Self-Organizing Map (SOM)

- create a 1D/2D lattice of artificial neurons
- define no of neurons in each dimension
- assign random weights in same dim as input
- select random data point from input
- choose winning neuron based on similarity with input
- update weights of winning and neighboring neurons (based on learning rate & neighborhood function)
- repeat for all data points until convergence

Principle Component Analysis (PCA)

Algorithm
- obtain distance matrix
- construct matric matrix
- compute eigen values and eigen vectors of matric matrix
- compute cartesian coordinates

PCA focuses on maximizing variance, and compromises resolution of proximal clusters.

A distance matrix doesn't reveal the underlying dimensionality of the space in which these points exist.

t-distributed Stochastic Neighbor Embedding (tSNE)

Algorithm
- **Similarity score** for all points against all points are computed.
- The points are then randomly placed on 2- or 3-dimensional space.
- Using an **optimization** method, points are moved step by step based on the similarity score until convergence is achieved

Retains resolution for close clusters, while scaling the farther clusters to fit in frame.

Used in combination with PCA.

K-Means Clustering

- choose no of clusters (k)
- randomly initialise k cluster centroids
- calculate distance between each data point and each centroid (euclidean or manhattan distance)
- assign data point to cluster whose centroid is closest
- recalculate centroid by taking mean of all data points assigned to the cluster
- iterate until stopping criteria met:

1. max no of iterations reached
2. centroids no longer change significantly

By **Nimisha** (lemonbuzz)

cheatography.com/lemonbuzz/

Not published yet.
Last updated 11th October, 2023.
Page 1 of 1.