CPSC 340:
Machine Learning and Data Mining

Hierarchical Clustering

BONUS SLIDES

UBClustering Algorithm

- Let's define a new ensemble clustering method: UBClustering.

1. Run k-means with 't' different random initializations.

2. For each object i and j:
   - Count the number of times x_i and x_j are in the same cluster.
   - Define p(i,j) = count(x_i in same cluster as x_j)/t.

3. Put x_i and x_j in the same cluster if p(i,j) > 0.5.

- Like DBSCAN merge clusters in step 3 if i or j are already assigned.
   - You can implement this with a DBSCAN code (just changes "distance").
   - Each x_i has an x_j in its cluster with p(i,j) > 0.5.
   - Some points are not assigned to any cluster.

UBClustering Algorithm

It looks like DBSCAN, but far-away points will be assigned to a cluster if they always appear in same cluster as other points.

Bonus Slide: Divisive (Top-Down) Clustering

- Start with all objects in one cluster, then start dividing.
- E.g., run k-means on a cluster, then run again on resulting clusters.
  - A clustering analogue of decision tree learning.