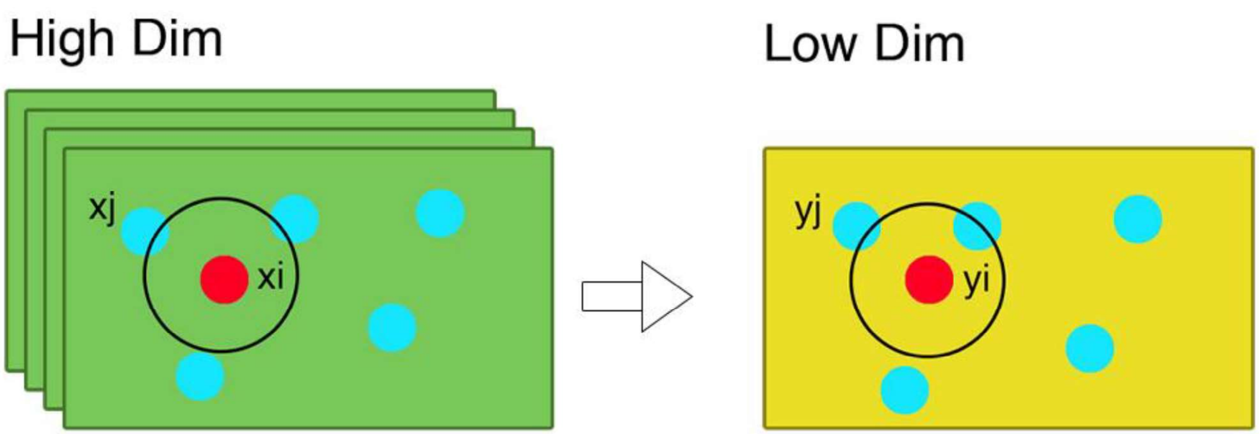
t-SNE

- Stands for **t-**distributed **S**tochastic **N**eighbour **E**mbedding
- A non-linear dimensionality reduction technique which is mainly used for visualization purposes of high dimensional data.
- Maps high-dimensional data to 2 or 3 dimensions, making it easier for us to visualize.
- Tries to preserve local structure in the data
    - This is done by preserving the distribution of the data found in higher dimension to the lower dimension as much as possible.
    - This means that points that are closer to each other in the higher dimension are close in the lower dimension too, and the points far apart from each other in the higher dimension are far apart in the lower dimension too

This file is meant for personal use by unmesh_redkar@hotmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.

How does it do this?

- Compute a distribution that measures pairwise similarities in original data (high dim)
- Find a ‘close’ lower dim distribution of pairwise similarities
- Use this mapping to transform higher dimensional data into lower dimensional data.



- Finding this ‘close’, technically involves minimizing the divergence between two distributions. The minimization is usually achieved using stochastic gradient descent and Kullback-Leibler divergence is usually the measure of divergence.

This file is meant for personal use by unmesh_redkar@hotmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.

Steps involved in t-SNE

- Scale the dataset
- Compute the similarity of data points in high dimensional space
- Compute the similarity of data points in corresponding lower dimensional space
- Minimize the distance between these two distributions using KL divergence
- The lower dimensional distribution obtained when the algorithm converges is the lower dimensional representation of the original data

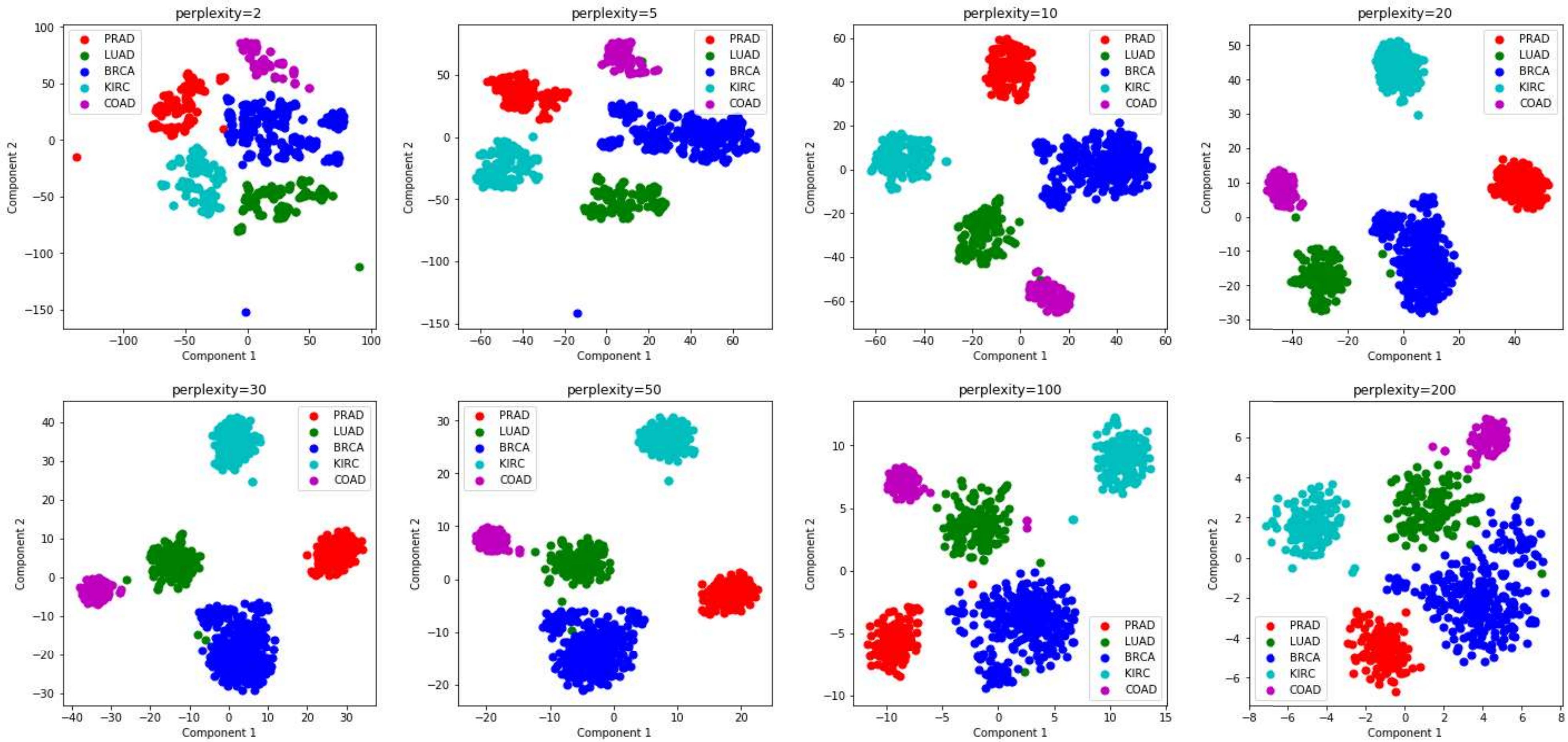
This file is meant for personal use by unmesh_redkar@hotmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.

Choosing the right Perplexity

- We find patterns by identifying observed clusters based on similarity of data points with multiple features. However, after this process, the input features are no longer identifiable, and you cannot make any inference based only on the output of t-SNE. Hence it is mainly a data exploration and visualization technique.
- The hyper parameter ‘perplexity’ in some sense captures the balancing act between local and global aspects of the data. Has a significant effect on the resulting visualization.
- Perplexity can also be thought of as the guess on number of neighbors
- The range of values recommend for perplexity is between 5 to 50 .

This file is meant for personal use by unmesh_redkar@hotmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.

Different Perplexity Value



This file is meant for personal use by unmesh_redkar@hotmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.

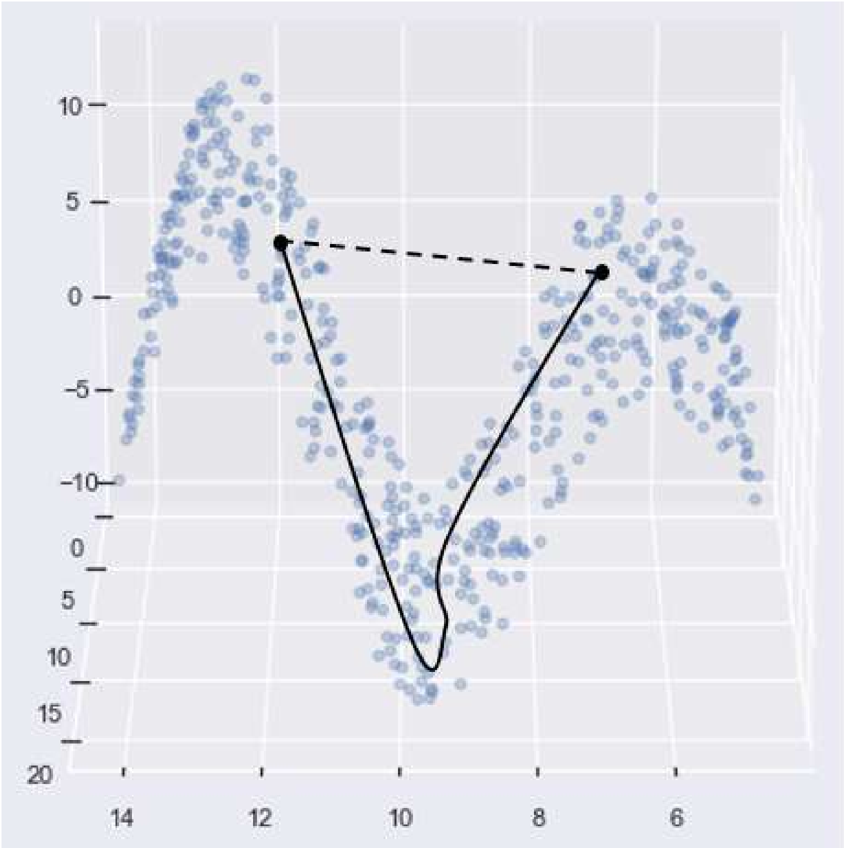
PCA vs t-SNE

PCA	t-SNE
Linear	Non-linear
Maximizes variance capture to find the structure within the data	Relies on probability distributions to find the structure within the data
Seeks to preserve all pairwise distances	Seeks to preserve small pairwise distances
Tries to preserve global structure only	Tries to preserve local structure only
Computation time is low	Computation time is high

This file is meant for personal use by unmesh_redkar@hotmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.

Why t-SNE over PCA for visualization?

- The visualization illustrates how distances between points are computed by PCA (dotted line) and t-SNE (solid line).
- The points, which are actually distant as per structure of data, are interpreted as similar (close) by PCA
- However, t-SNE captures the local structure correctly and interprets them as dissimilar (far apart)



This file is meant for personal use by unmesh_redkar@hotmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.