Cluster Genes Using K-Means and Self-Organizing Maps

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

This example demonstrates two ways to look for patterns in gene expression profiles by examining gene expression data from yeast experiencing a metabolic shift from fermentation to respiration.

This demonstration uses data and functions from the [Bioinformatics Toolbox™](https://www.mathworks.com/products/bioinfo.html).

Load Data

load filteredyeastdata

rng('default') % For reproducibility

The available information for this example consists of the yeast genes and their expression levels in yeastvalues at different times.

Clustering Genes Using a Hierarchical Cluster Tree

clusters = clusterdata(yeastvalues,'maxclust',16,'distance','correlation','linkage','average');

figure(1)

for c = 1:16

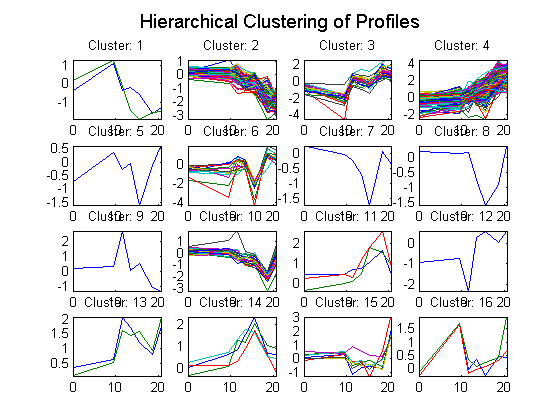
subplot(4,4,c);

plot(times,yeastvalues((clusters == c),:)');

axis tight

end

suptitle('Hierarchical Clustering of Profiles');



Use Principal Component Analysis and K-Means to Cluster in Lower Dimensions

figure(2)

[~,score,~,~,explainedVar] = pca(yeastvalues);

bar(explainedVar)

title('Explained Variance: More than 90% explained by first two principal components')

ylabel('PC')

% Retain first two principal components

yeastPC = score(:,1:2);

figure(3)

[clusters, centroid] = kmeans(yeastPC,6);

gscatter(yeastPC(:,1),yeastPC(:,2),clusters)

legend('location','southeast')

xlabel('First Principal Component');

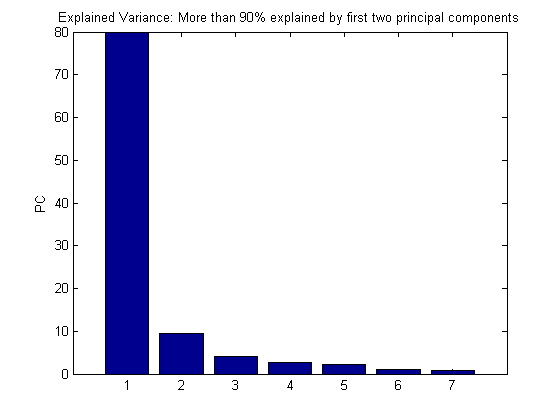
ylabel('Second Principal Component');

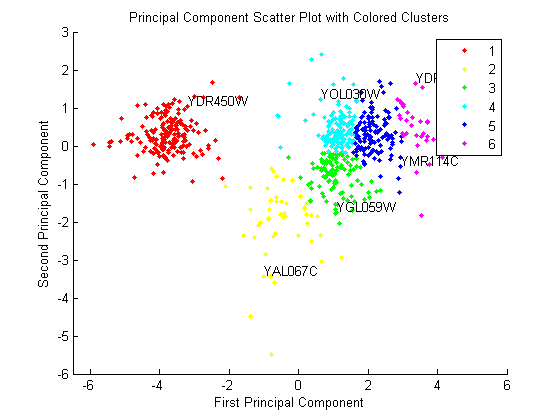
title('Principal Component Scatter Plot with Colored Clusters');

% Label one gene in each cluster

[~, r] = unique(clusters);

text(yeastPC(r,1),yeastPC(r,2),genes(r),'FontSize',11);





Use Principal Component Analysis and Self-Organizing Maps to Cluster in Lower Dimensions

This section uses the self-organizing maps functionality from [Neural Network Toolbox™.](https://www.mathworks.com/products/neural-network.html)

net = newsom(yeastPC',[4 4]);

net = train(net,yeastPC');

distances = dist(yeastPC,net.IW{1}');

[d,center] = min(distances,[],2);

% center gives the cluster index

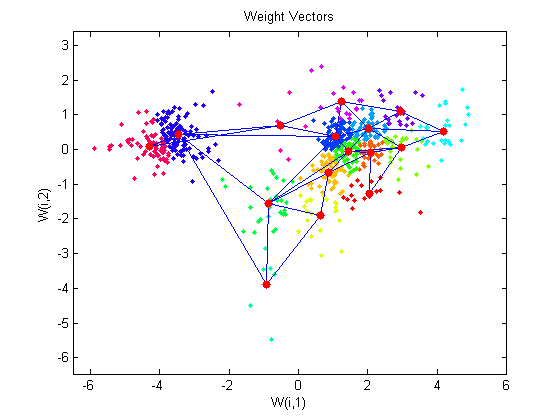
figure

gscatter(yeastPC(:,1),yeastPC(:,2),center); legend off;

hold on

plotsom(net.iw{1,1},net.layers{1}.distances);

hold off



This example explores two different approaches to cluster genes. For a more comprehensive demonstration, please visit our [Gene Expression Profile Analysis](https://www.mathworks.com/help/bioinfo/examples/gene-expression-profile-analysis.html) documentation.