

Recap

T-SNE

Goal of T-SNE

Preserve probability of pair of point to be Neighbours

It is widely used Visualization

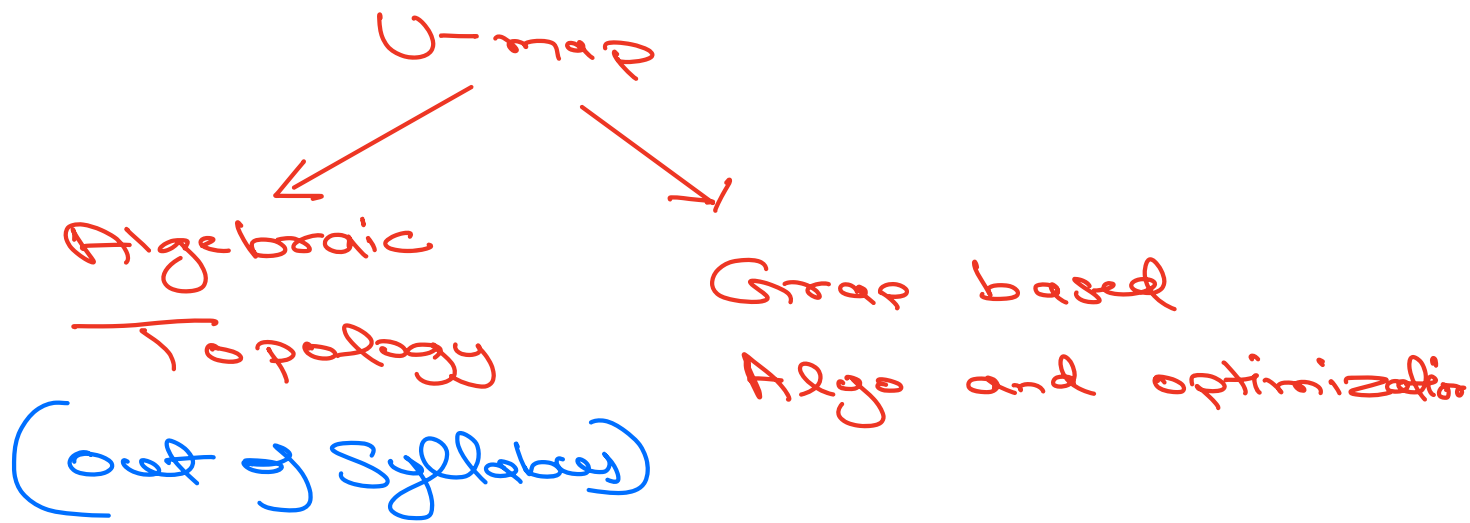
High Dimension	Low Dimension
Normal Distribution $P(i, j) \uparrow$	+ - Distribution [to avoid crowding] $Q(i, j) \uparrow$

Loss Function:

KL-Divergence \Rightarrow

$$KL(P, Q)_{div} = \sum_i \sum_j P_{ij} \log \left(\frac{P_{ij}}{Q_{ij}} \right)$$

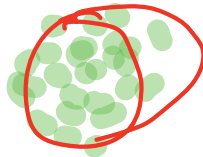




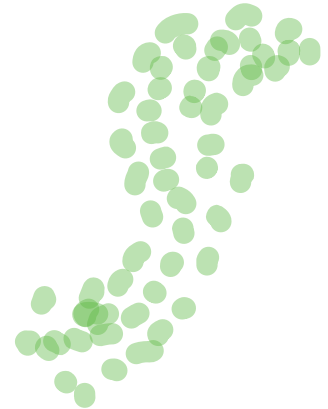
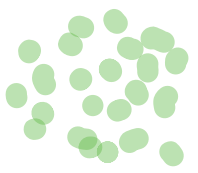
manifold \mathbb{D}



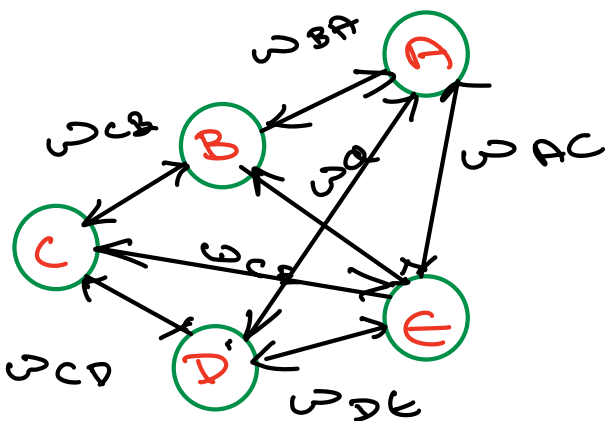
\mathbb{D}



\mathbb{D}

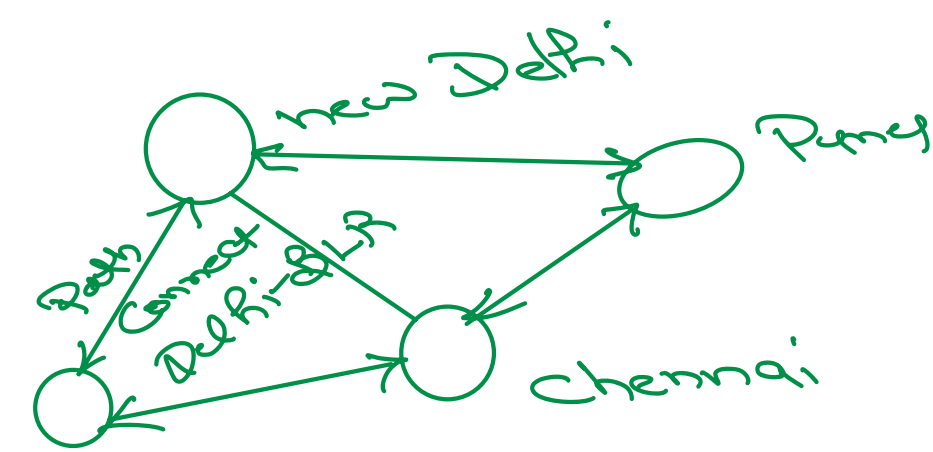


Graph



Nodes

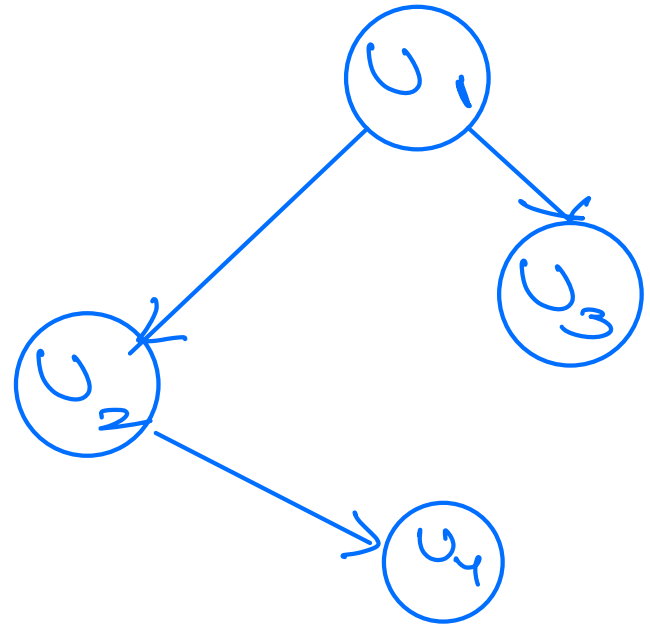
Edges can have weights



$E \Rightarrow$ Roads
 $w \Rightarrow$ Distance

BLR

Weighted Graph

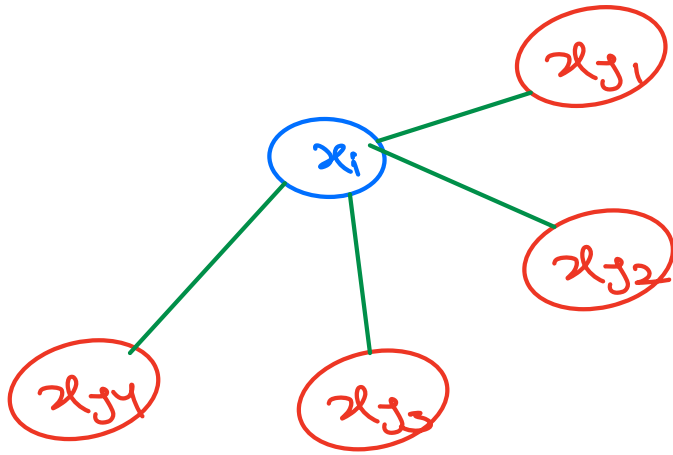


Unweighted

Algo of map

- ① Pick n -neighbours (Hyperparameter)
- ② Build Graph
 - ↳ For every point, the nodes n -neighbours

Dataset $[x_{j1}, x_{j2}, x_{j3}, x_{j4}, x_{j5}]$

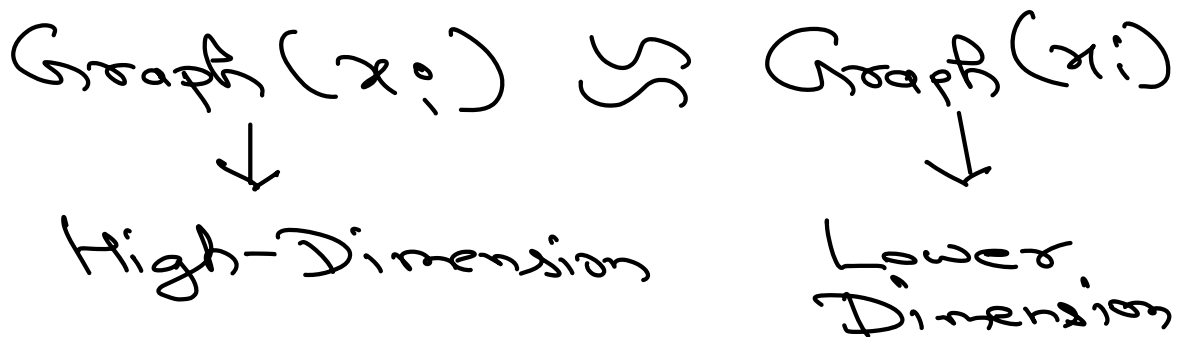


h -neighbors = 4

$$w_{ij} \Rightarrow \frac{1}{\text{Dist}(x_i, x_j)}$$

$$\frac{1}{\|x_i - x_j\|}$$

③ Project the points into Lower Dimension



How do i Compare Graphs
of two different Dimension

A new Loss
Loss \Rightarrow

$$-\sum_{i=1}^N \left[w^H_i \log \left(\frac{w^H_i}{\sum w^H_i} \right) + (1 - w^H_i) \log \left(\frac{1 - w^H_i}{1 - \sum w^H_i} \right) \right]$$

Problem - Statement

House price prediction

① Preprocessing

Scaler

BaseLine LR model

PCA and Rebuild Model

Outlier Detection

Clustering \rightarrow Build Model