Manifold Technique

1) t-SNE

t-SNE purpose is to take high-dimensional data (for ex:- images or text with many features), and reduce it to 2D or 3D, while keeping similar points close together and dissimilar points far apart.

Steps

1.) Measure Similarity in Higher Dimensions.

→ Imagine you have data points $x_1, x_2 \dots x_n$ in higher dimensions

→ Calculate how 'similar' each pair of points x_i and x_j are in this high-dimensional space.

To do this, use a Gaussian (bell curve) function.

$$P_{ij} = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma^2)}{\sum_{k \neq i} \exp(-\|x_i - x_k\|^2 / 2\sigma^2)}$$

→ This gives higher probabilities to closer points. The "closer" two points are, the higher their p_{ij} value.

→ σ controls the spread of the Gaussian, which can be adjusted to consider more or fewer neighbors.

2. Make similarity in Low dimensions.

→ Take the same data points but map them to a lower-dimensional space (like 2D), where each point y_i is the lower-dimensional version of x_i.

→ In low-dimensional space, calculate the similarity b/w each pair of points y_i and y_j.

Here instead of gaussian funcn, uses a Student's t-distribution

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

This formula also calculates how close the points y_i and y_j are, but it uses t-distribution, which helps deal with crowding problem.

Crowding problem refers to the difficulty of accurately representing the relative distances between points when projecting data from a high dimensional space unto low-dimensional Space.

3) Compare the two Spaces

→ t-SNE now compares how similar points are in high-dimensional space (the p_{ij}'s) to how similar they are in low-dimensional space

→ The goal is to minimize the difference between the high-dimensional similarities and the low-dimensional similarities.

4) Kullback-Leibler (KL) Divergence:

- To compare the two similarity distributions (high dimensional and low dimensional), t-SNE uses KL divergence, which measures how different two probability distributions are:-

$$KL(P \| Q) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

If the distributions p_{ij} and q_{ij} are very similar, this divergence will be small.

5) Optimization (Adjusting the points).

→ t-SNE then adjusts the positions of the points in the low-dimensional space y_i's, using gradient descent $\left(\frac{\partial KL}{\partial y_i}\right)$ to minimize the KL divergence.

→ This process continues until the points in low-dimensional space best preserve the local structure of the high-dimensional data.

Locally Linear Embedding (LLE)

Non linear dimensionality reduction technique

→ Locally Linear Embedding (LLE) is a manifold learning technique that reduces the dimensionality of high-dimensional data while preserving the local structure of the data.

→ LLE Assumes that each data point and its neighbours lie on a locally linear patch of the manifold.

→ It aims to map this local structure to a lower-dimensional space.

Steps in LLE

1) Finding the Nearest Neighbours

For each data point x_i, identify its K nearest neighbors in high dimensional space. This can be done using euclidean distance

$$d(P.Q) = \sqrt{\sum_{i=1}^{n} (y_i - x_i)^2}$$

$$d(P,Q) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$

$P = (x_1, y_1)$

$Q = (x_2, y_2)$

2). Compute Reconstruction weights.

For each point, calculate the weights w_{ij} such that each point can be written as a linear combination of neighbors. The weights are chosen to minimize the reconstruction error:

$$E(W) = \sum_{i} \| x_i - \sum_{j \in \text{neighbours of } i} w_{ij} x_j \|^2 .$$

Also, the sum of the weights must be equal to 1 for each point.

$$\sum_{j} w_{ij} = 1.$$

for x_1:-

we need to find w_{12} and w_{13} where

$x_1 \approx w_{12} x_2 + w_{13} x_3$

Step 3: Solve for Low-Dimensional Embedding

Once the weights are computed, we aim to find the low-dimensional representation that preserves these relationships. We solve an eigenvalue problem:

$$MY = \lambda Y$$

where M is a matrix derived from the weights W. The smallest non-zero eigenvalues give the 2D coordinates.

Matrix M construction -

- Diagonal Entries : Each diagonal entry M_{ii} represents the amount by which the point cannot be reconstructed from its neighbors.

Ex :- $M_{11} = 1 - (w_{12} + w_{13})$

$M_{31} = - w_{31}$

$M_{32} = - w_{32}$

- Off diagonal Entries

↓

Each off-diagonal entry M_{ij} represents the Negative contribution of point j in reconstructing point i.

Locally Linear Embedding (LLE) seeks a lower-dimensional projection of the data that preserves local neighbourhoods' distances.

When to use t-SNE -

→ If your primary goal is to visualize high-dimensional data and discover clusters or patterns, t-SNE is often better choice due to its ability to represent complex relationships visually.

When to use LLE -

→ If preserving Local linear relationships is critical and the dataset exhibits more linear characteristics, LLE may be a better options.