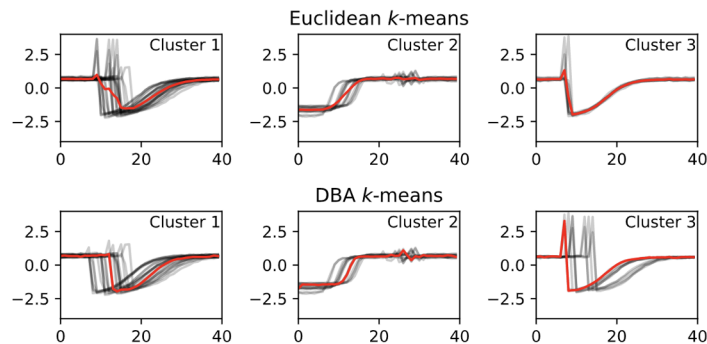# Time Series Clustering

K-means

# k-means clustering for time series: Similarity function

- Many variants of the algorithm are available, two of them:

  - Euclidean k-means
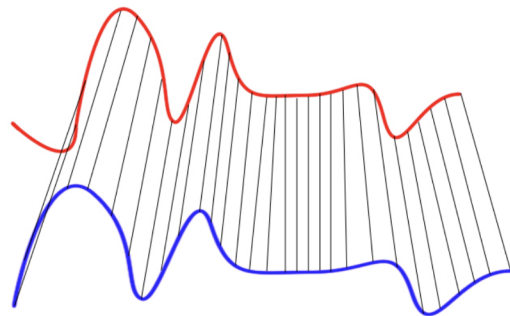
  - DBA-k-means (for DTW Barycenter Averaging)

Training set series for a cluster (in black) and centroid of the cluster (in red).

Dynamic Time Warping(DTW)

- Euclidean distance - not invariant to time shifts.

- Alternate similarity measure between time series, which may vary in speed.

- Based on the Levenshtein distance (edit distance)
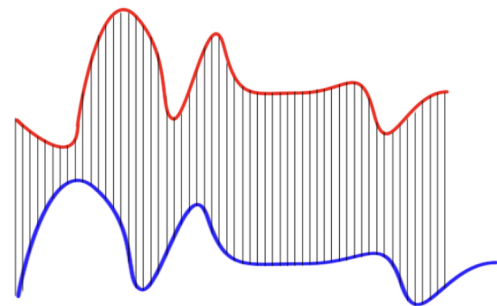
Euclidean Matching

Dynamic Time Warping Matching

Similarity Functions

Let A = $< a_1, a_2, \ldots, a_T >$ and B = $< b_1, b_2, \ldots, b_T >$ be two sequences, and let δ be a distance between elements (or coordinates) of sequences.

- Euclidean Distance: $D(A,B) = \sqrt{\delta(a_1,b_1)^2 + \cdots + \delta(a_T,b_T)^2}$

- Dynamic time warping: $D(A_i,B_j) = \delta(a_i,b_j) + \min \begin{Bmatrix} D(A_{i-1},B_{j-1}) \\ D(A_i,B_{j-1}) \\ D(A_{i-1},B_j) \end{Bmatrix}$

- Numerical Example [YT video]

k -means and Dynamic Time Warping

- Cluster centers (aka centroids) are computed with respect to DTW.

- Clusters gather **time series of similar shapes**

  - DTW deals with time shifts.

- Capable of retrieving **sensible average shape** of the temporal shifts in the cluster.

Further Readings

- https://www.cs.unm.edu/~mueen/DTW.pdf

- https://dl.acm.org/doi/10.1145/2939672.2945383