Tuesday, 11/17/2020

EE 660

MACHINE LEARNING FROM SIGNALS: FOUNDATIONS AND METHODS

Prof. B. Keith Jenkins

Lecture 25

Lecture 25 **EE 660** **Nov 17, 2020**

Announcements

- H9W12 is due on Friday
- Draft of the final report template has been posted (Week 13)
- Learning experience (course) evaluations are open - check your email

Today's topics

- Unsupervised learning (part 2)
  - Similarity / dissimilarity measures (part 2)
  - Hierarchical clustering
    - Concept and assumption
    - Agglomerative Hierarchical Clustering

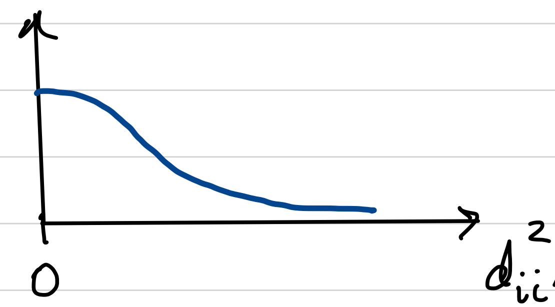
Let $s(\underline{x}_i, \underline{x}_{i'}) = s_{ii'}$ denote a similarity function

Can choose $s(\underline{x}_i, \underline{x}_{i'}) = g[d_{ii'}]$, g = any monotonically decreasing fcn.

e.g.:

$$s(\underline{x}_i, \underline{x}_{i'}) = (\max_{k,k'} d_{kk'}) - d_{ii'}$$

$$\text{or} \quad = \exp\left[\frac{-d_{ii'}^2}{\sigma^2}\right]$$



Other similarity measures:

For binary features $x_j \in \{0, 1\}$ (e.g., feature j contains an attribute or not).

$$s(\underline{x}_i, \underline{x}_{i'}) = \frac{\underline{x}_i^T \underline{x}_{i'}}{\underline{x}_i^T \underline{x}_i + \underline{x}_{i'}^T \underline{x}_{i'} - \underline{x}_i^T \underline{x}_{i'}}$$

For signals with (spatial or temporal) structure: correlation coefficient

e.g.: (Pearson)

Let j be index over spatial location or time

$$r_{ii'} = \frac{\sum_{j=1}^{D} (x_{ij} - \bar{x}_i)(x_{i'j} - \bar{x}_{i'})}{\left[\sum_j (x_{ij} - \bar{x}_i)^2 \sum_j (x_{i'j} - \bar{x}_{i'})^2\right]^{1/2}}$$

in which $\bar{x}_i = \frac{1}{D} \sum_{j=1}^{D} x_{ij}$

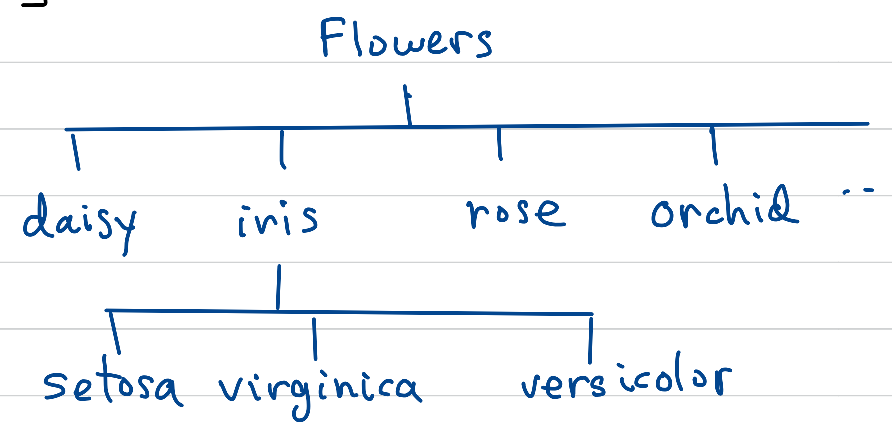
Note: $-1 \le r_{ii'} \le +1$ always.

$\Rightarrow$ Can let $s_{ii'} = r_{ii'}$ or $s_{ii'} = \frac{r_{ii'} + 1}{2}$

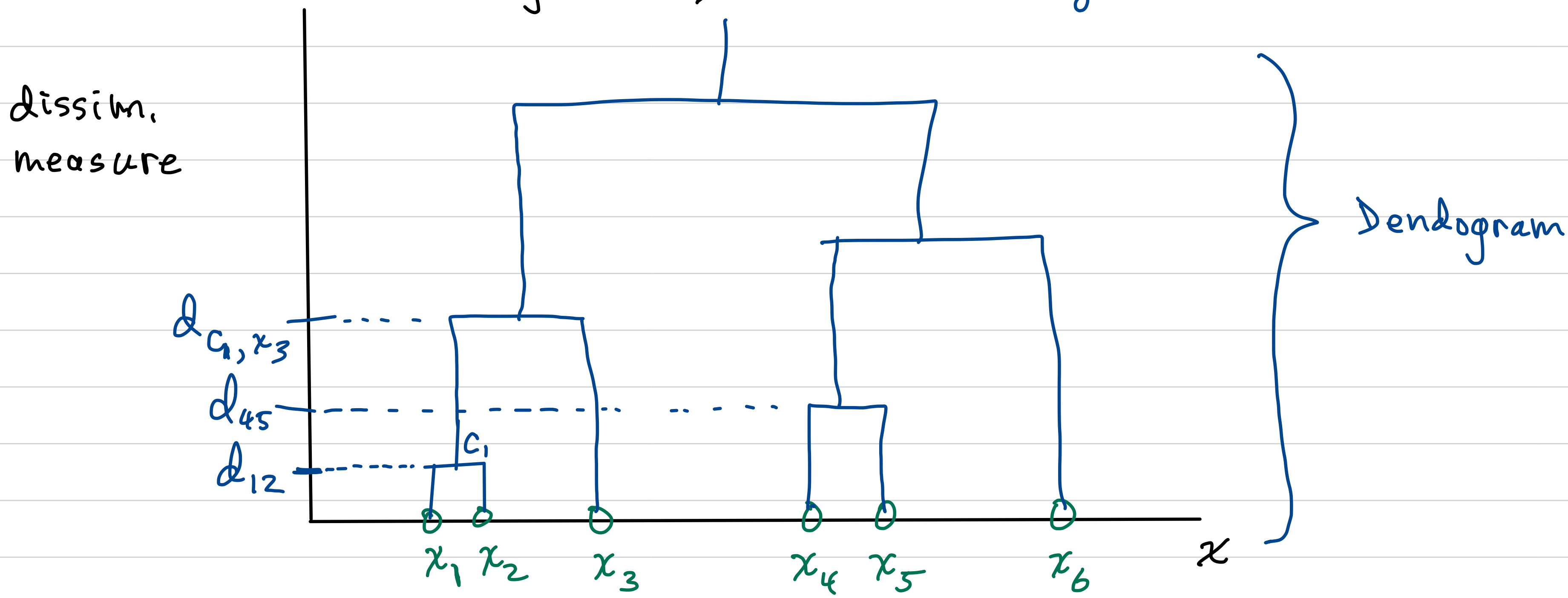
Or, $d_{ii'} = \frac{1 - r_{ii'}}{2}$

Hierarchical Clustering [Murphy 25.5]

Sometimes data is hierarchical in nature. →



Ex. of hierarchical clustering in 1D ↓



∃ Agglomerative (bottom-up) and divisive (top-down) approaches.

<u>Underlying assumption:</u> If 2 data points are in the same cluster at one level, then they are in the same cluster at all higher levels.

Agglomerative hierarchical clustering procedure

Let δ_{jk} = distance or dissimilarity between clusters C_j and C_k. [need dissim. measure]
$\hat{K}$ = current # of clusters

1. Choose halting condition (H.C.)
2. Initialize $\hat{K} = N$; cluster $C_i = \{\underline{x}_i\}$, $i = 1, 2, \cdots, N$; iteration $m = 1$.
3. Repeat until H.C. is met:
    4. Find nearest (most similar) pair of clusters:
    $$j', k' = \underset{j,k}{\operatorname{argmin}}\ \delta_{jk}, \quad \text{and} \quad \delta' = \min_{j,k} \delta_{jk} \quad \text{(resolve ties randomly)}$$
    5. Optionally output $m, \hat{K}, \delta', j', k'$
    6. If H.C. is based on δ', test for it (halt if true) [e.g., $\delta' \geq \delta_{halt}$]
    7. Merge clusters $C_{j'}$ and $C_{k'}$ to form new cluster C_ℓ [Apply merge rule]
    8. Update $\hat{K} = \hat{K} - 1$, iteration $m = m + 1$
    9. If H.C. is based on $\hat{K}$, then test for it (halt if true) [e.g., $\hat{K} \leq K_{halt}$]
10. Output final # of clusters $\hat{K}_{final}$, final clusters C_ℓ, $\ell = 1, 2, \cdots, \hat{K}_{final}$, minimum dissimilarities $\delta'(m)$.

If $\hat{K}_{final} = 1$, then the resulting hierarchy is a dendogram.

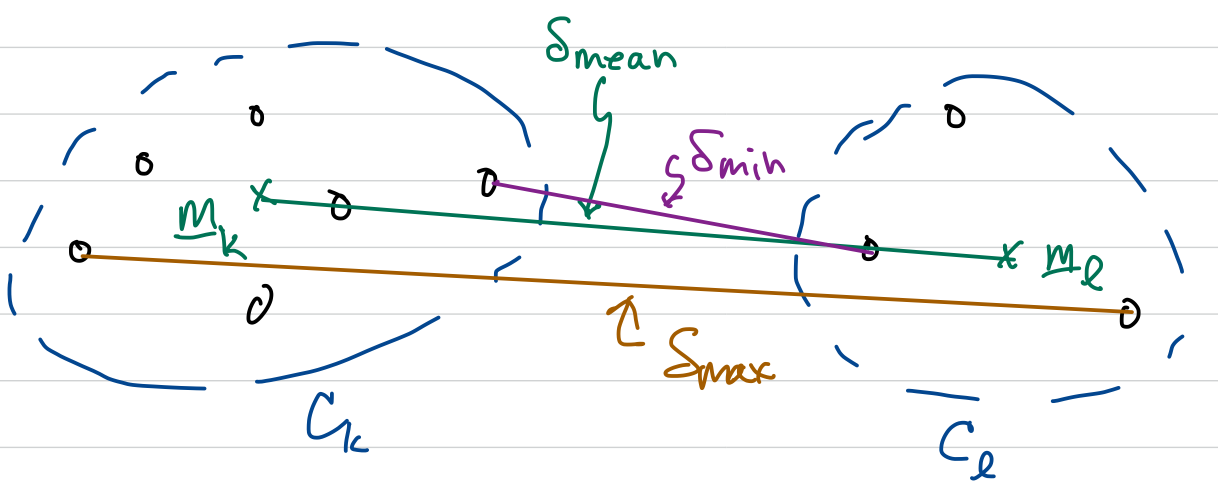
Useful distance or dissimilarity measures between clusters C_k, C_ℓ.

$$\delta_{mean}(C_k, C_\ell) \triangleq \| \underline{m}_k - \underline{m}_\ell \|_2$$

$$\delta_{min}(C_k, C_\ell) \triangleq \min_{\substack{\underline{x} \in C_k \\ \underline{x}' \in C_\ell}} \| \underline{x} - \underline{x}' \|_2$$

$$\delta_{max}(C_k, C_\ell) \triangleq \max_{\substack{\underline{x} \in C_k \\ \underline{x}' \in C_\ell}} \| \underline{x} - \underline{x}' \|_2$$

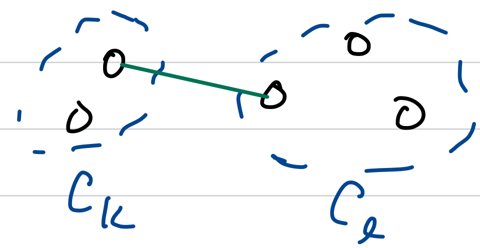
$$\delta_{avg}(C_k, C_\ell) \triangleq \frac{1}{N_k N_\ell} \sum_{\underline{x} \in C_k} \sum_{\underline{x}' \in C_\ell} \| \underline{x} - \underline{x}' \|_2$$



Nearest Neighbor Algorithm (Single Linkage Alg.)

Use δ_{min}

Merge rule: Join two clusters by connecting the closest pair $\underline{x}, \underline{x}'$.



Ex:

$\hat{K} = N = 11 \qquad m = 1$

$\hat{K} = 10 \qquad m = 2$

$\hat{K} = 9 \qquad m = 3$

$\vdots$

Example: (Nearest neighbor clustering)

H. C.:

$\delta' \geq \delta_{halt}$