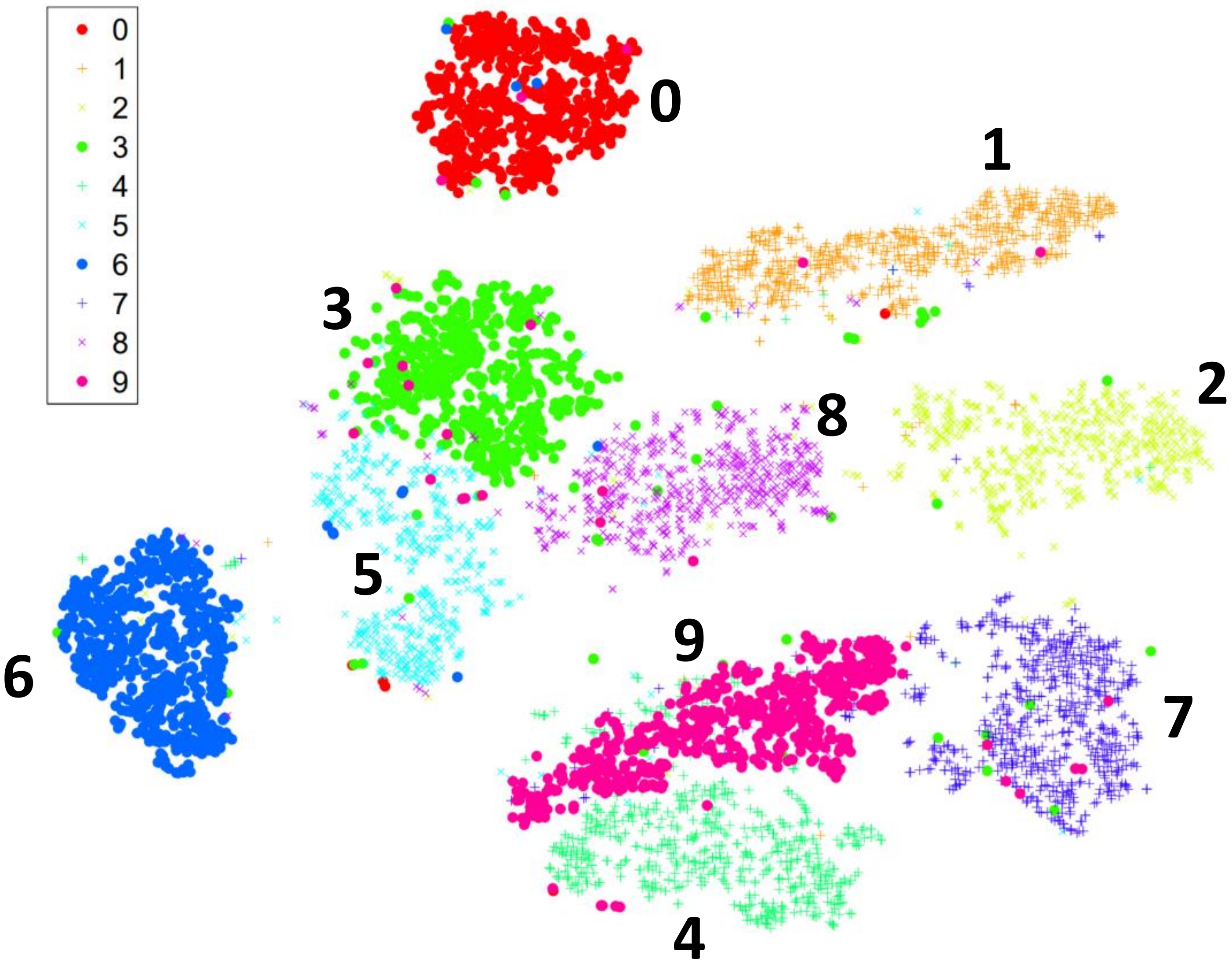
The MNIST datasets contains **60,000 grayscale images of handwritten digits**. They randomly selected 6,000 of the images for computational reasons. The digit images have 28×28 = 784 pixels (i.e., dimensions).





http://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf

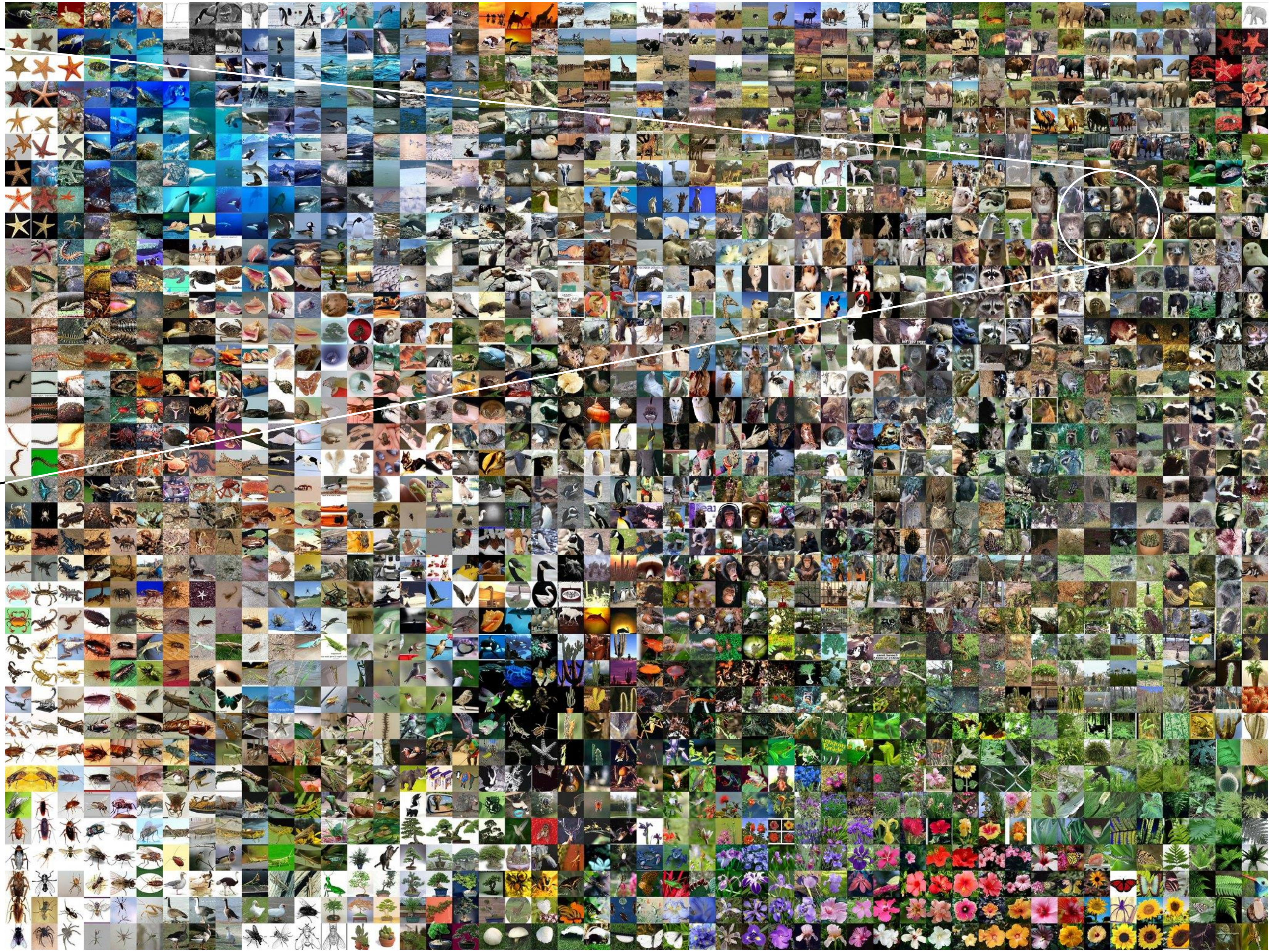
t-SNE result



QUEEN'S UNIVERSITY BELFAST

SCHOOL OF ELECTRONICS, ELECTRICAL ENGNIEERING AND COMPUTER SCIENCE

Applying t-SNE on colour images



2D visualisation of a large image dataset of various objects using t-SNE dimensionality reduction technique

https://github.com/genekogan

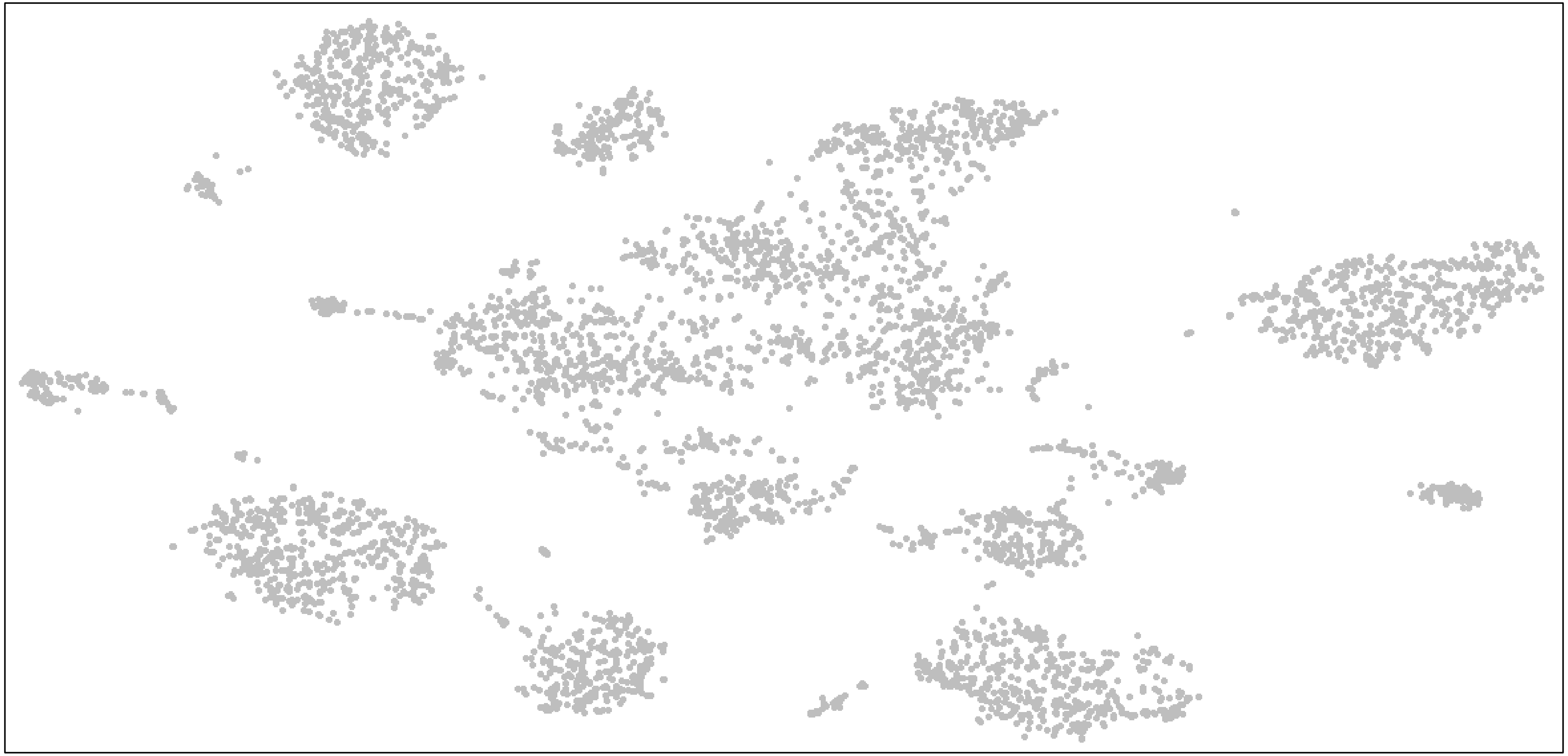
Data summarisation & visualisation

QUEEN'S UNIVERSITY BELFAST

SCHOOL OF ELECTRONICS, ELECTRICAL ENGNIEERING AND COMPUTER SCIENCE

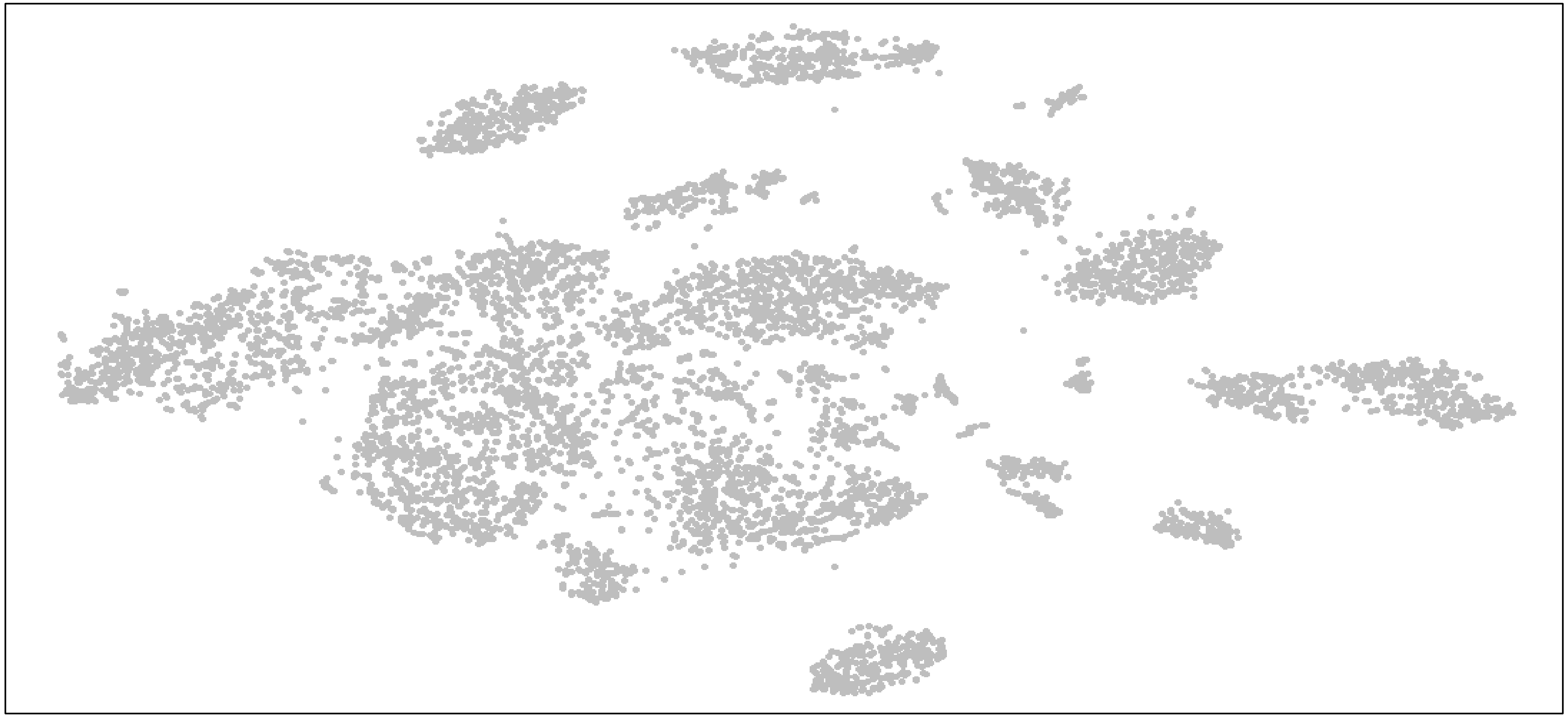
Unsupervised clustering of 9126 solid tumours (16,335 genes) – transcriptomic map of tumour

2D visualisation of a large dataset of 30 solid tumours using *t-SNE* dimensionality reduction technique



Data summarisation & visualisation

Unsupervised clustering of 7302 solid tumours (440 genes)



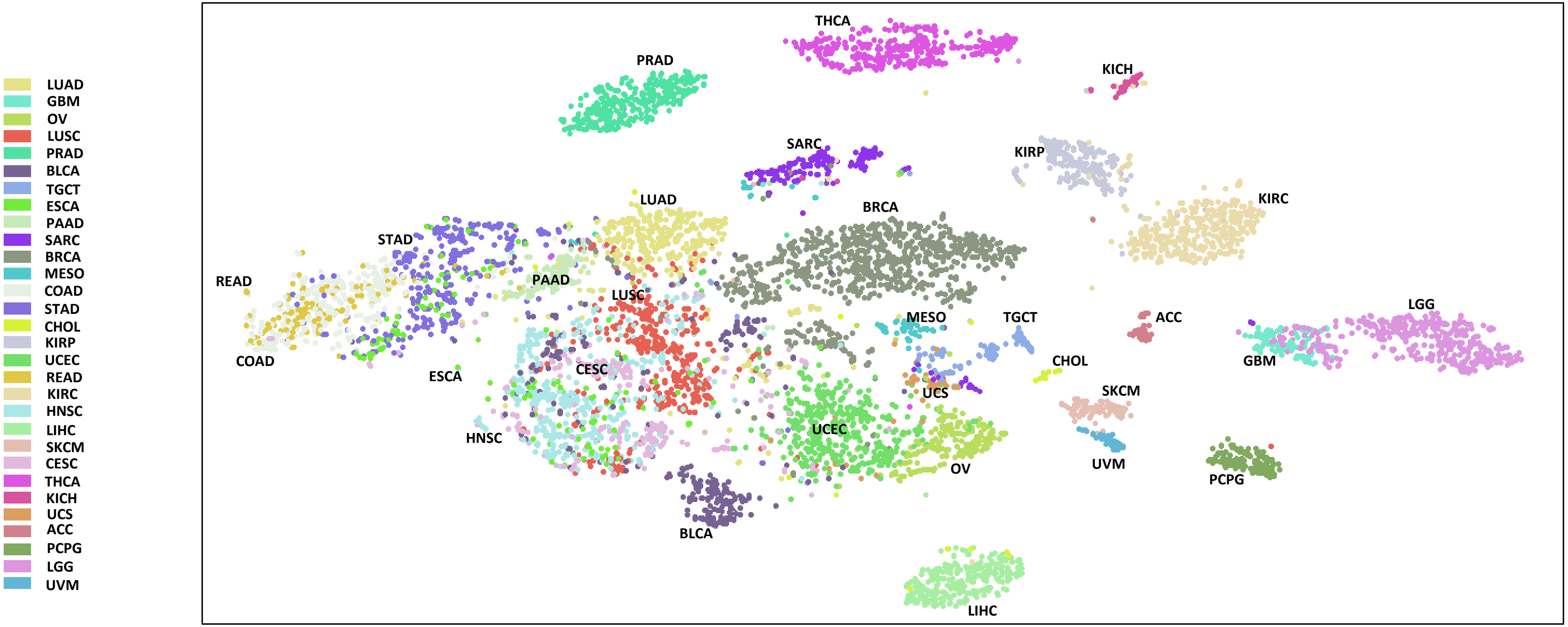
2D visualisation of a large dataset of 30 solid tumours using *t-SNE* dimensionality reduction technique

Visualisation using t-SNE

Unsupervised clustering of 7302 solid tumours (440 genes)

a

b



Establishing the gene expression-based immune solid tumours reference cohort. a, Overview of the 30 non-hematologic/solid tumour cohorts. **b**, Unsupervised clustering of reference cohort samples (n=7,302) using t-SNE dimensionality reduction technique. Individual samples are colour-coded in the respective class colour (n=30) and labelled with the class abbreviation.

tsne() function in R

Usage

```
tsne(X, initial_config = NULL, k = 2, initial_dims = 30, perplexity = 30,
     max_iter = 1000, min_cost = 0, epoch_callback = NULL, whiten = TRUE,
     epoch=100)
```

Arguments

`X`	The R matrix or "dist" object
`initial_config`	an argument providing a matrix specifying the initial embedding for X. See Details.
`k`	the dimension of the resulting embedding.
`initial_dims`	The number of dimensions to use in reduction method.
`perplexity`	Perplexity parameter. (optimal number of neighbors)
`max_iter`	Maximum number of iterations to perform.

More details about t-SNE

https://www.youtube.com/watch?v=g72uroShwmI

http://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf

Any Questions?