

t-SNE

t-Distributed Stochastic Neighbor
Embedding (t-SNE)

Icke-linjär dimensionreduceringsteknik
för data exploration och visualisera
högdimensionell data

Ger en känsla för hur datan är fördelad

Dimensionsreducering är att presentera
högdimensionell data i 2-3 dimensioner

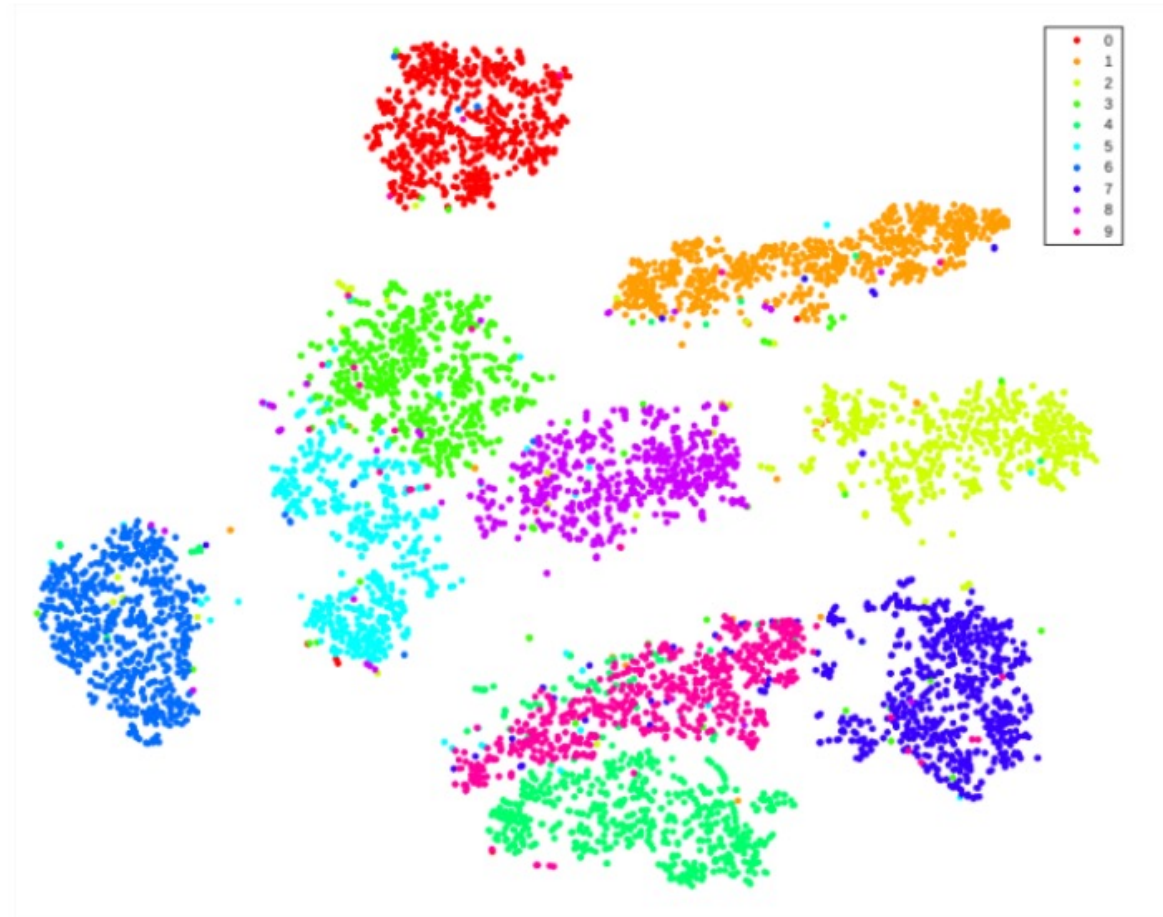


Figure 1 : Illustration of t-SNE on MNIST dataset

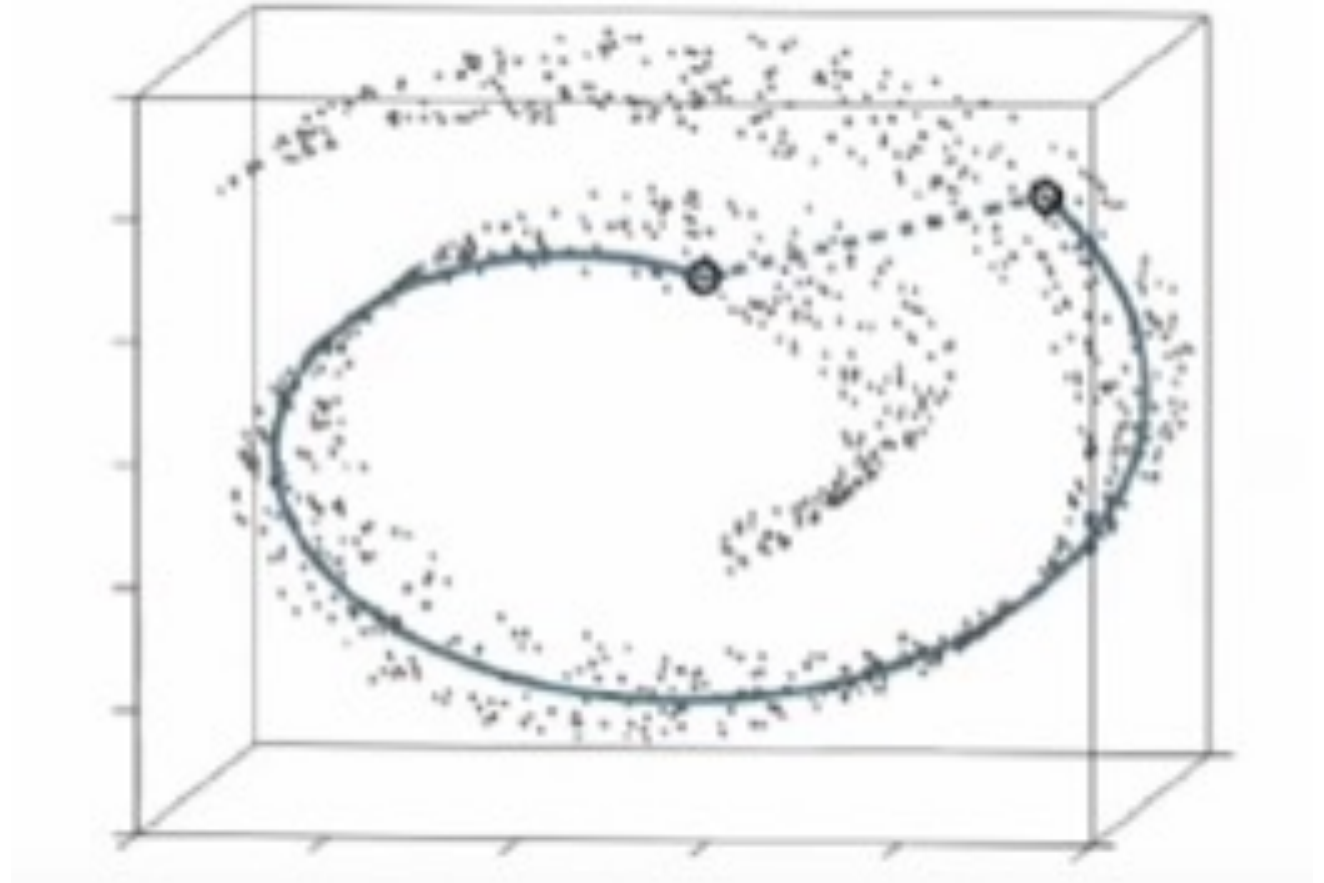
Jämförelse PCA

PCA är linjär
dimensionsreducering

PCA maximerar varians och
behåller stora parvisa distanser.

Dvs saker som är olika hamnar
långt isär

Detta kan vara problematiskt i de
icke-linjära fallen



t-SNE

- t-SNE hittar mönster i datan baserad på likheter i datapunkternas features
- Likheten beräknas med conditional probability (betingad sannolikhet) att punkt A kommer att välja punkt B som sin granna
- t-SNE försöker minimera skillnaden mellan dessa conditional probabilities (likheter) både i högdimensionella fallet och i lågdimensionella fallet
- På så sätt kan den presentera datapunkterna så bra som möjligt i låga dimensioner
- Algoritmen involverar många beräkningar så den tar mycket plats och tar lång tid att köra
- Algoritmen har kvadratisk tidskomplexitet efter antal datapunkter