Connectivity-Based Clustering (Hierarchical Clustering)

[View all machine learning examples](https://www.mathworks.com/solutions/machine-learning/examples.html)

This example demonstrates a clustering technique that takes advantage of inherent connectivity in the data. Clustering algorithms that cluster data with any connectivity information could generate spurious clusters that may have no practical meaning.

Generate Swiss Roll Data

rng(10); % for reproducibility

N = 1500;

noise = 0.05;

t = 3\*pi/2 \* (1 + 2\*rand(N,1));

h = 11 \* rand(N,1);

X = [t.\*cos(t), h, t.\*sin(t)] + noise\*randn(N,3);

Cluster Data Using K-Means and Agglomerative Clustering

figure('units','normalized','Position',[0.2 0.4 0.55, 0.35]),

subplot(1,2,1)

c = clusterdata(X,'linkage','ward','maxclust',6);

scatter3(X(:,1),X(:,2),X(:,3),[],c,'fill','MarkerEdgeColor','k');

view(-20,5)

title('Agglomerative Clustering')

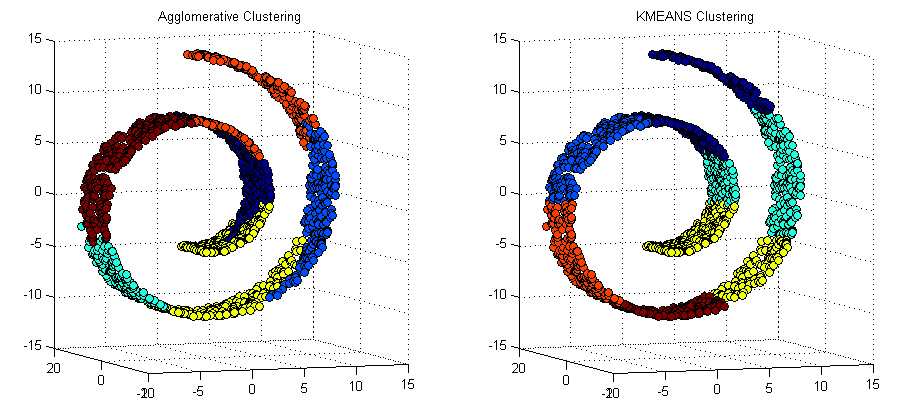
subplot(1,2,2)

c = kmeans(X,6);

scatter3(X(:,1),X(:,2),X(:,3),[],c,'fill','MarkerEdgeColor','k');

title('KMEANS Clustering')

view(-20,5)



Two different clustering techniques were attempted here. Both of these clustering algorithms create clusters that "leak" into different "layers" of the spiral. In other words, some points have been incorrectly clustered. For example, in the second plot, the blue cluster contains points that appear in a different layer.

To instruct a clustering algorithm to give preference to its neighboring points on the same "layer" rather than those that are located nearby but on a different "layers", we can provide a nearest neighbor connectivity matrix as a data matrix to clustdata. Since the points on the same layer have more common neighbors they will be grouped together

Use Nearest Neighbors to Compute the Linkage

% Compute 40 nearest neighbors

[idx,Dist]=knnsearch(X,X,'k',40);

% Create the adjacency matrix for linkage

D = zeros(size(X,1));

for ii = 1:length(X)

D(ii,idx(ii,2:end)) = 1;

end

% Cluster the data with structure defined by neighbors

cLinks = linkage(D, 'ward');

c = cluster(cLinks, 'maxclust', 6);

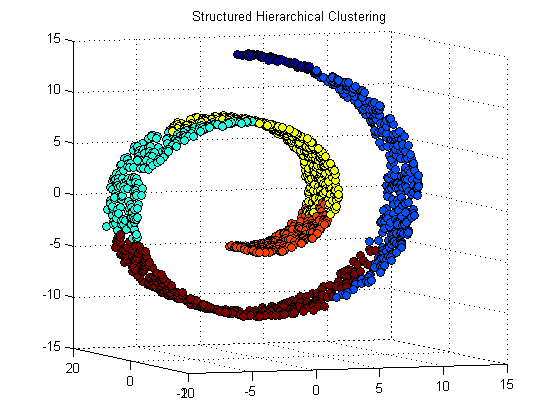
% Visualize

figure(2)

scatter3(X(:,1),X(:,2),X(:,3),[],c,'fill','MarkerEdgeColor','k');

title('Structured Hierarchical Clustering')

view(-20,5)



Dataset, Reference, and Licensing

This demonstration uses the Swiss Roll dataset from the following reference:

S. Marsland, Machine Learning: An Algorithmic Perpsective, Chapter 10, 2009.

Example from [scikit-learn.org](http://scikit-learn.org/stable/auto_examples/cluster/plot_ward_structured_vs_unstructured.html#example-cluster-plot-ward-structured-vs-unstructured-py) License: [BSD clause](https://www.mathworks.com/content/dam/mathworks/mathworks-dot-com/products/demos/machine-learning/swiss_roll/scikit-learn_bsd_license_copy.txt).