t-SNE

t-SNE

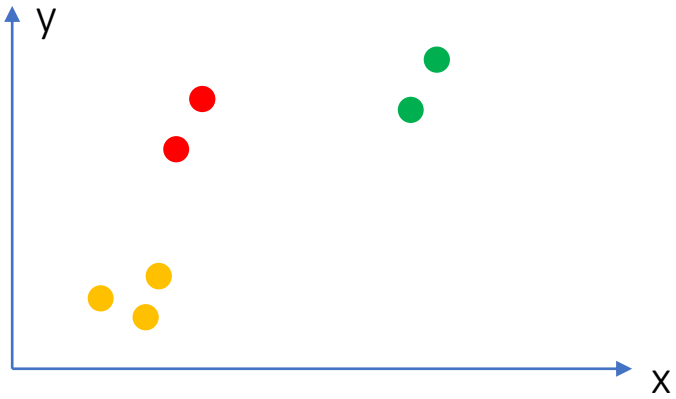
Introduction

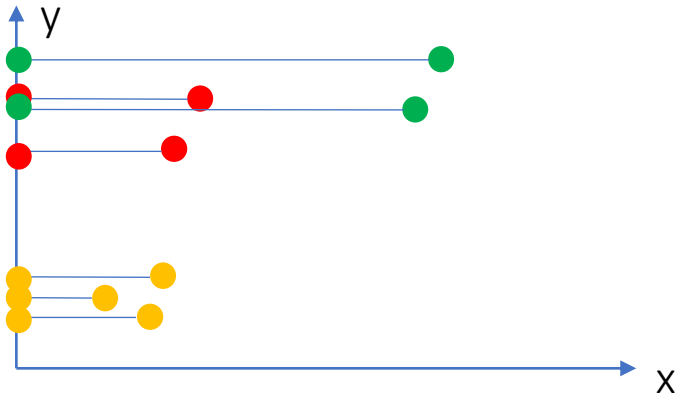
- t-Distributed Stochastic Neighbor Embedding
- dimensionality reduction technique
- uses non-linear projection
- Overcomes crowding problem
- Takes local and global structure into account
- Keeps the distance of points in low dimension constant in high dimension
- t-SNE preserves local similarities of observations

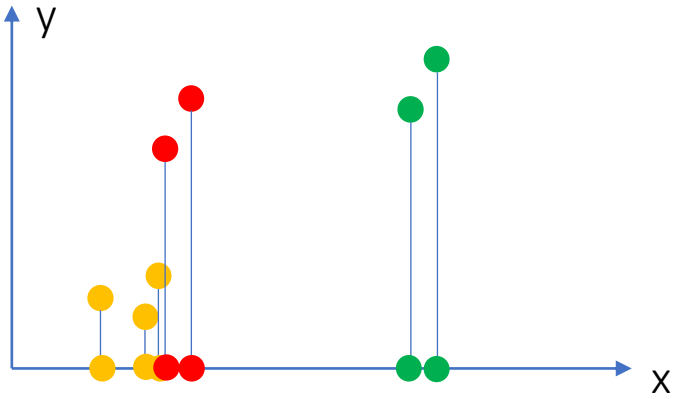
t-SNE

Crowding Problem

- Comes from curse of dimensionality
- Mapping from 2D to 1D creates crowed regions



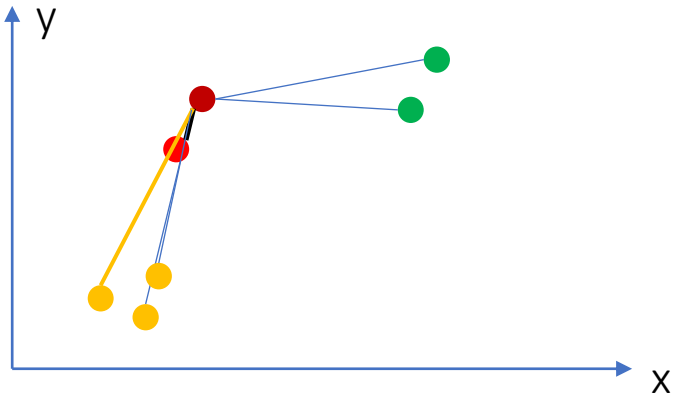




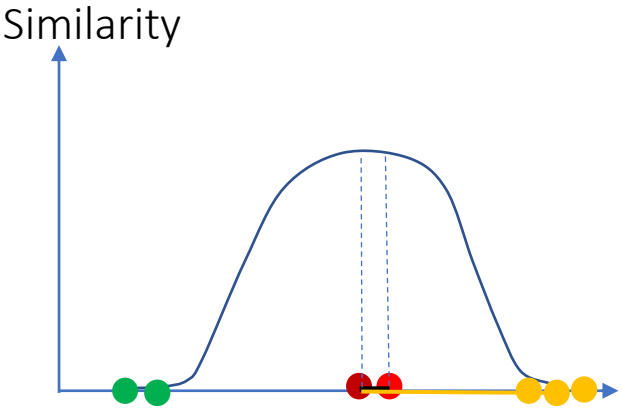
t-SNE

How it works

1. Distance measure of all points to all other points



2. Plot all distances of points on a Gaussian Curve

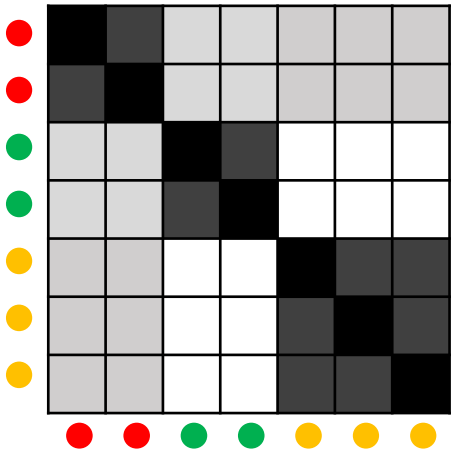


t-SNE

How it works

3. Derive similarity matrix (high dimensional space)

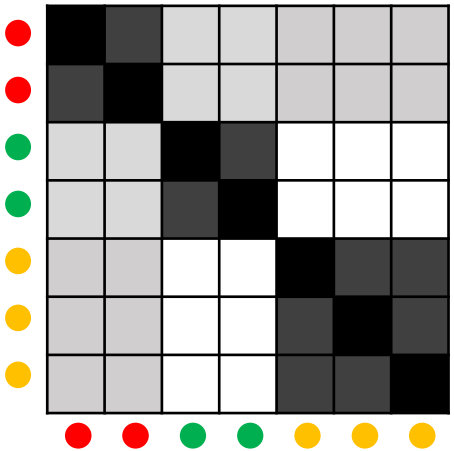
Similarity Matrix



Similarity
low
medium
high
perfect

4. Reproduce similarity matrix in low dimensional space by rearranging points (Student-t distribution)

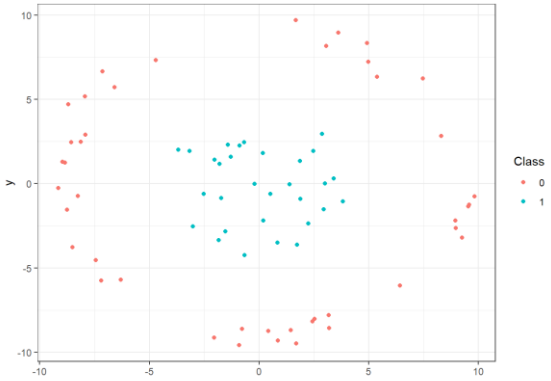
Similarity Matrix

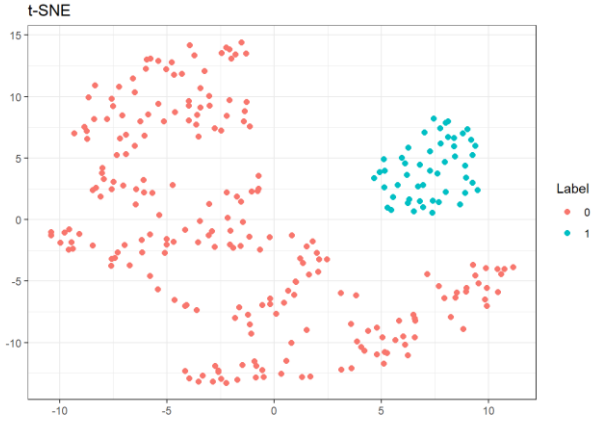


t-SNE

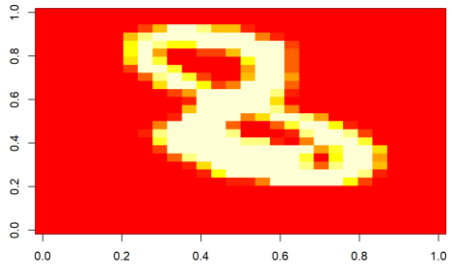
Examples

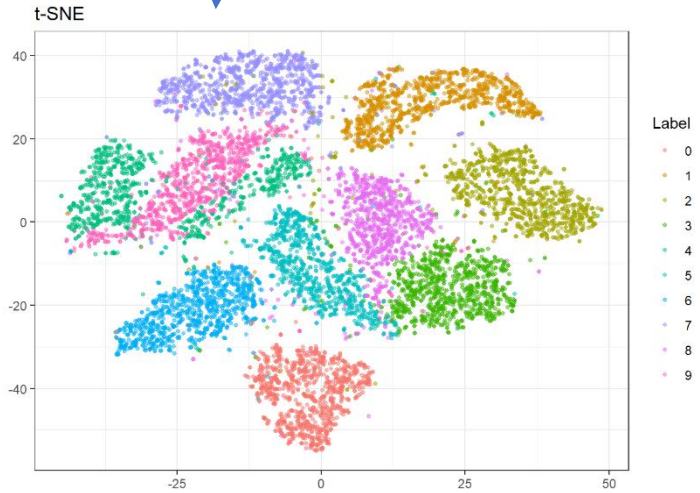
Spheres





Mnist





t-SNE

Advantages / Disadvantages

\+

- Can cover more complex structures than PCA

\-

- Non-convex loss function leads to non-deterministic solution
- Uses Euclidean distance, which assumes linearity
- Much higher computational effort than PCA