Time Series Data Clustering: A brief survey

Jim Howard

January 18, 2011

Contents

Contents

What

- Time series is a sequence of data points measured at successive times.

$$X = \{x^{(1)}, x^{(2)}, ..., x^{(N)}\}$$

$$x^{(i)} = \{x_1^{(i)}, x_2^{(i)}, ..., x_M^{(i)}\}$$

  - Indexing the t element of time series i would be noted as $x_t^{(i)}$

- Cluster

$$C = \{c^{(1)}, c^{(2)}, ..., c^{(K)}\}$$

- Center: $\overline{c^{(i)}}$

Why?

Contents

P-Norm / Root Mean Square

$$d_M(x^{(i)}, x^{(j)}) = \sqrt[p]{\sum_{t=1}^{M} (x_t^{(i)} - x_t^{(j)})^p}$$

- P is typically 1, 2 or inf

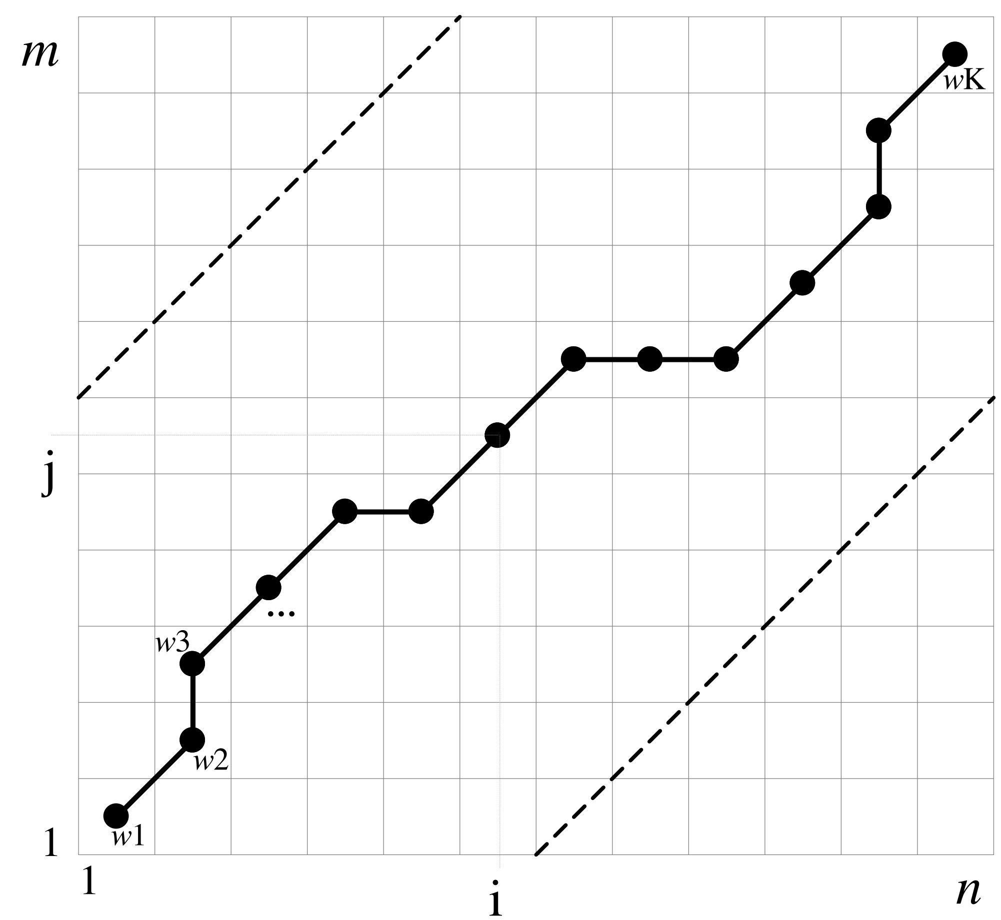
$$d_{rms}(x^{(i)}, x^{(j)}) = \frac{\sqrt[2]{\sum_{t=1}^{M} (x_t^{(i)} - x_t^{(j)})^2}}{M}$$

- Root Mean Square is the normalized form of the 2-norm

Dynamic Time Warping (DTW)

- Finds the minimum cost warping path between two time series
- Can be used on time series of different lengths

$$DTW(x^{(i)}, x^{(j)}) = \forall\, W \;\; min\{ \frac{\sqrt{\sum_{z=1}^{Z} w_z}}{Z} \}$$



Short Time Series Distance

- Looks only at the euclidian distance between the slope calculated between each pair of points within a time series

$$d_{STS}(x^{(i)}, x^{(j)}) = \sum_{t=1}^{M-1} ((x_{t+1}^{(i)} - x_t^{(i)}) - (x_{t+1}^{(j)} - x_t^{(j)}))^2$$

Contents

- What and Why
- Time Series Distance Measures
- Cluster Scoring
- Clustering Algorithms
- Data Modification
- Dataset and Results
- Conclusion

Inter-Cluster Distance

- Sum of the distance of all time series assigned to a cluster to that cluster center.

$$d_{ic} = \sum_{x \in c^{(i)}} d(x, \bar{c^{(i)}})$$

Silhouette (Matlab function)

- Average of the ratio of the inter-cluster distance and the distance to the next closest cluster center

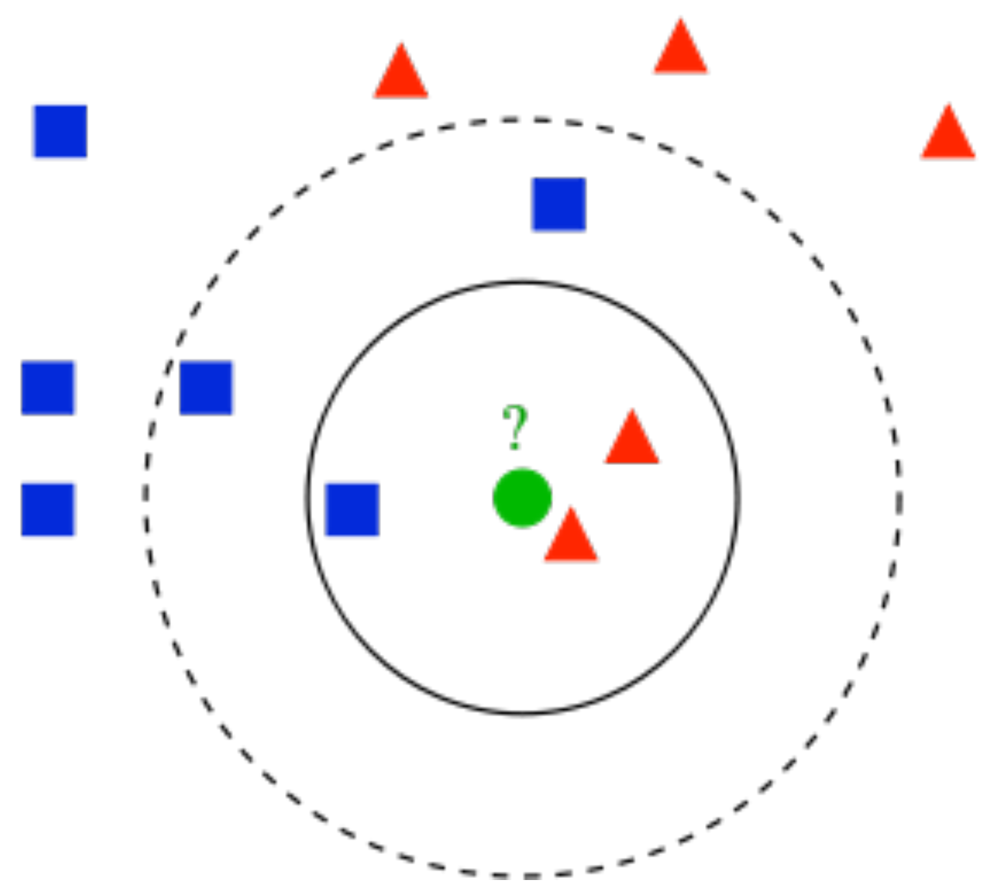
$$S = \frac{\sum_X \frac{d(x,\bar{c})}{d_n(x,C)}}{N}$$

- where $d_n(x, C)$ is the distance between a time series x and the second closest cluster C

Contents

k-Nearest Neighbor (kNN)

- Determines cluster based on majority vote
- Requires classified data



  - k = 3 circle is red
  - k = 5 circle is blue

k-Means

$$d_{ic} = \sum_{x \in c^{(i)}} d(x, \bar{c^{(i)}})$$

$$d_{tic} = \sum_{c \in C} d_{ic}(c)$$

- Initialize K clusters
- repeat
  - calculate cluster centers
  - determine new time series assignment
  - calculate d_{tic}
- until $d_{tic,new} - d_{tic,old} < threshold$

Agglomerative Clustering

- Initialize the set of clusters with each cluster containing exactly one time series
- repeat
  - calculate the pair-wise average distance between each cluster
  - merge the two clusters with the shortest distance
- until the number of clusters is equal to 1

$$d_{pair-wise}(c^{(i)}, c^{(j)}) = \frac{1}{|c^{(i)}| * |c^{(j)}|} \sum_{x \in c^{(i)}} \sum_{y \in c^{(j)}} d(x, y)$$

Clustering Algorithms

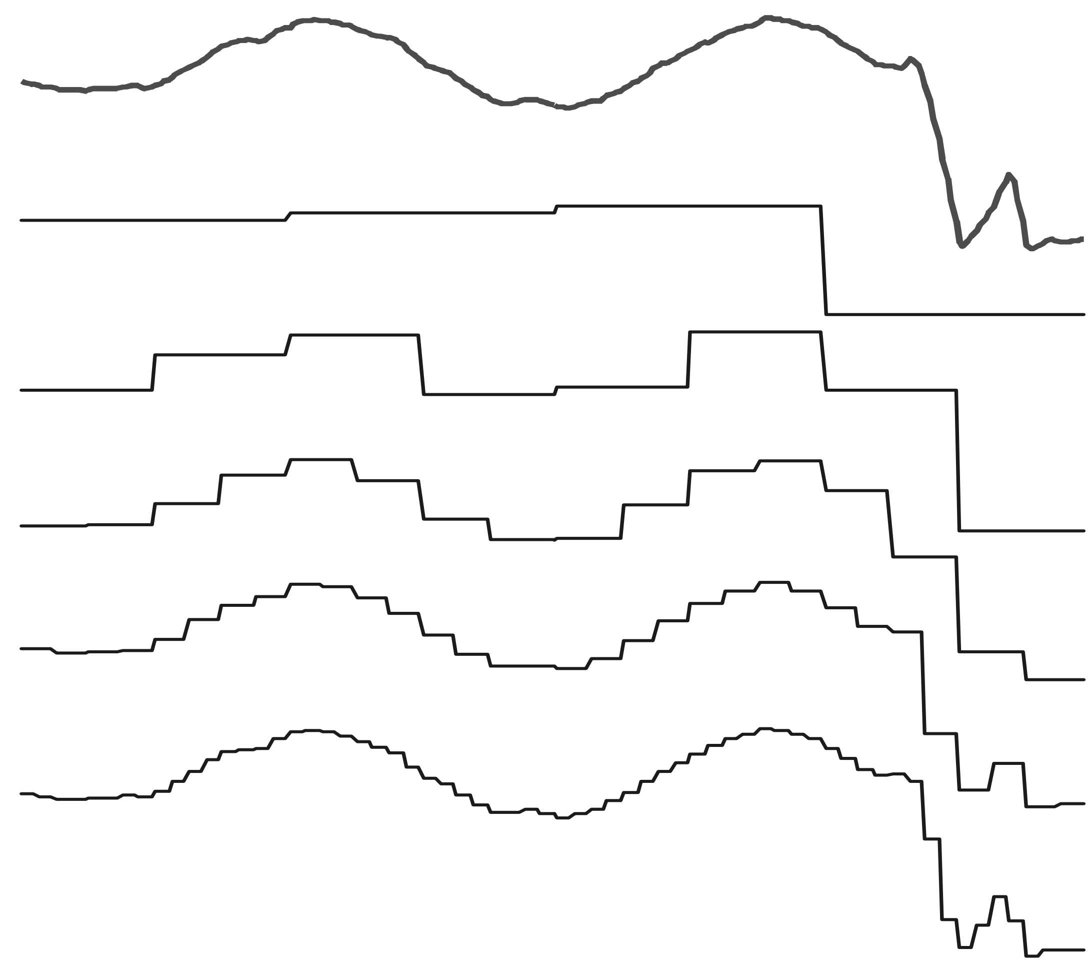
- kNN
  - Fast to run and implement
  - Requires classified data
- kMeans
  - Fast training time
  - Must know the number of models to stop on
  - Gets stuck in local minima frequently (varies based on initial random clusters)
- Agglomerative Clustering
  - Slowest training time
  - Useful when ideal number of clusters is not known
  - Deterministic

Contents

- What and Why
- Time Series Distance Measures
- Cluster Scoring
- Clustering Algorithms
- Data Modification
- Dataset and Results
- Conclusion

Haar Wavelet

- Takes the average of adjacent time series values.



- Example:

$$x = \{5, 3, 8, 4\}$$

one level transform

$$x = \{4, 6\}$$

Hidden Markov Models (HMM)

- Represent a cluster as a HMM trained on the set of time series assigned to that cluster.
- Distance Measure is then the log likelihood of a time series to a given HMM
- Typically require some initial “smart” guess of clusterings for usage with clustering algorithms (kMeans especially)

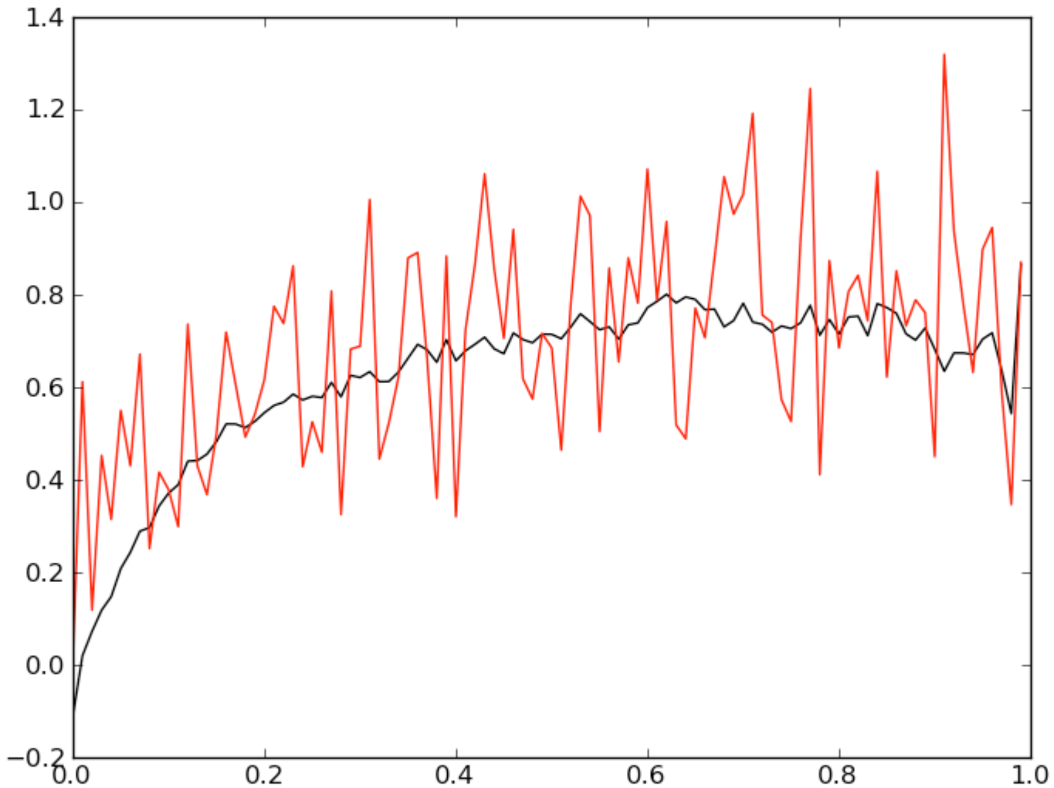
Locally Weighted Linear Regression

- applies linear regression to each point in the time series

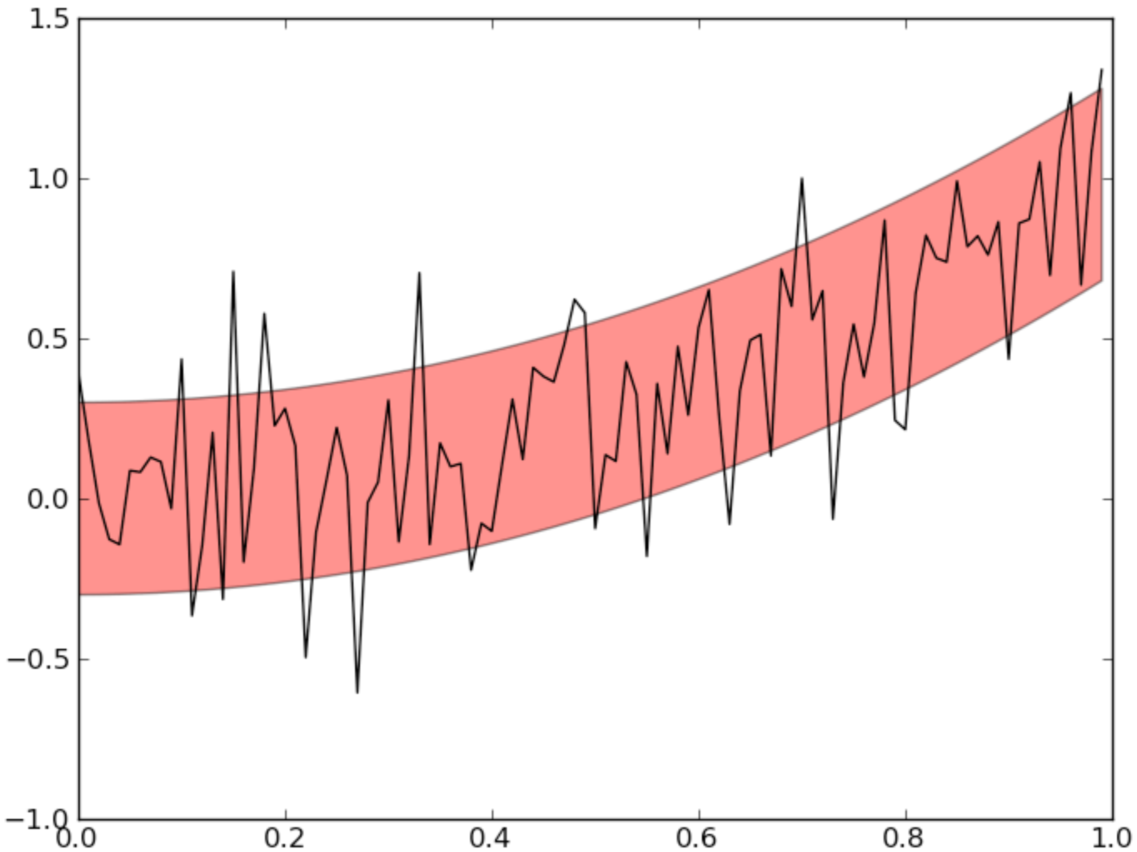
$$w^{(i)} = exp(-\frac{(x^{(i)} - x)^2}{2\tau^2})$$

- Uses batch gradient descent for point level linear regression.

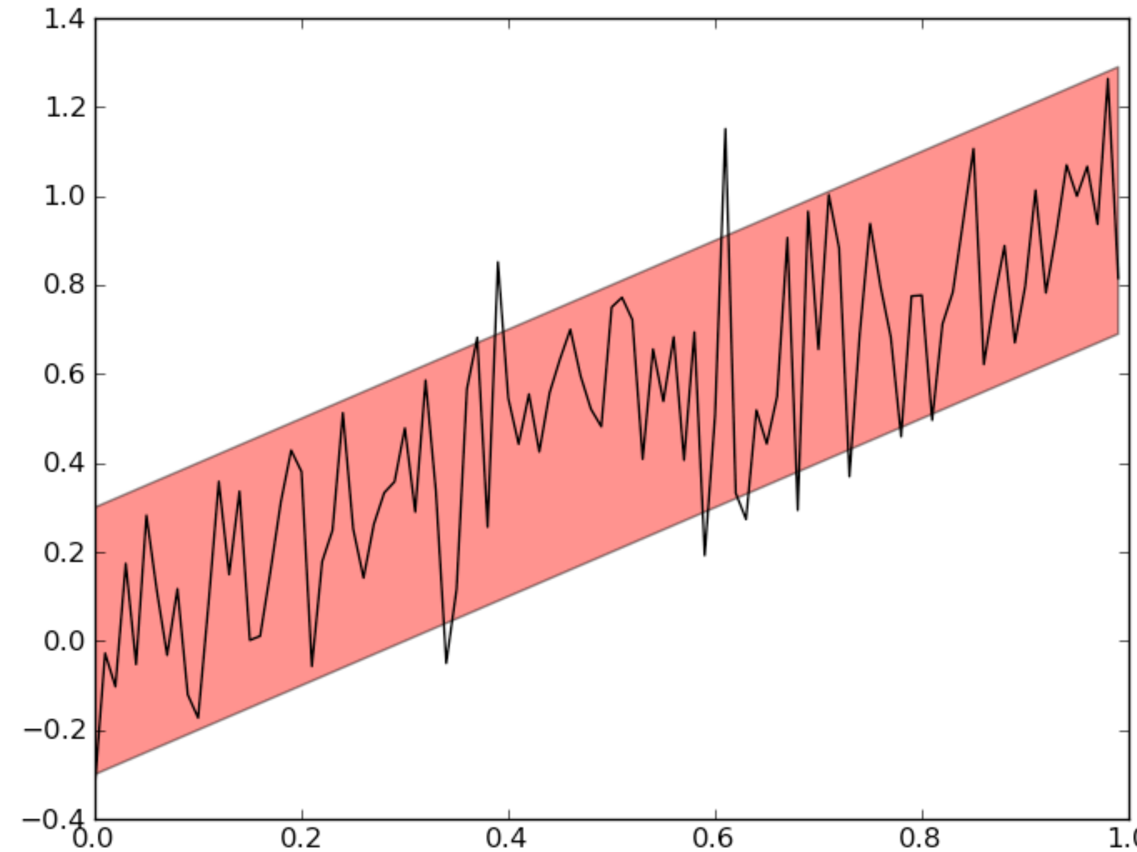
- Drop the offset



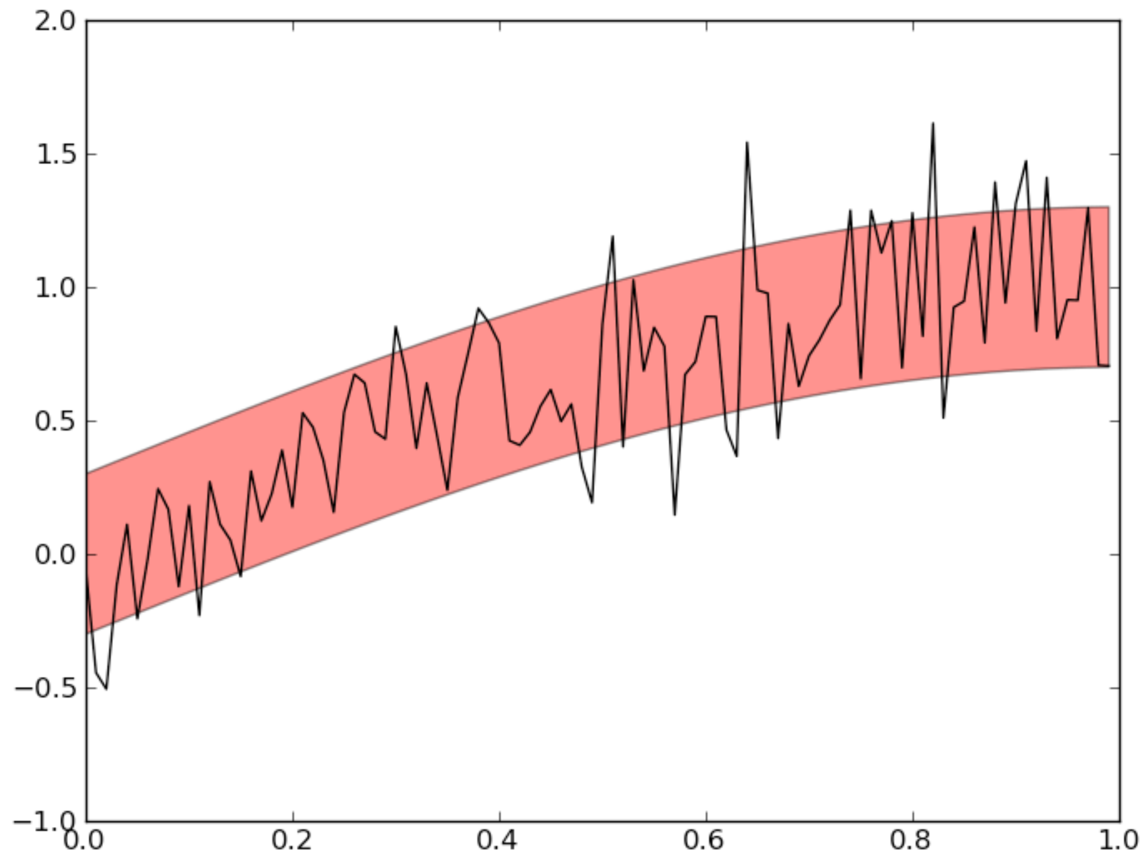
Contents



1.5
1.0
0.5
0.0
−0.5
−1.0
0.0
0.2
0.4
0.6
0.8
1.0



1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
0.0
0.2
0.4
0.6
0.8
1.0



2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
0.0
0.2
0.4
0.6
0.8
1.0

Results

Clustering Algorithm	Pre-processing	Distance Metric	Train Accuracy	Test Accuracy	Silhouette	Inter-cluster dist
kNN	Haar	L2	0.67	0.57	NA	NA
kNN	Haar	Linf	0.68	0.55	NA	NA
kNN	Haar	DTW	0.63	0.57	NA	NA
kMeans	Haar	L2	0.59	0.59	0.65	NA
kMeans	Haar	Linf	0.57	0.61	0.78	NA
kMeans	Haar	DTW	0.59	0.61	0.33	NA
kNN	None	L2	0.57	0.71	NA	NA
kNN	None	Linf	0.53	0.56	NA	NA
kNN	None	DTW	0.55	0.58	NA	NA
kMeans	None	L2	0.59	0.57	0.76	2.62
kMeans	None	Linf	0.59	0.61	0.78	2.63
kMeans	None	DTW	0.57	0.54	0.76	2.64

Results (Cont)

Clustering Algorithm	Pre-processing	Distance Metric	Train Accuracy	Test Accuracy	Silhouette	Inter-cluster dist
Agglom	None	L2	0.58	0.57	NA	NA
Agglom	None	Linf	0.56	0.55	NA	NA
Agglom	None	DTW	0.63	0.57	NA	NA
KNN	LWLReg	DTW	0.69	0.51	NA	NA
KMeans	LWLReg	L2	0.57	0.59	0.04	0.4
KNN	LWLReg	L2	0.69	0.56	NA	NA

Contents

Conclusions

- Despite the numerous different distance measures, nothing seems to considerably beat DTW (confirmed by Ding, Trajcevski, et al, 2008)
- Each clustering algorithm should be selected based on the clustering task.