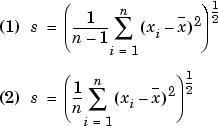
Sorting + Clustering Algorithm Specifics

Scott Grimes

(Note: Data should band-pass filtered prior to exportation from spike2 (artifacts are also removed))

STEP 1

The standard deviation for each channel in the file (1- electrode, 2-stereotrode, 4-tetrode) is computed using:



If the absolute value of the data is greater than USER\_GIVEN\_CUTOFF\*std+mean of the data: a spike is detected and saved using user-defined size of spike in ms to the left and right of the threshold.

Spikes are aligned to one another using their maximum absolute value.

STEP 2

The principal component loadings are determined, and the first N loadings are used for sorting where N is given by:

Latent = the eigenvalues of the covariance matrix of NxP (N spikes, P points)

M = cumsum(latent)/sum(latent)

N = first cell of M with value greater than .98

(using N principal components accounts for 98% of the variance of the spikes).

STEP 3

K-Means clustering is used with an initial number of clusters 15 (or higher, determined by the user).

K-Means:

Given an initial set of *k* means **m**1(1),…,**m***k*(1) the algorithm alternates between two steps:

**Assignment step**: Each observation is assigned to the cluster with the nearest mean:

S_i^{(t)} = \left\{ \mathbf x_j : \big\| \mathbf x_j - \mathbf m^{(t)}_i \big\| \leq \big\| \mathbf x_j - \mathbf m^{(t)}_{i^*} \big\| \text{ for all }i^*=1,\ldots,k \right\} 

**Update step**: The new means of the observations in each cluster are calculated

\mathbf m^{(t+1)}_i = \frac{1}{|S^{(t)}_i|} \sum_{\mathbf x_j \in S^{(t)}_i} \mathbf x_j 

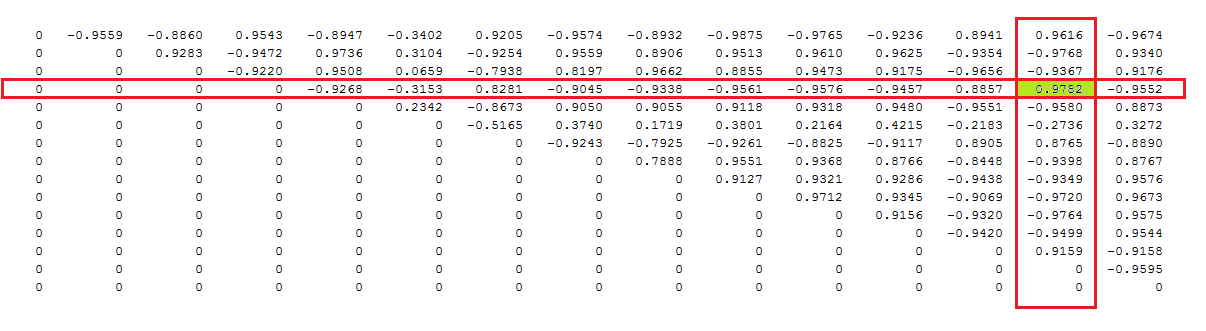
This continues until the observations in each cluster become stable. (This process is also known as Lloyd’s algorithm)

STEP 4

A similarity matrix is computed for M clusters where Mij is the similarity between clusters i and j.

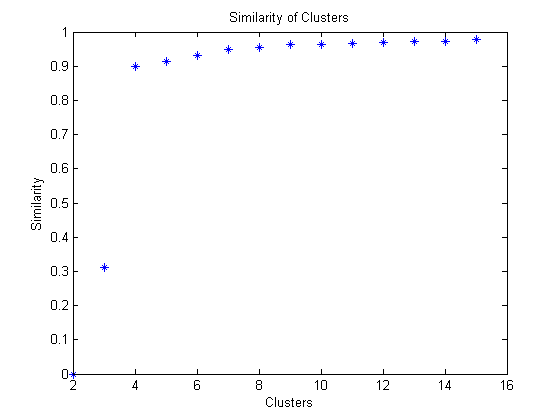


For M-1 iterations the clusters with the greatest similarity are combined, and a new similarity matrix is computed…



EX. Cluster 14 and cluster 4 are combined.

STEP 5

When the number of clusters is reduced from M to 2, the maximum similarity between clusters for every iteration of M is calculated.

The first number given by:

Absolute value of y > mean(y)+3\*std(y), where y = diff( vector containing max similarities from cluster 15 to 2)

Becomes the number of clusters and the cluster indexes are determined.