Multi-Dimensional Scaling

Used for similarity/dissimilarity data (in general) and models this as distances.
Comes in two flavors.
Metric - tries to maintain actual distances and nonmetric which tries to maintain the rank of data rather than distances - x is closer to y than z.

t-SNE

t-distributed Stochastic Neighbor Embedding.
Probably the most popular.
Treats affinities of data points as probabilities.
Focuses on local clusters which is nice when the data is comprised of multiple manifolds.
KL divergence of joint probabilities in original and embedded space minimized by gradient descent - not convex so different starts/restarts get different answers.
Parameters = perplexity - number of local neighbors to consider in algorithm (more neighbors = more linear bias), learning rate, iterations, angle (again whether to emphasize local or global structure).

Check out the remaining flashcards on our LinkedIn page.

ShotNOTES

K-Means

Separates data into K groups with equal variance.
Minimizes the within cluster sum of squares (distance from mean of cluster).
Assumes spherical clusters.
Relies on distance measures, so curse of dimensionality can be a problem.
Uses a simplified EM algorithm.
Nonconvex, so multiple starts needed.

Affinity Propagation

Send messages between pairs of items asking how similar pairs of items are. Ends when a key number of exemplar items are chosen. These are the cluster centers. Chooses number of clusters itself.

Preference parameter controls how many clusters are created.
O(N**2*T) where T is number of iterations. Memory complexity is O(N**2).

Hierarchical Clustering