



311 Introduction to Machine Learning

Summer 2024

Instructor: Ioannis Konstantinidis

Clustering



- K means
- Hierarchical
- Dendrograms

What is clustering?



Unsupervised Learning

Clustering is **unsupervised classification**: no predefined classes (only X, no y)

Cluster analysis

- Grouping a set of data objects into clusters
- Assign (instead of predict) values for y

Cluster: a collection of data objects

- Similar to one another within the same cluster
- Dissimilar to the objects in other clusters

K-Means clustering

accepting (word
article).
focus n point
converging rays of light,
heat, waves of sound, meet;
centre of activity or
intensity; pl focuses, foci; v
adjust, cause to converge;
concentrate; a focal
pertaining to focus

The k-Means Clustering Method

Given k , split samples into k groups of equal variance, minimizing a criterion known as the *inertia* or within-cluster sum-of-squares (WCSC):

$$\sum_{i=0}^n \min_{\mu_j \in C} (||x_i - \mu_j||^2)$$

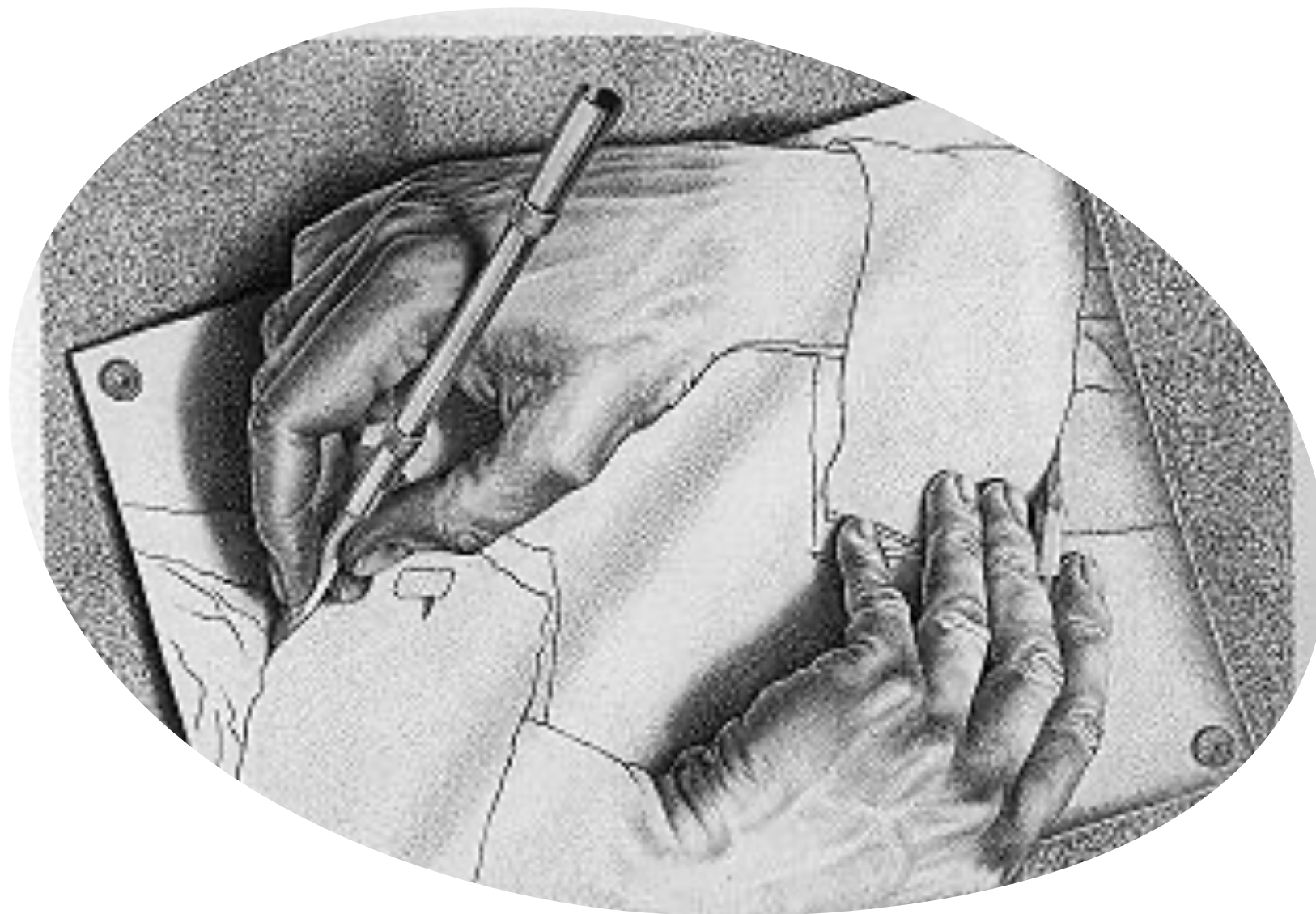
Initialize: Choose k initial centroids.

1. Assign each sample to its nearest centroid.
2. For each group created in Step 1, compute
 - the centroid of all of the samples assigned to that group.
 - the difference between the old and the new centroids for that group

Repeat until the new centroids do not change significantly.

Cautions

- Given enough time, K-means will always converge, however this may be to a local minimum.
- This is highly dependent on the initialization of the centroids.
- As a result, the computation is often done several times, with different initializations of the centroids.



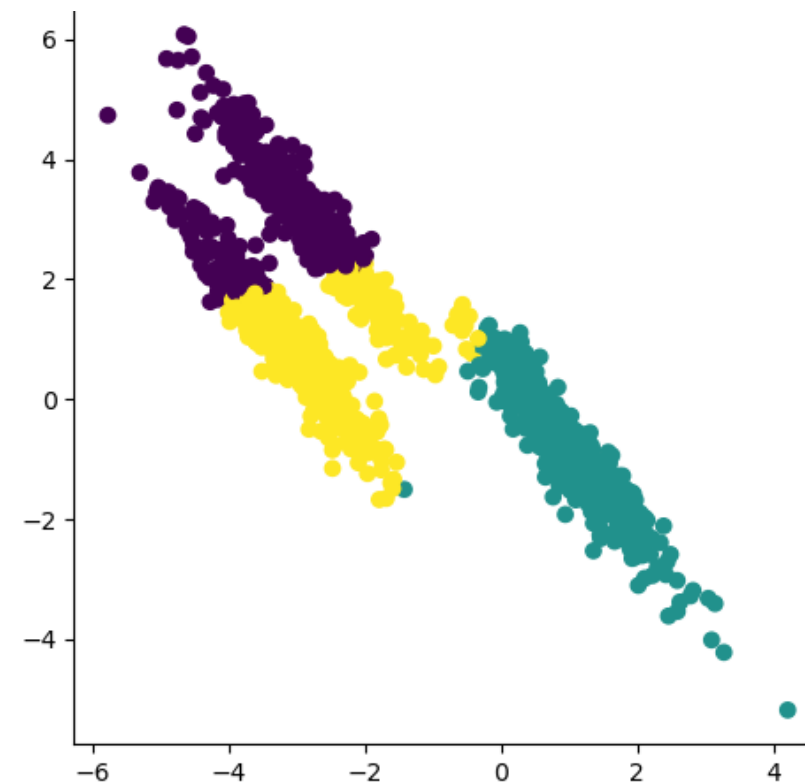
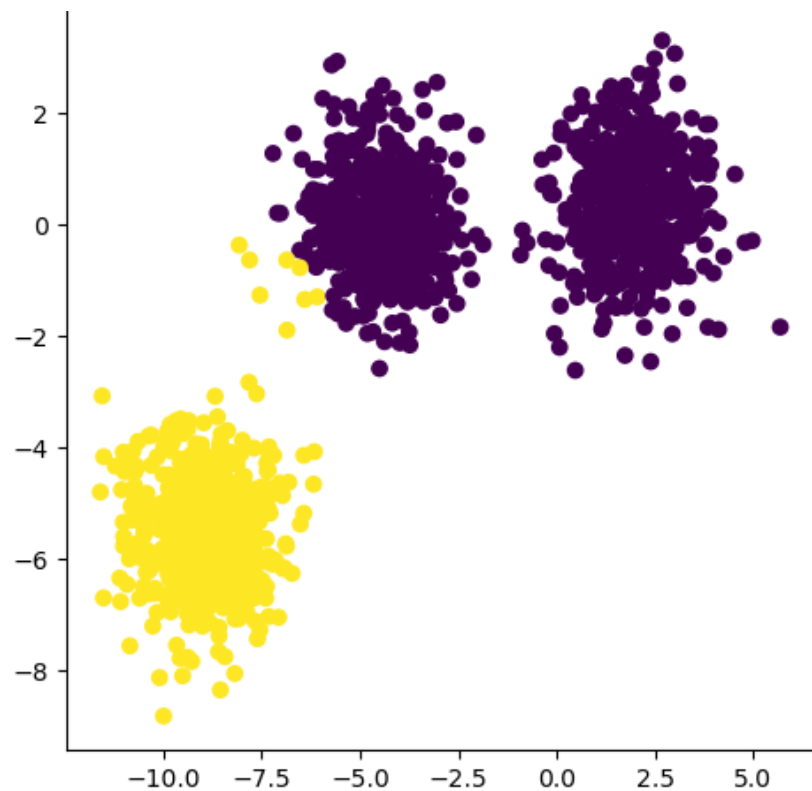
Hands-on
Example:

K means

Kmeans()

- *n_clusters*: The number of clusters to form = the number of centroids to generate
- *Init*: Method for initialization
 - Default is 'k-means++' : selects initial cluster centers for k-mean clustering in a smart way to speed up convergence.
- *n_init=10*: Number of time the k-means algorithm will be run with different centroid seeds. The final results will be the best output of *n_init* consecutive runs in terms of inertia.

Failure mode examples



Hard Clustering vs Hierarchical Clustering

- Hard: Divide objects into a set number of groups (clusters)
- Hierarchical: organize clusters in a hierarchy

Hierarchical clustering

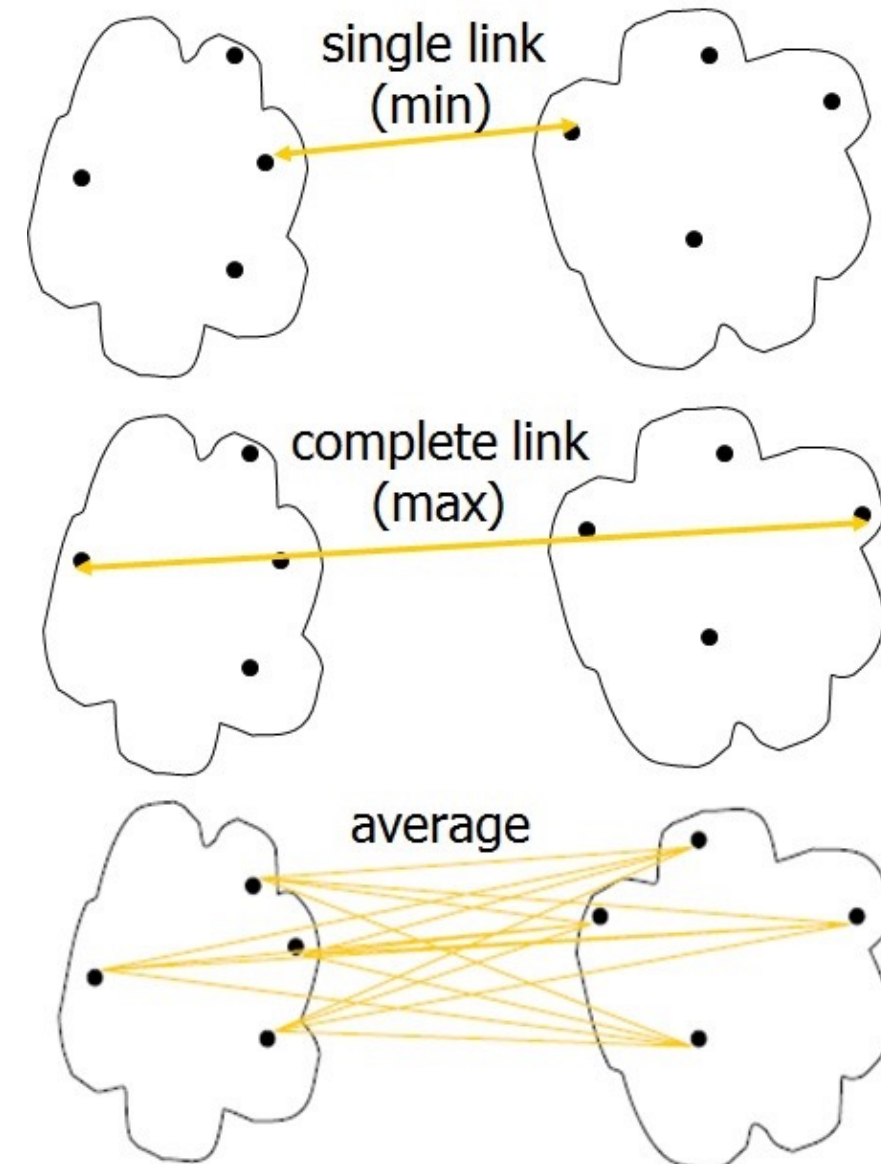
accepting (word
article).
focus n point
converging rays of light,
heat, waves of sound, meet;
centre of activity or
intensity; pl focuses, foci; v
adjust, cause to converge;
concentrate, a focal
pertaining to focus

Hierarchical clustering

- Top-down approach:
 - start with **all samples** in the dataset as one cluster
 - **divide** that cluster into subclusters based on criteria
 - repeat for each subcluster until done
- Bottom-up approach:
 - starts with **every single sample** in the dataset as its own cluster
 - **merge** samples into superclusters based on criteria
 - repeat for each supercluster collection until done

Nesting criteria: Linkage

- Single link: the distance between the nearest pair of points from each cluster
- Complete link: the distance between the farthest pair of points from each cluster
- Average link: the mean distance between all the pairs of points from each cluster
- Ward: Same distance concept as K-Means

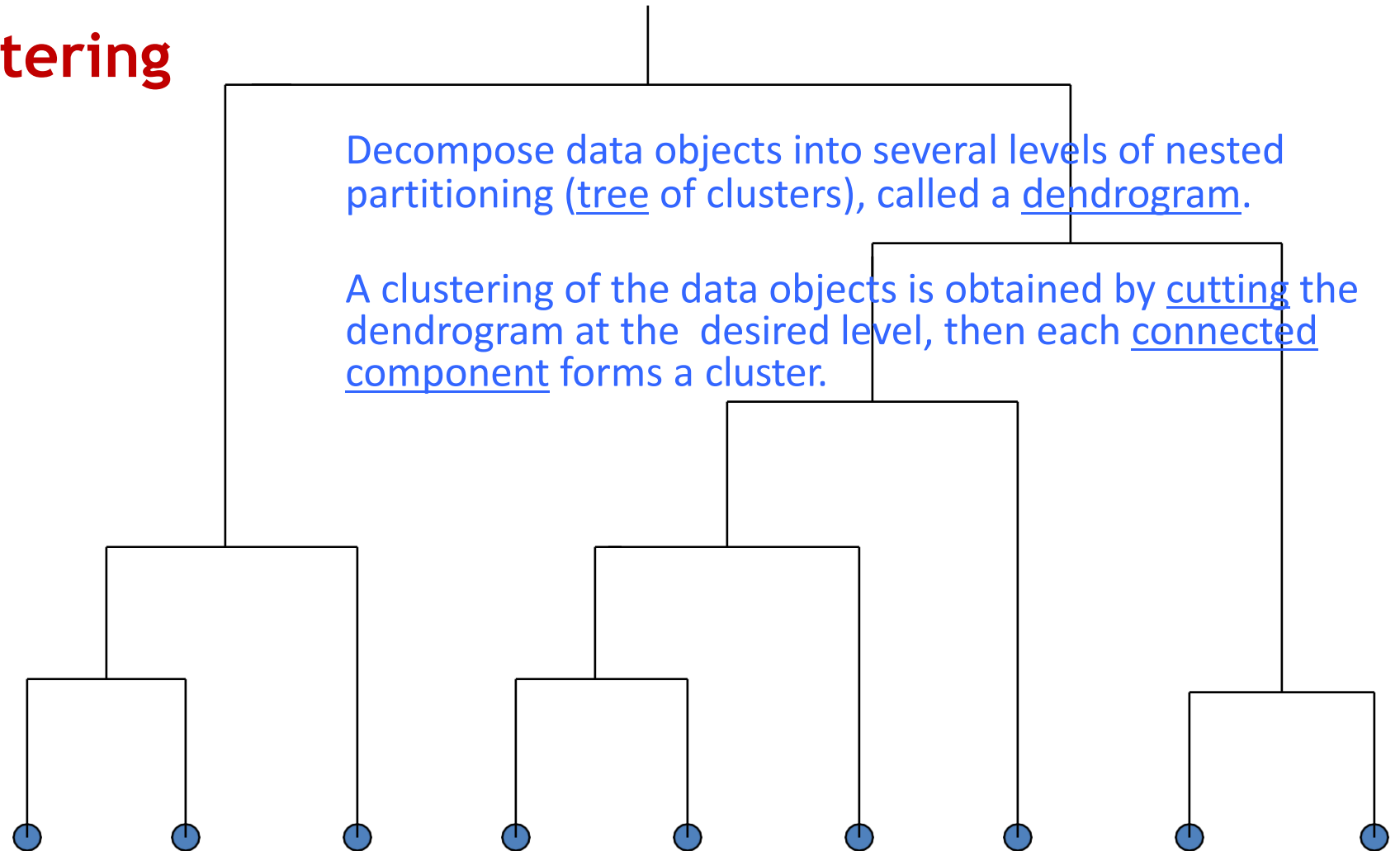


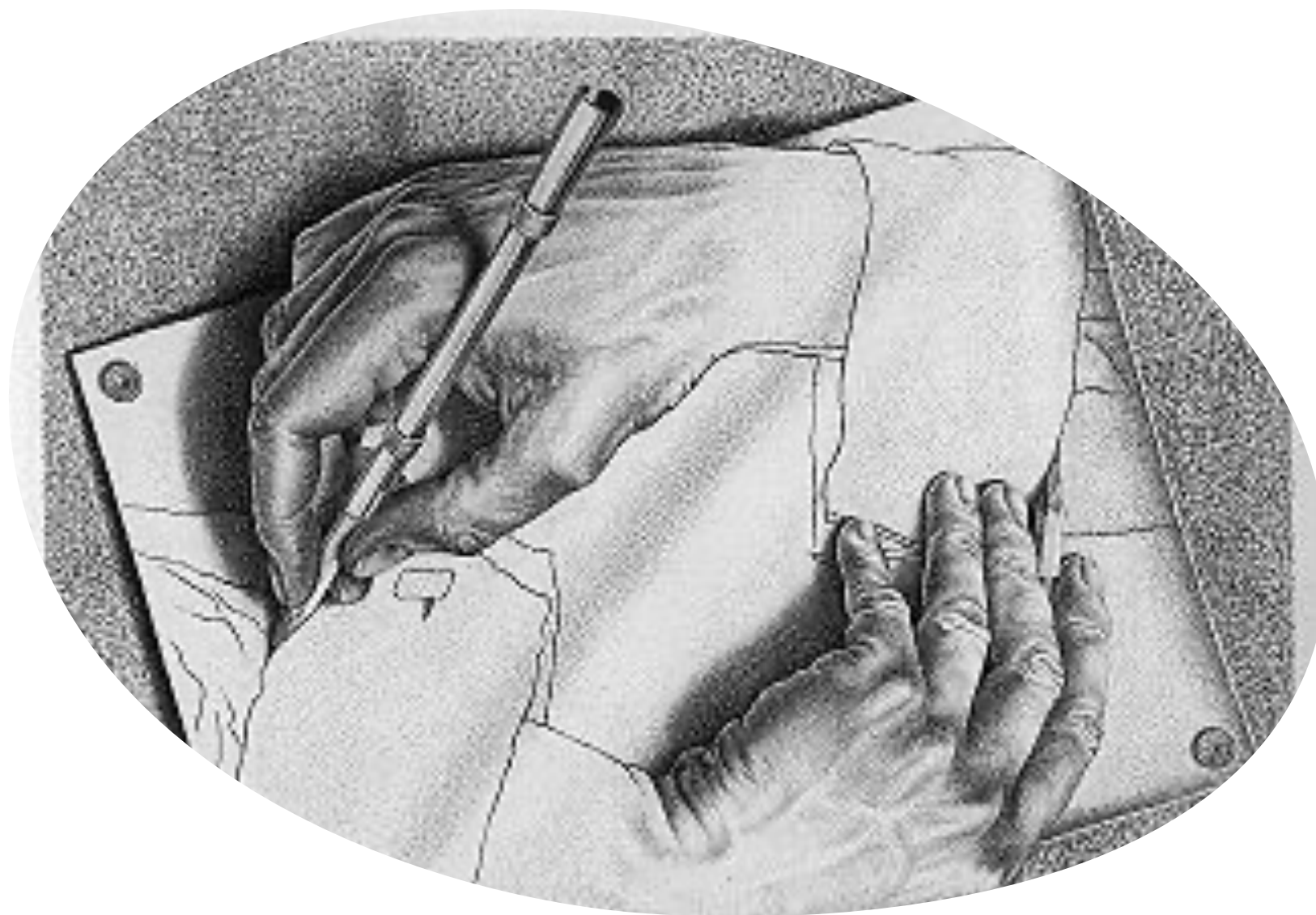
AgglomerativeClustering()

- *Affinity*: Metric used to compute the linkage. Can be “euclidean”, “l1”, “l2”, “manhattan”, “cosine”, or “precomputed”.
- *Connectivity*: Connectivity matrix. Defines for each sample the neighboring samples following a given structure of the data.
- *Linkage*: Which linkage criterion to use. The linkage criterion determines which distance to use between sets of observation. The algorithm will merge the pairs of cluster that minimize this criterion.

Hierarchical Clustering

- We can visualize the results of clustering as a dendrogram.
- This helps us in deciding when we want to stop clustering further (how "deep") by setting "depth" with some threshold.





**Hands-on
Example:**

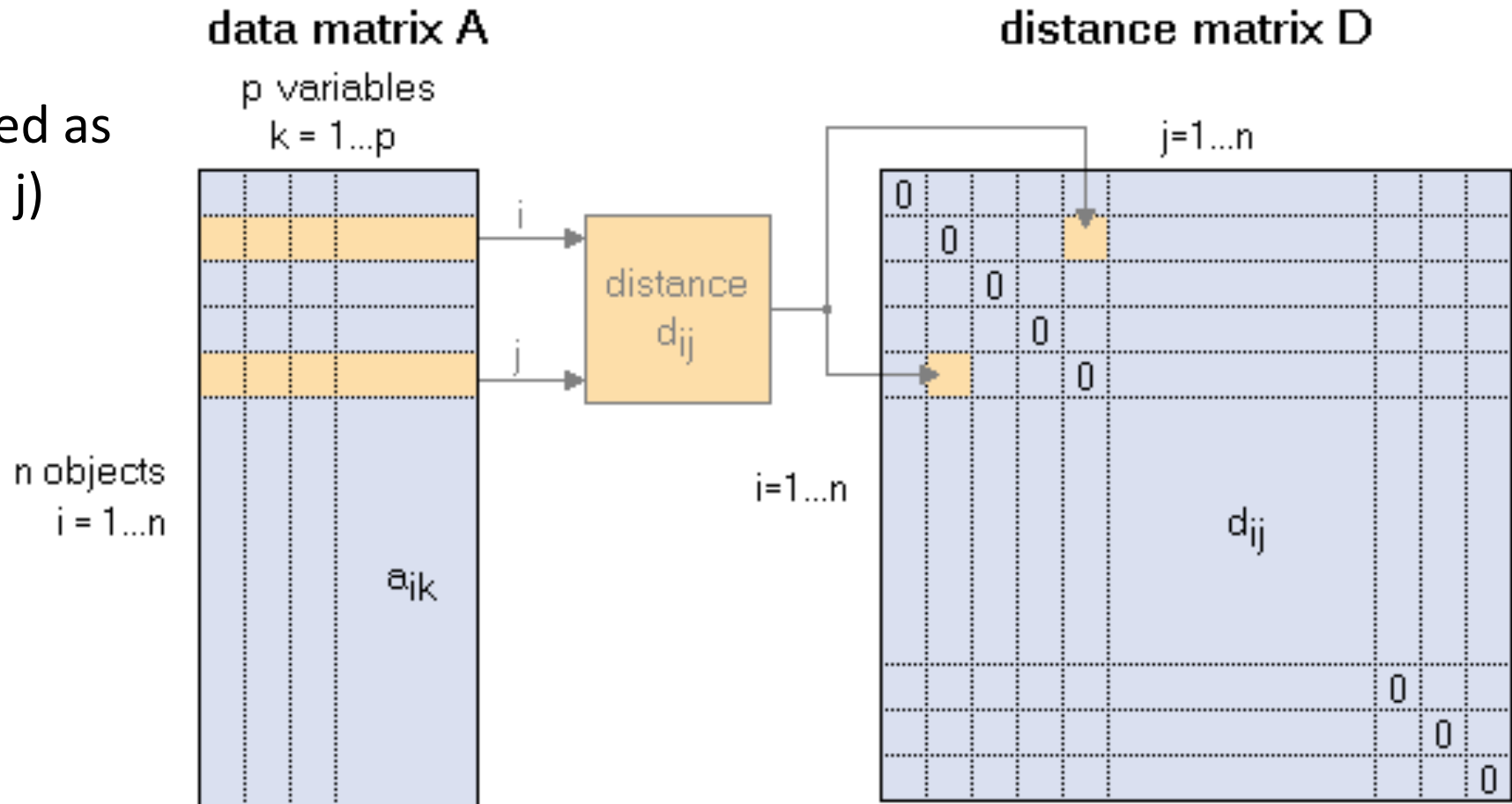
**Hierarchical
Dendrograms**

Similarity matrix

Dissimilarity is expressed as
a distance function $d(i, j)$

Could be Minkowski,
e.g.,

- Euclidean,
- manhattan



Final logistics



Exam Logistics

Course Feedback

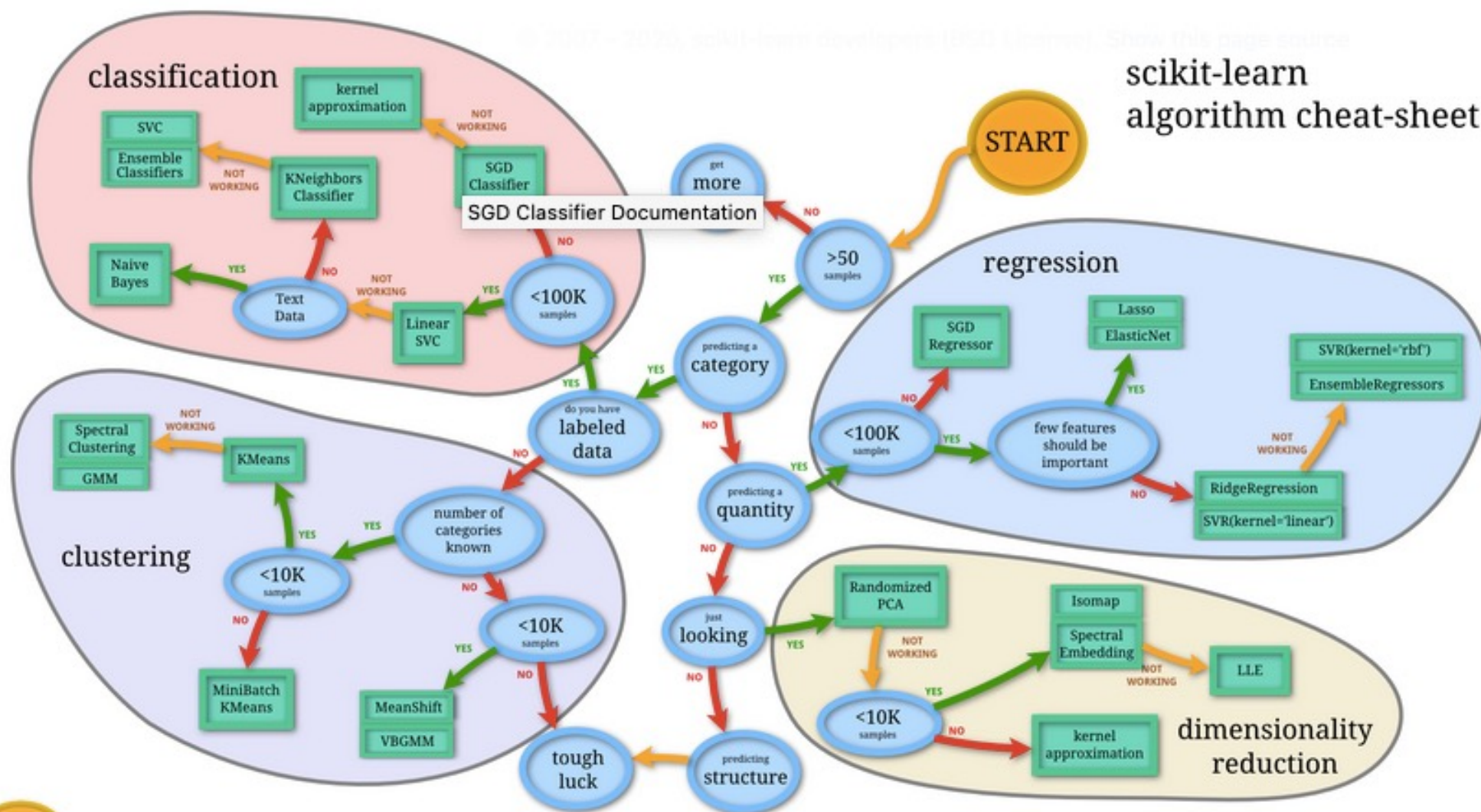
Please fill out the feedback form in Moodle!



Where to go from here?



Sandi Hemmerlein | Painted Rock at Hwy 58 Corrido Plain



A large target graphic with concentric circles in shades of red and a dark blue center. The text 'That's all Folks!' is written in a white, cursive script across the center of the target.

That's all Folks!



Stay in touch...

... and let us know how you end up
using what you learned!