Parameter Tuning in t-SNE

Content

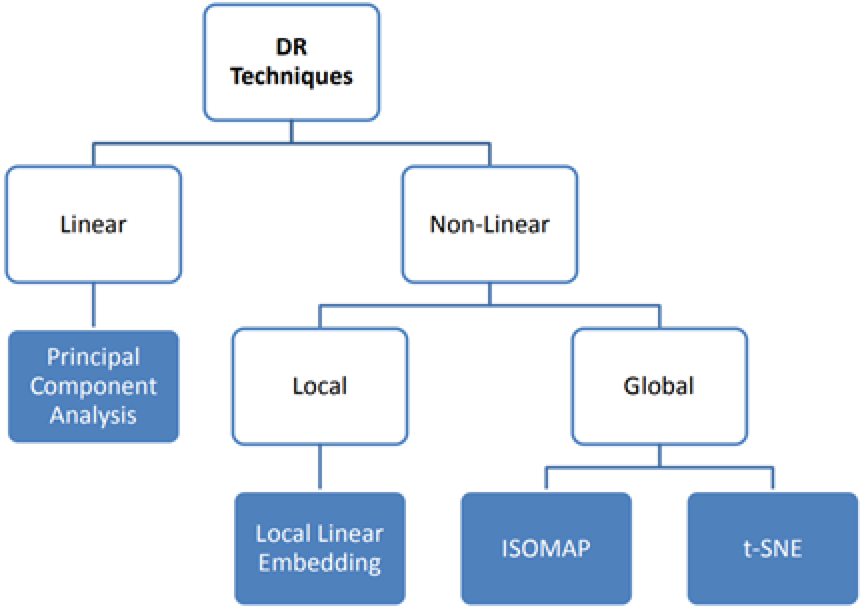
1. **What is t-SNE** (a small recap)
   a. Motivation
   b. Theory
2. **Parameter Tuning**
   a. Theory
3. **Pros and Cons of t-SNE**
4. **Code Example**
5. **Summary**

What is t-SNE

T-distributed Stochastic Neighbor Embedding

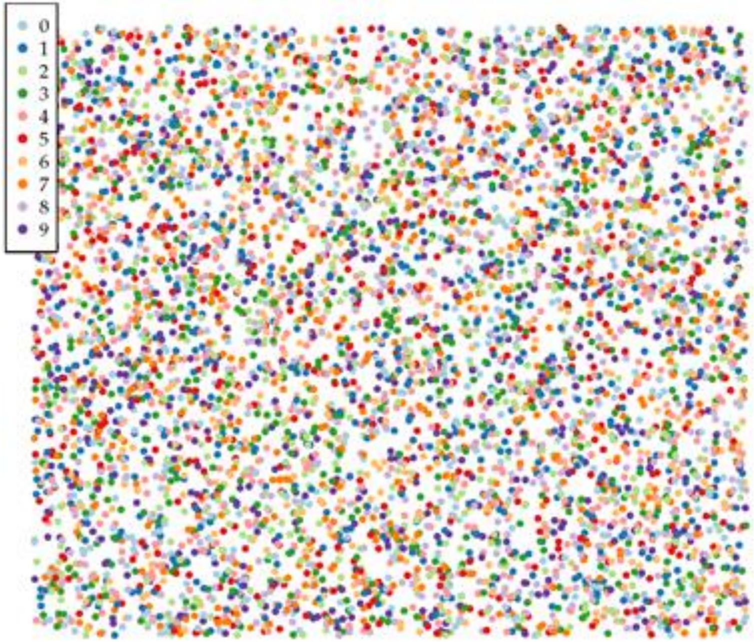
Motivation

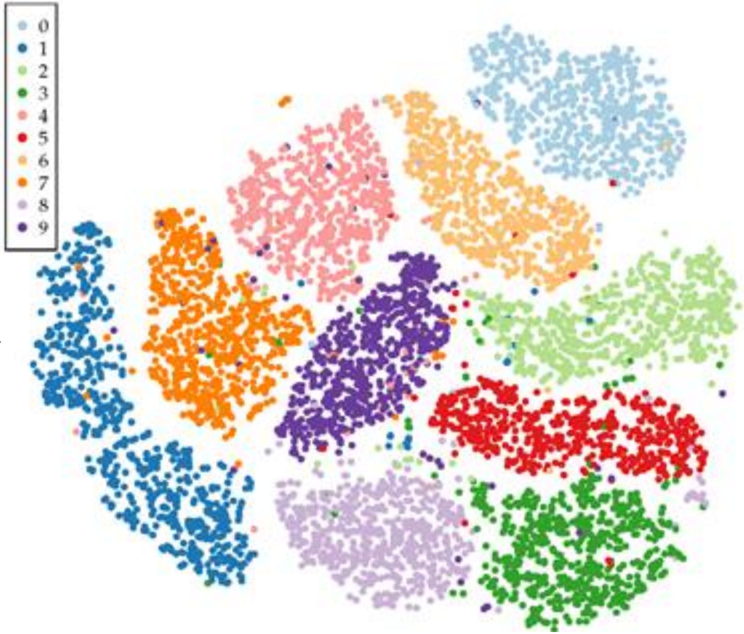
- **The problem to be solved: High-dimensional data**
  - Can mean high computational cost to perform learning
  - Often leads to over-fitting
  - **Hard to interpret**
- **The solution?**
  - ➢ Can use *Dimension Reduction*, to **visualize** high-dimensional data in a low-dimensional space



Medium. "Importance of Dimensionality Reduction!!"
Fabian Kleiser. "High Dimensional Data Visualization"

Example - MNIST handwritten numbers





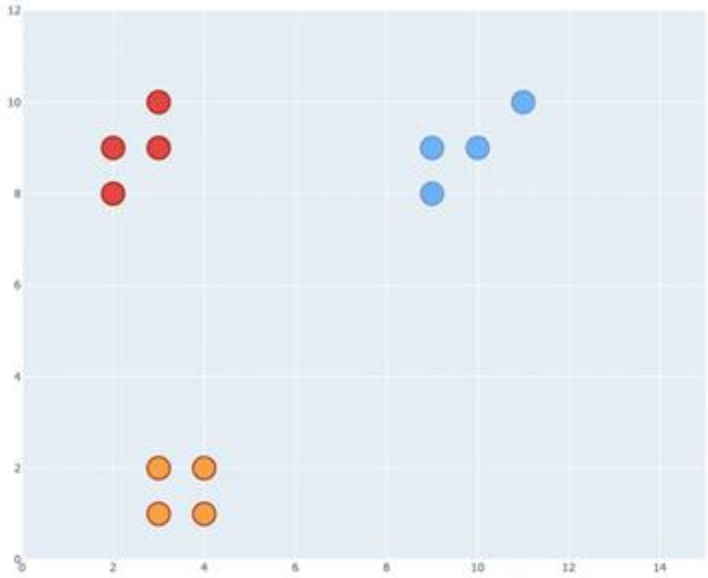
t-SNE Explained

- Example from 2D to 1D
- *(But in practice usually from highD to 2D or 3D)*

Towards data science, “t-SNE clearly explained”

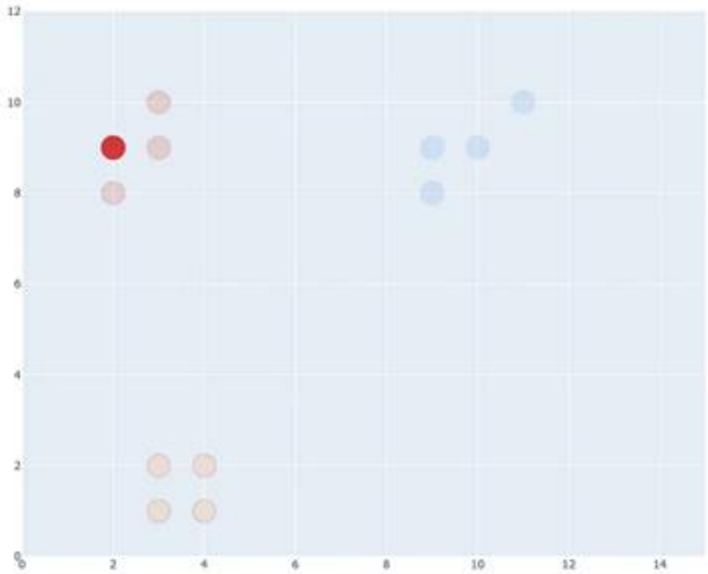
Youtube Video Explaining in more detail: Josh Starmer, “StatQuest: t-SNE, Clearly Explained”

The Data

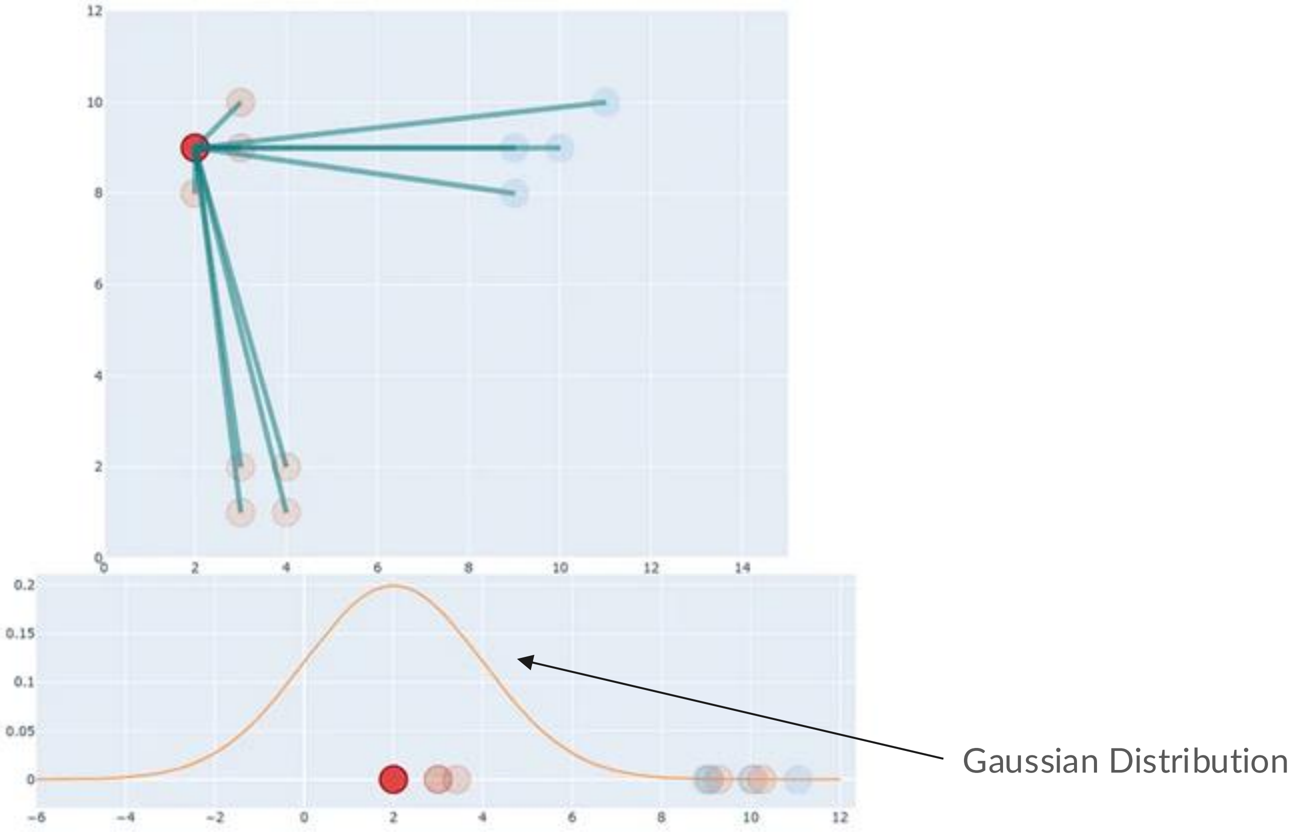


Towards data science, "t-SNE clearly explained"

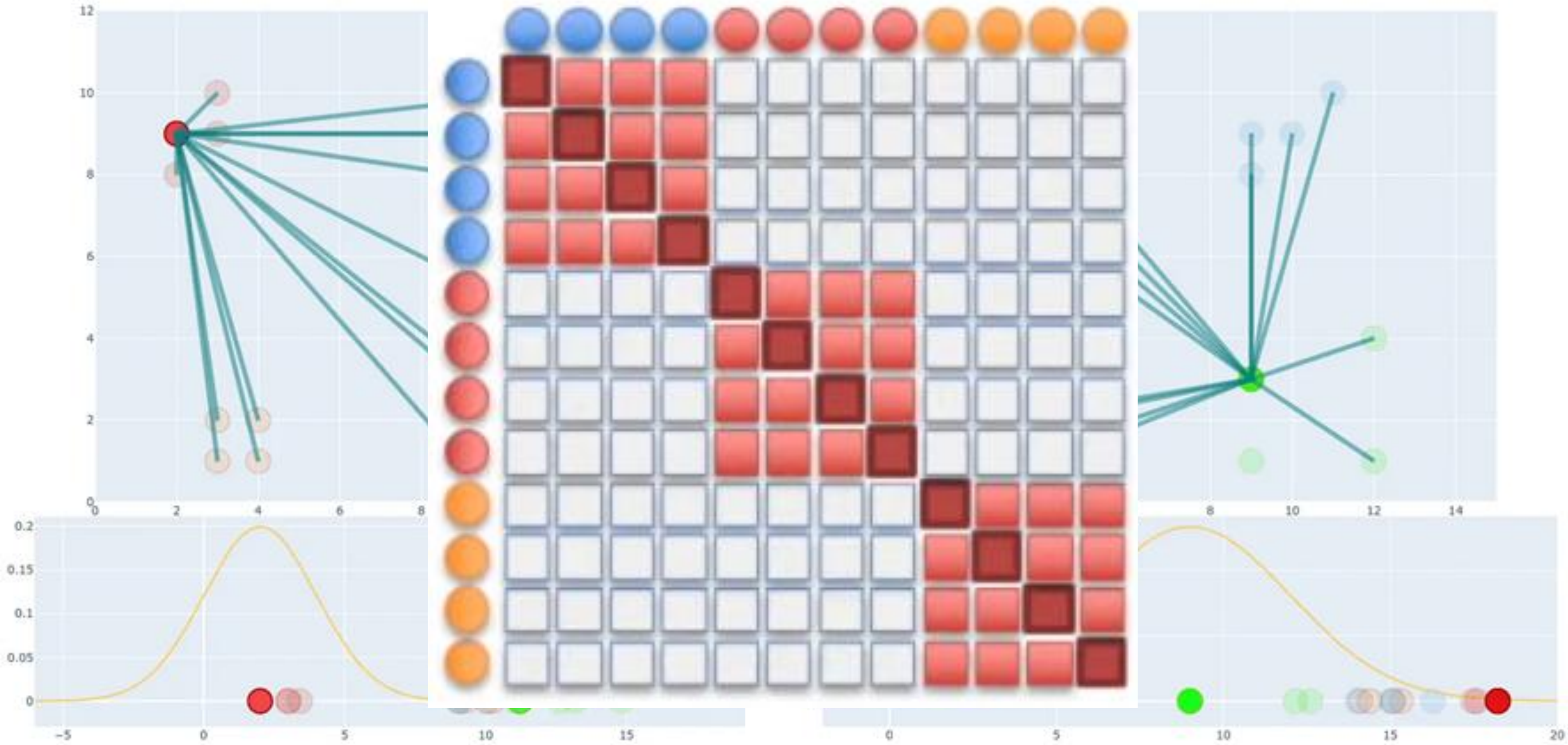
Compute Pairwise Similarities in High-Dimensional Space



Towards data science, "t-SNE clearly explained"



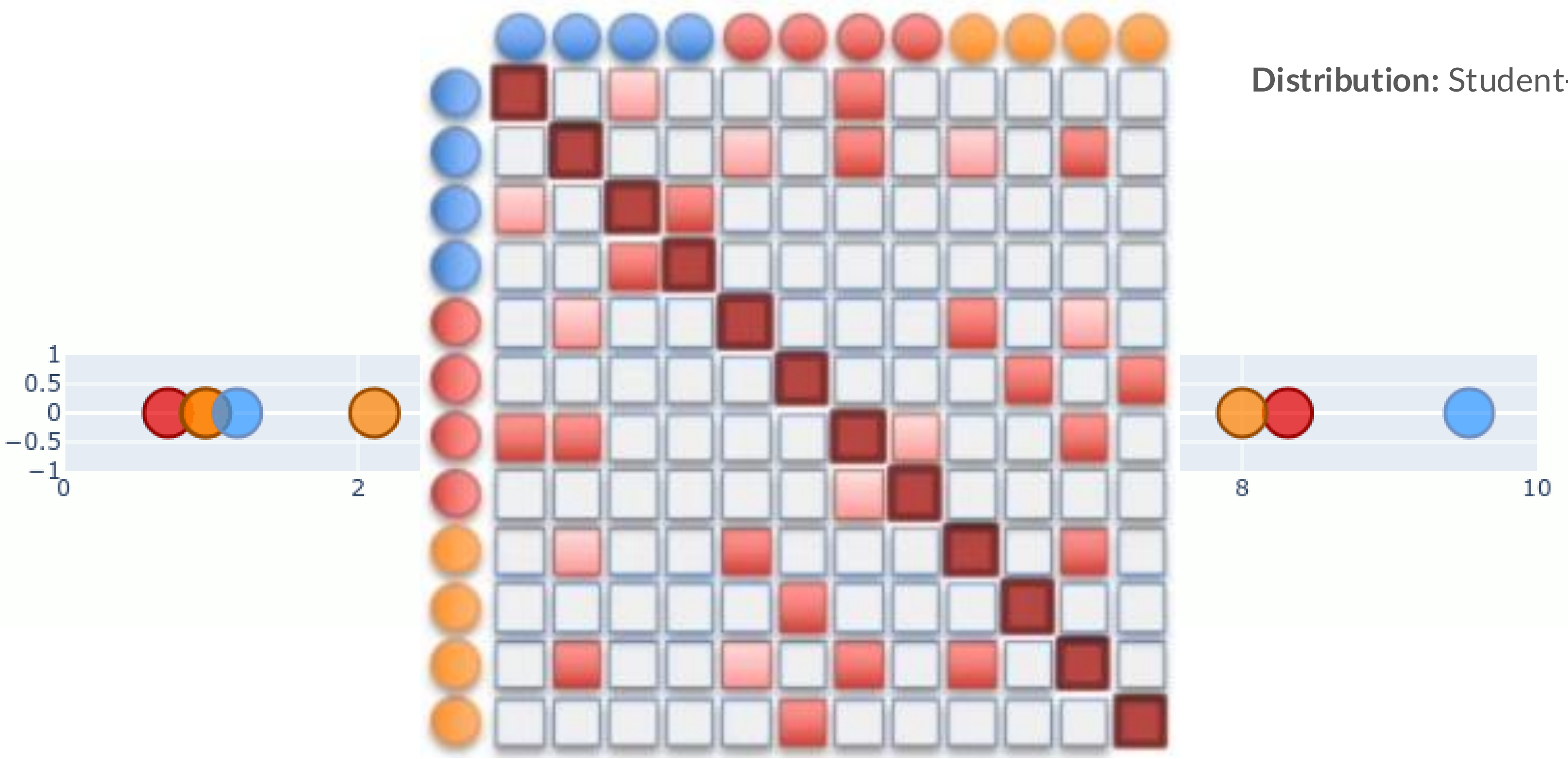
Towards data science, "t-SNE clearly explained"



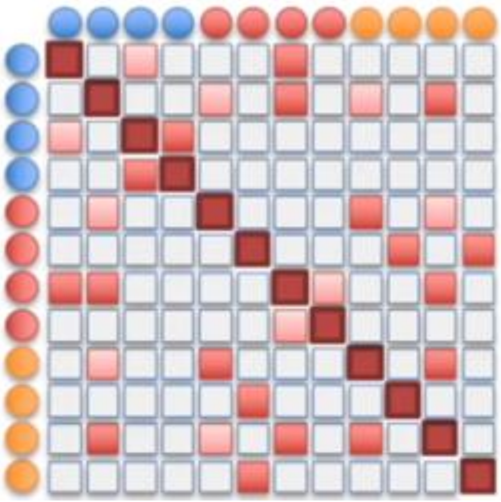
[Towards data science, “t-SNE clearly explained”]

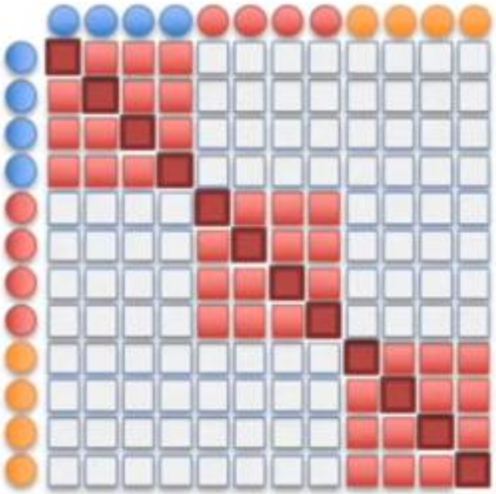
Compute Pairwise Similarities in Low-Dimensional Space

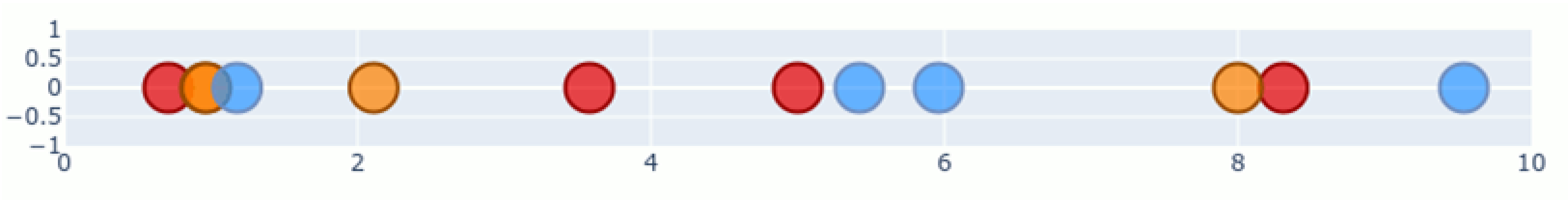
Distribution: Student-t



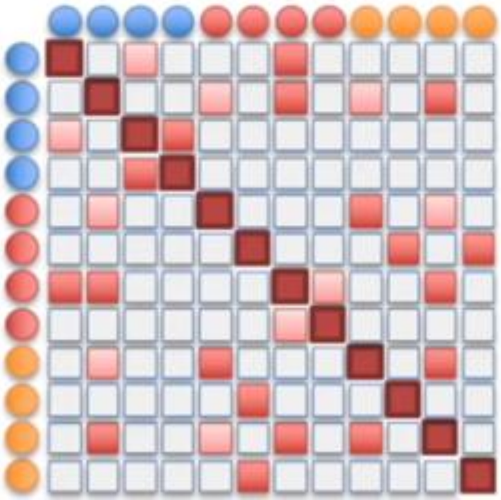
Minimize the KL-Divergence (Cost Function)

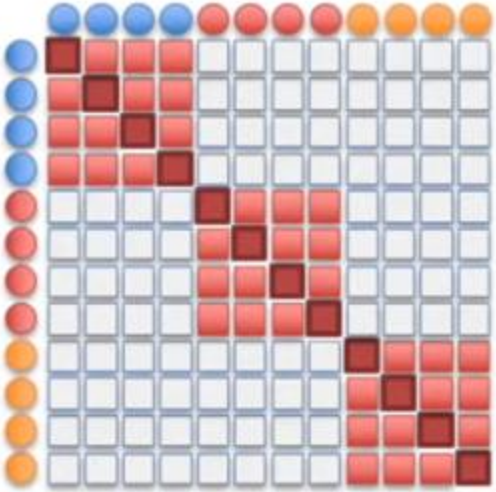


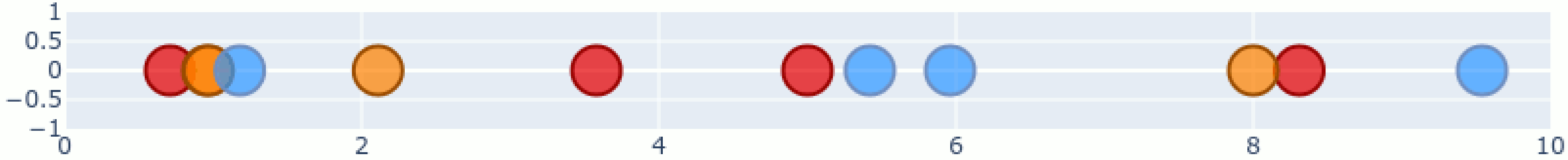




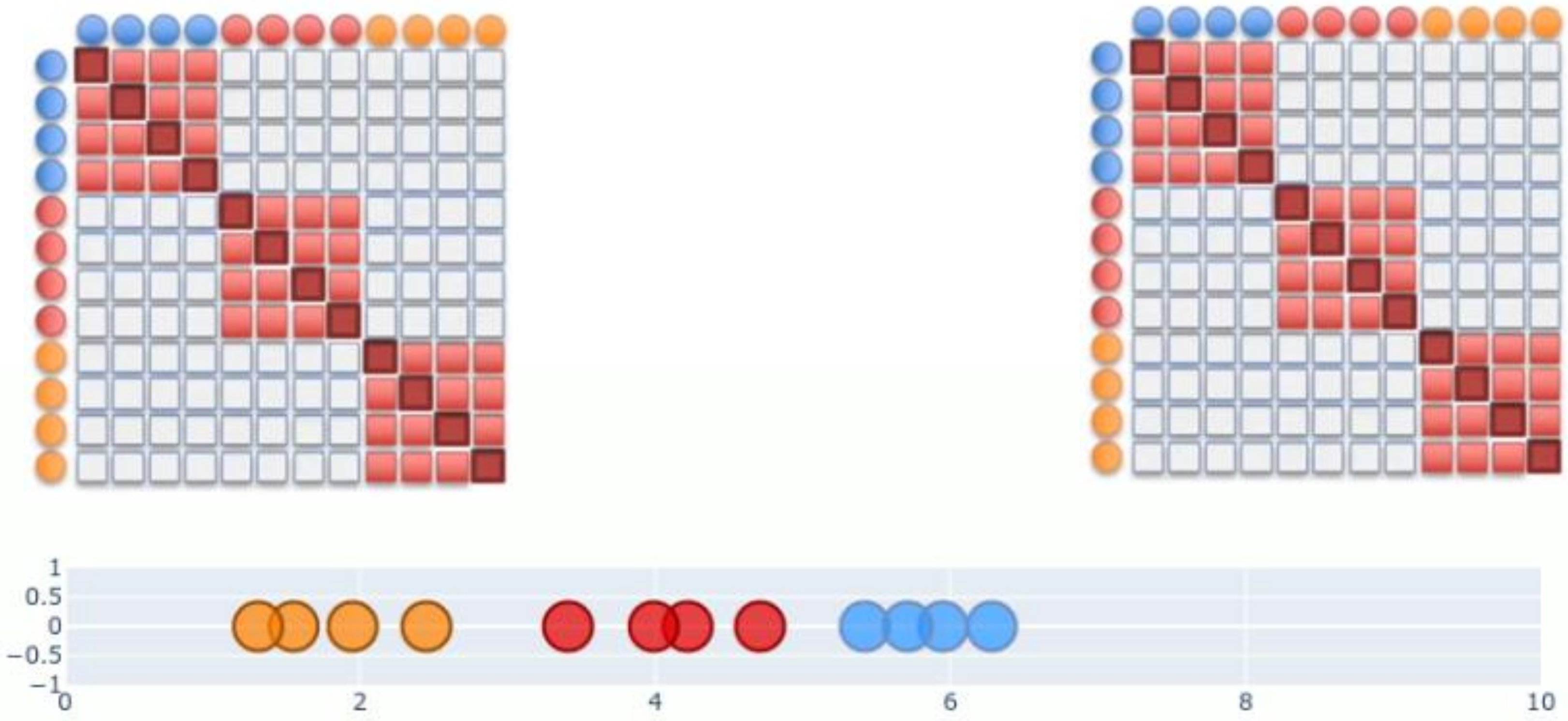
Optimization through Gradient Descent







Towards data science, "t-SNE clearly explained"



Towards data science, t-SNE clearly explained

The Parameters

- **Cost function parameter:**
    - Perplexity: *How many nearest neighbors are considered when calculating similarities*
- **Optimization parameters:**
    - Number of iterations: *How long optimization process*
    - Learning rate: *How large steps when updating points' positions during gradient descent*
    - Momentum: *Used to accelerate the convergence*
- And the dimensions you want to reduce to, of course

Parameter Tuning

parameter tuning (theory)

- Minimize the Kullback-Leibler divergences
- Run the algorithm with the same data, changing the perplexity
- Recommended perplexity between 5 and 50
- Number of iterations

Pros & Cons of t-SNE

Pros & Cons

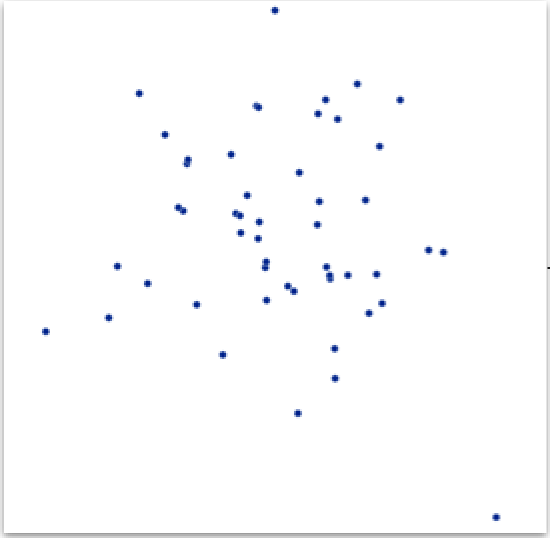
Pros

- Creates very good two-dimensional “maps” of high-dimensional data
- Keeps local structures of data while revealing global structure as well
- Outperforms most of the visualization techniques (UMAP is comparable, faster tho)
- Do not suffer from “crowding problem”

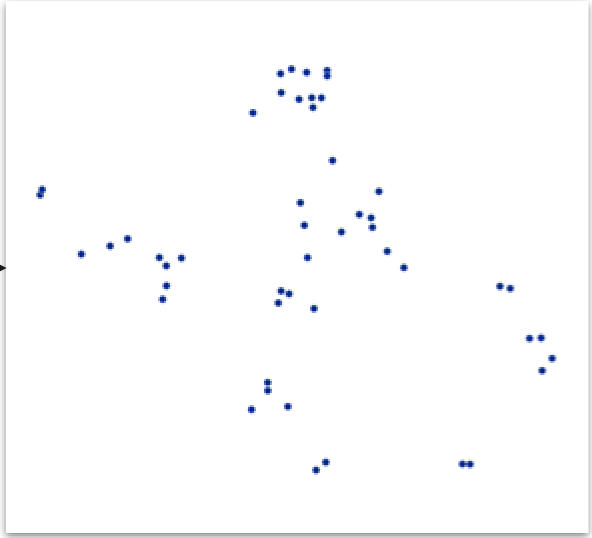
Cons

- Behavior not clear for dimensionality reduction (i.e. where low-dimensional space $d > 3$)
- Non-convexity of cost function
- Does not perform well on data with high intrinsic dimensionality (i.e. number of relevant coordinates needed to describe the data-generating process accurately)
- Fine tuning can be challenging & expensive
- Can be misleading

Example for misleading plots



Original unit Gaussian distribution



After t-SNE
(Hallucination of clusters)

Coding example

Example: MNIST datasett

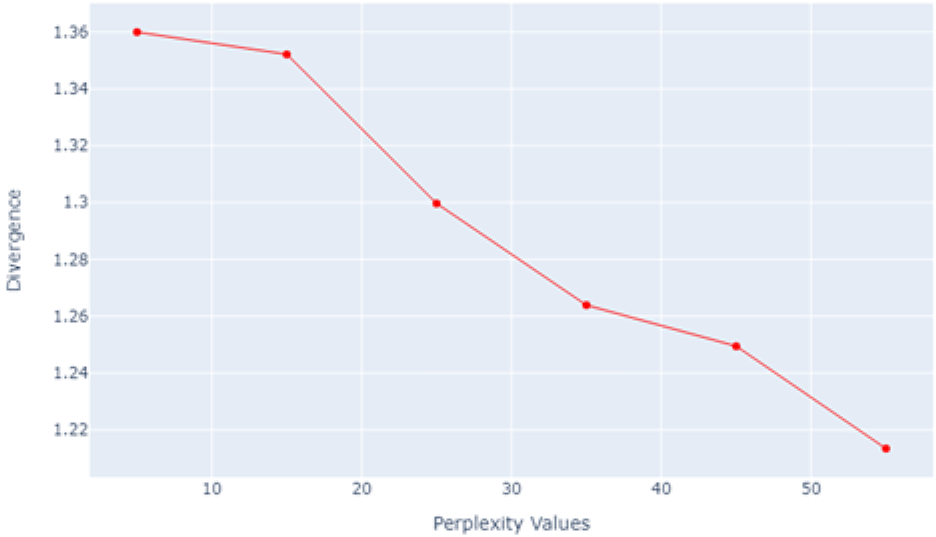
t-SNE from scikit-learn

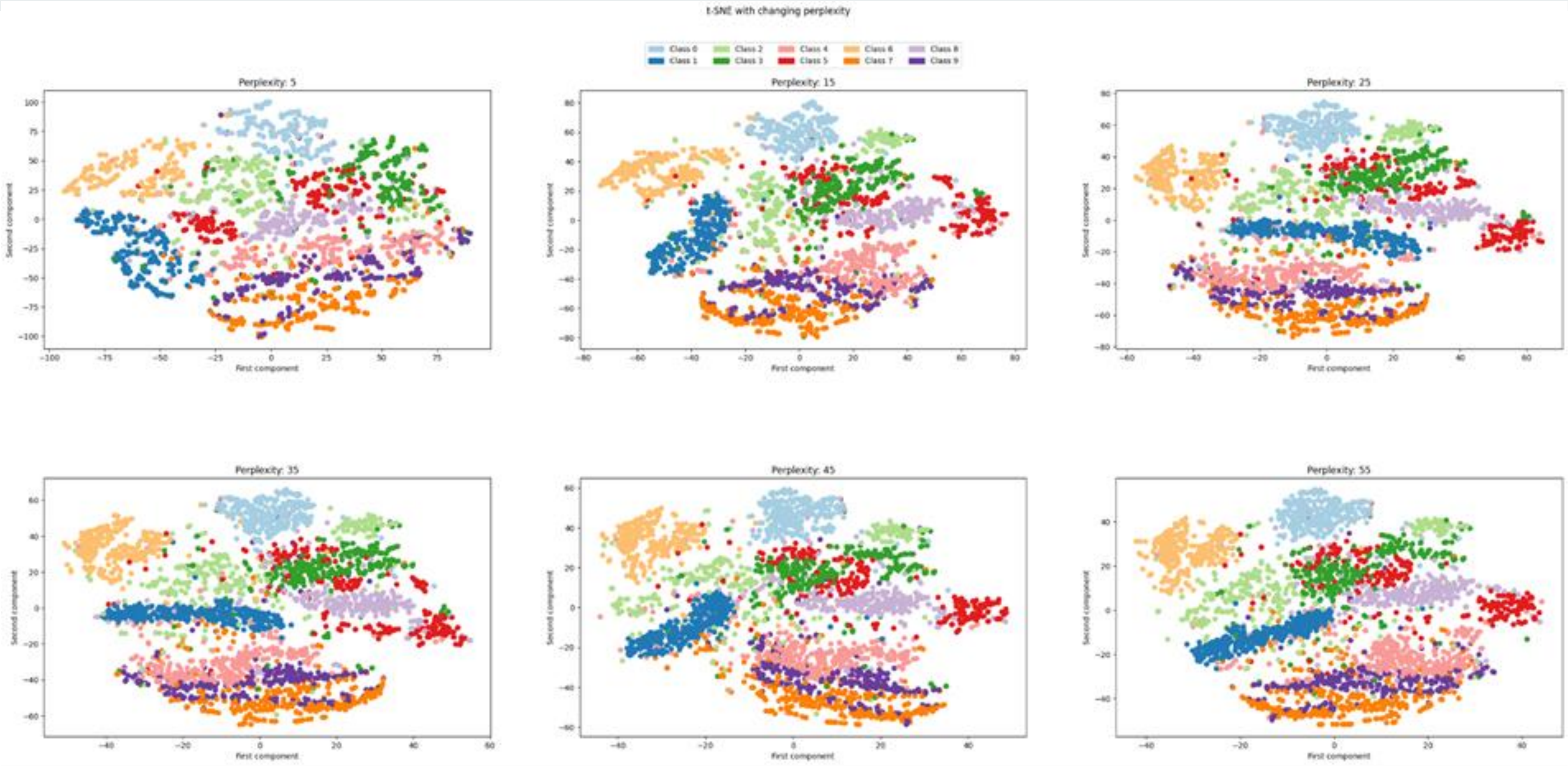
Number of iterations: 1000

Learning rate: ‘auto’

Number of components: 2

Perplexity:
[5, 15, 25, 35, 45, 55]




t-SNE with changing perplexity
Class 0
Class 1
Class 2
Class 3
Class 4
Class 5
Class 6
Class 7
Class 8
Class 9
Perplexity: 5
Perplexity: 15
Perplexity: 25
Perplexity: 35
Perplexity: 45
Perplexity: 55
Second component
First component

Summary

Summary

- t-SNE
  - Method to visualize high-dimensional data in a low-dimensional space
  - CAN be used for dimension reduction as well
- Tunable Parameters
  - Cost function parameter
  - Optimization parameters
- Notes
  - Just test a lot :,)
  - Non-deterministic
  - Don’t use the same parameters on similar datasets

Sources

- https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf
- https://scikit-learn.org/stable/modules/manifold.html#t-sne
- https://lvdmaaten.github.io/publications/papers/JMLR_2008.pdf
- Parameter Tuning - interactive example
  - https://distill.pub/2016/misread-tsne/
  - https://github.com/karpathy/tsnejs
- Parameter Tuning
  - https://www.datacamp.com/tutorial/introduction-t-sne?dc_referrer=https%3A%2F%2Fwww.google.com%2F