Lecture 5 Feb/09

Clustering:

Dis between points: d (p1, p2)

Dis between clusters: D (p1, p2)

Single-Link distance

Sensitive to noise between cluster, what’s it makes sense?

Complete-Link Dist:

Less susceptible to noise

Average-Link Dist:

Less susceptible to noise and outliers

Ward’s distance:

Merge the clusters of large variances (C12) VS not merge them (C1 & C2)

Example: (A, B, C, D point) : merge A&B then & D for Dist(A&B, D) =2

Hierarchical Clustering:

Use cost function to capture the difference between two cluster.

Density-Based Clustering:

How many points do we need to make sense it is a density-Base Cluster:

Define the Min\_pts firtly; e.g. Min\_pts = 3;

ε

**Core point**: ε-neighborhood contains at least min\_pts

**Border point**: in the ε-neighborhood (can be regarded as core point if its ε-neighborhood contains at least min\_pts)

Begin with a point and check the points in its ε-neighborhood until can’t find a core point in its neighborhood. Have begin another cluster check.