

Lecture 27 announcements

- Project reports are due Tuesday, 12/4, 2:00 PM.
 - Template for the final report will be posted.
- On-campus final exam is Tuesday, 12/11, 2:00 PM - 4:00 PM, in SGM 124.
- DEN students will be notified by denexam@usc.edu of their exam locations.
- My office hours this week will be Friday 11 am - ~~12 noon~~ ^{12:30 PM.} (No office hours Wednesday.) Piazza and TAs office hours will be held also.

Lecture 27 outline

- Unsupervised Learning (part 3)
 - Agglomerative hierarchical clustering (part 2)
 - Nearest neighbor
 - Farthest neighbor
 - Clustering metrics: how to choose K

HIERARCHICAL CLUSTERING (part 2)

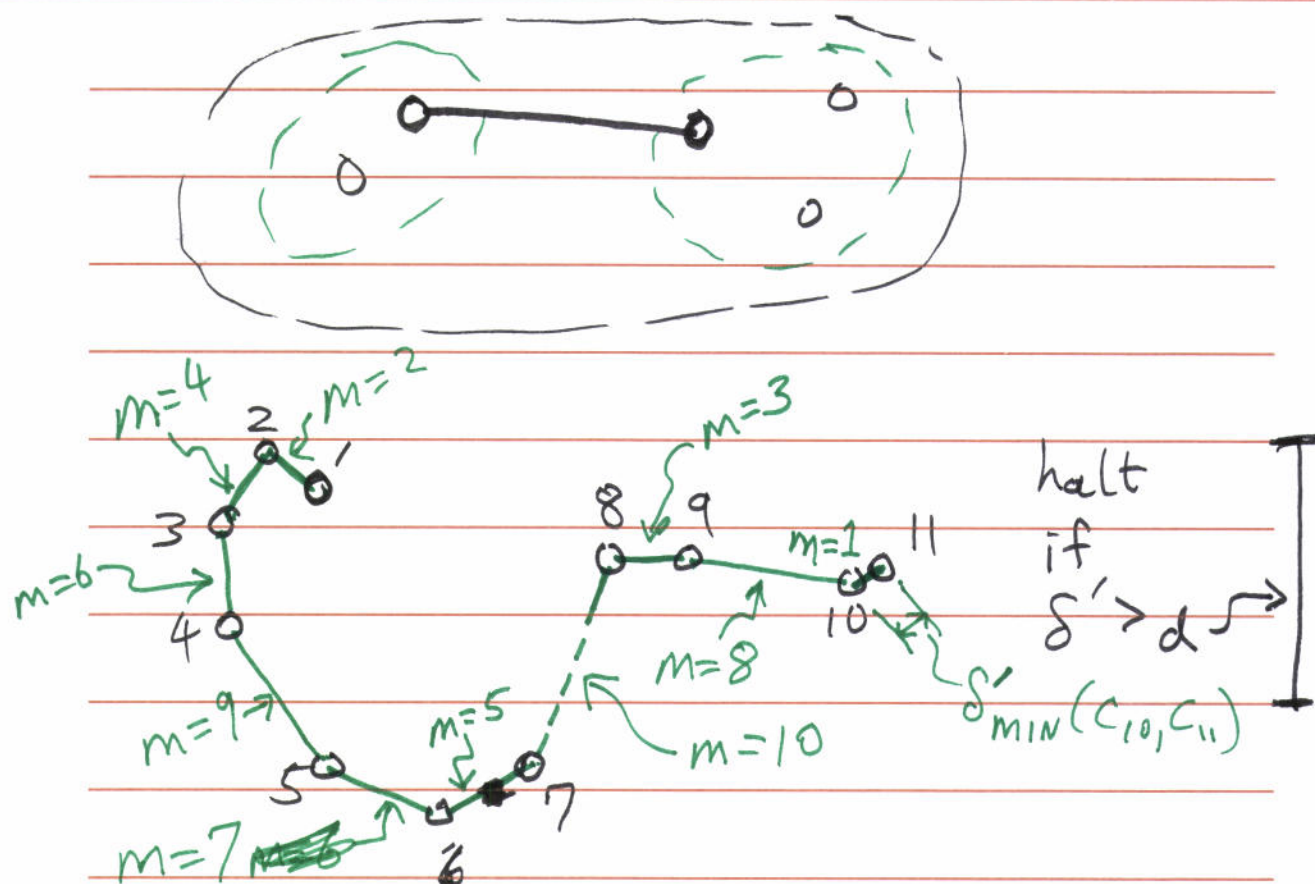
- STILL AGGLOMERATIVE.

NEAREST NEIGHBOR ALGORITHM (NN)

- ALSO "SINGLE LINKAGE" ALG.

→ USE δ_{\min}

MERGE RULE: JOIN TWO CLUSTERS BY
CONNECTING THE CLOSEST PAIR OF POINTS
(1 PT. FROM EACH CLUSTER).



INITIALIZE: $\hat{K} = N = 11$, $m = 1$
 $C_i = \{x_i\}$, $i = 1, \dots, 11$.
 $j', k' = 10, 11$.

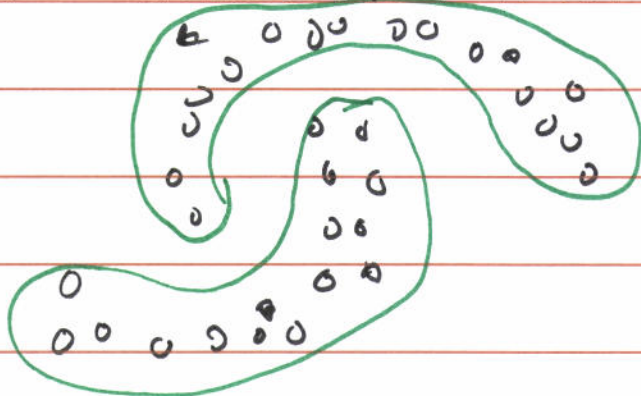
$$\hat{K} \leftarrow \hat{K} - 1 \Rightarrow \hat{K} = 10$$

~~INITIALIZE~~: $C_i = \{x_i\}$, $i = 1, \dots, 9$
 $C_{10} = \{x_{10}, x_{11}\}$

$m = 2$: $j', k' = 1, 2$.

N.N. (S.L.) TENDS TO BE GOOD AT LINKING

LONG STRINGS OF POINTS:



CHARACTERISTICS:

RESULTING GRAPH IS A TREE (NO CLOSED LOOPS)
CONTINUE TO $\hat{K}_{\text{FINAL}} = 1 \Rightarrow$ SPANNING TREE

USE $S_{\text{min}} \Rightarrow$ MINIMAL SPANNING TREE

(A SPANNING TREE THAT HAS MIN.
TOTAL LENGTH OF ALL EDGES).

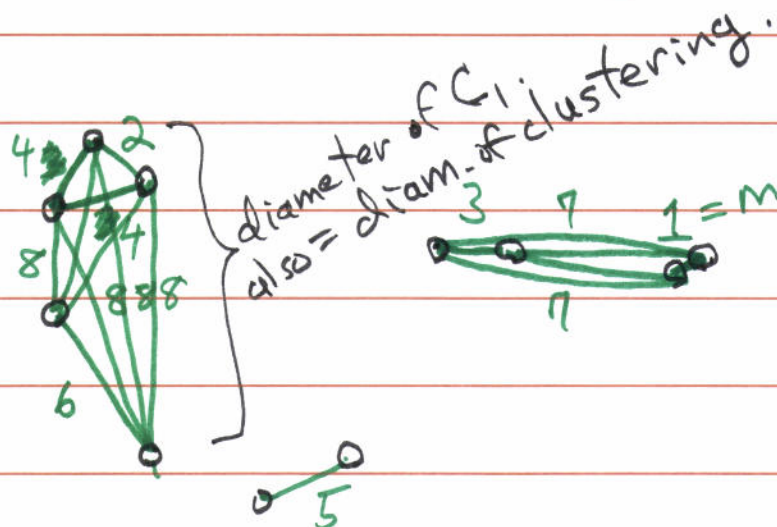
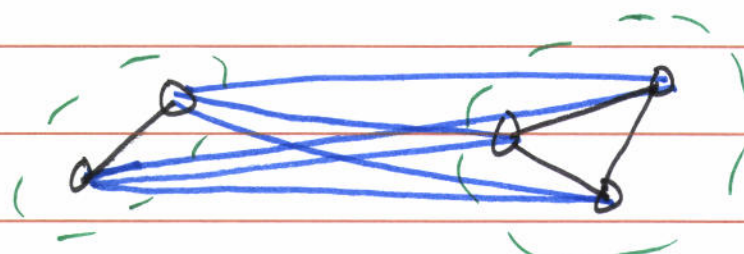
FURTHEST NEIGHBOR (FN) ALGORITHM

- ALSO "COMPLETE LINKAGE" ALG.

USE δ_{\max} .

MERGE RULE: CONNECT ALL NODES IN ONE
CLUSTER TO ALL NODES IN OTHER CLUSTER.

\Rightarrow EACH CLUSTER IS REPRESENTED BY A
FULLY CONNECTED SUBGRAPH.



HALT
WHEN
 $\hat{K} = 3.$

5
FN (C.L.) — PREFERS COMPACT, DENSE CLUSTERS.

CHARACTERISTICS:

DEF DIAMETER OF A CLUSTER: LENGTH OF
THE LONGEST EDGE.

DIAMETER OF A CLUSTERING (PARTITIONING):
LENGTH OF LONGEST EDGE OVER ALL
CLUSTERS.

F.N. ALG., AT EACH STEP, MERGES THE 2
CLUSTERS THAT GIVE THE SMALLEST

INCREASE IN DIAMETER OF THE CLUSTERING.

CRITERIA FOR CHOOSING K IN U.L.

[Ref: Xu PAPER, SEC. II M]

FOR PROBABILISTIC MIXTURE MODELS

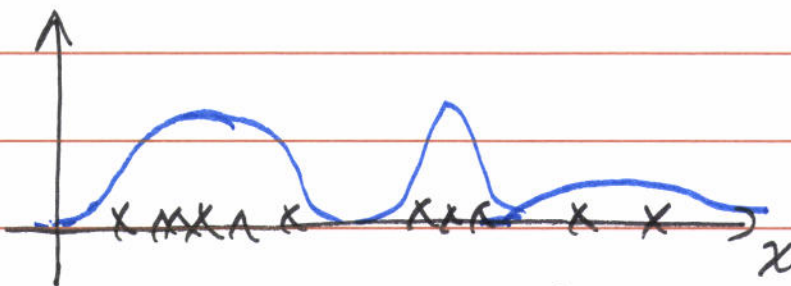
WE HAVE $\underline{\theta}$ = UNKNOWN PARAMETERS

\mathcal{H} = HIDDEN LABELS (CLUSTER ASSIGNMENTS)

$$\hat{\underline{\theta}}_{MLE}^{(K)} = \underset{\underline{\theta}}{\operatorname{argmax}} \underbrace{p(\mathcal{D} | \mathcal{H}, \underline{\theta})}$$

$$\prod_{i=1}^N p(x_i | \mathcal{H}, \underline{\theta})$$

$$\text{EACH } p(x_i | \mathcal{H}, \underline{\theta}) = \sum_{k=1}^K \pi_k^{(K)} p(x_i | \mathcal{H}, \underline{\theta})$$



→ Use $p(\mathcal{D} | \mathcal{H}, \hat{\underline{\theta}}_{MLE}^{(K)})$ AS CLUSTERING QUALITY MEASURE.

$$\hat{K} = \underset{K}{\operatorname{argmax}} p(D|\mathcal{H}, \hat{\theta}_{\text{MLE}}^{(K)})$$

optimal \uparrow
K