**t-SNE**

## Aaryan

## CO21BTECH11001

t – distributed stochastic neighbor gradient (t-SNE) is a technique that visualizes high-dimensional data by giving each datapoint a location in a two or three-dimensional map.

Let’s say we want to convert the high dimensional data  
 into a 2 or 3 dimensional data .

First, we assume a random initial solution .

The aim of the algorithm is to preserve as much of the significant structure of the high-dimensional data as possible in the low-dimensional map.  
To do so, we calculate the probability that the neighbor of is by fitting a Gaussian centered at .  
where is a parameter representing the variance, which is assumed to be constant of whole dataset.

Also, we calculate the probability that the neighbor of is by fitting a Student t-distribution centered at .

**Note:**  because we are only interested in modeling pairwise similarities.

Now, we want to find such that the mismatch between and is minimized.

A well-known measure to find the distance between two distributions is he Kullback-Leibler divergence.  
So, our cost function is the sum of Kullback-Leibler divergences over all datapoints.

t-SNE minimizes C using a gradient descent method.  
Here is the algorithm –  
 Sample initial solution.  
 Maximum number of iterations = T  
 **for** :  
   
 **end**where is the learning rate and is momentum at iteration t, which is added to speed up the optimization.

**Questions –**

1. What is the major difference between SNE and t-SNE algorithms?  
   **Ans.** SNE uses a Gaussian distribution to model the low dimensional data, while t-SNE uses a Student t-distribution for the same.
2. Why is Student t-distribution used in t-SNE algorithm?  
   **Ans.** It is used to avoid the Crowding Problem faced in SNE algorithm.
3. Which of the following is/are correct?  
   (a)   
   (b)  
   (c)   
   (d)  
   **Ans.** (a), (c)
4. What is the advantage of using t-SNE over PCA?  
   **Ans.** t-SNE is a non-linear method while PCA doesn’t work well for non-linearly correlated data.
5. What is the disadvantage of using t-SNE?  
   **Ans.**a. t-SNE is a resource-intensive algorithm because it inspects every single data point and measures the distances between every pair of points. Therefore, it takes a bit long to run this algorithm.  
   b. t-SNE is not guaranteed to converge to a global optimum of its cost function.