Cs506 lec5

Agglomerative clustering algorithm
1. let each point in the data set be in its own cluster
2. Compute the distance between all pairs if clusters
3. Merge the two closest clusters
4. Repeat 3 and 4 until all points are in the same cluster

Single link distance : the minimum of all pairwise distances between a pint from one cluster and a point from the other cluster
pro: ez to handle clusters of different sizes
Con: don't do so well when there are overlaps in clusters (sensitive to noise points
Tends to create elongated clusters

Complete ;om distance:
Maximum of all pariwise distances between a point from one cluster and a point from the other cluster

Pro: better handling noise points
Con:tends to split up clusters
Better handling clusters of similar sizes

Average link distance

The average of all pariwise distances between a point from one cluster and a point from the other cluster.

Con: less susceptible to noise and outliers

Centroid distance

Wards distance:
Difference between the spread/variance of points in the merged cluster and the unmerged cluster

Density based clustering
Cluster together points that are densely packed together
Count the number of points in a region num of points in the radius of abslone

DBScan algorithm