Clustering Continued

By Jake Sauter

Kohonen Self Organizing Feature Maps (SOFM)

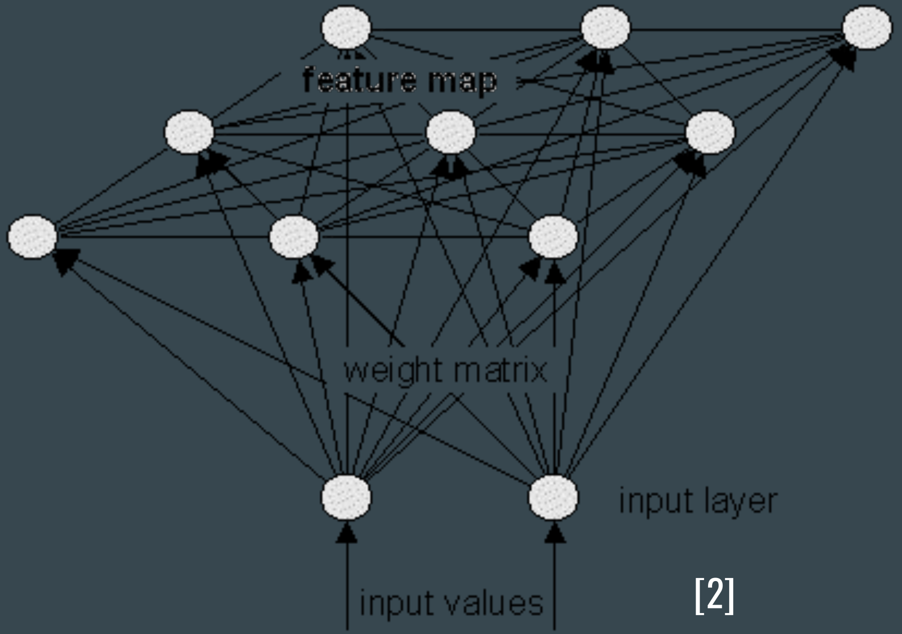
- SOFM is a type of clustering
- Novelty in that it the relationship between the clustered patterns actually contains information about the relationships and reciprocal positions of the patterns in the input space
- SOFM is designed to plot similar patterns next to one another, creating a **feature map**
- A feature map has the property that the distances and relationships measured on the feature map are proportional ot distances and relationships between patterns according to the similarity metric chosen

SOFMs

- The SOFM is actually a **neural network** technique
- Neural networks are graphs connecting simple processing nodes (**neurons**), with each connection possessing a specific connection strength (**weight)**
- Common SOFM architectures implement one dimensional and two dimensional networks

SOFMs

- The **input layer** of the network represents the dimensionality of the input space, and the **output layer** of the network represents the dimensionality of the desired output space



[2]

SOFM Training

- In order to form the output space, the SOFM must be **trained** with respect to a **training rule**
- The **training rule** in combination with a **learning rate** will inform us on how to update the weights of the network as training occurs

SOFM Training

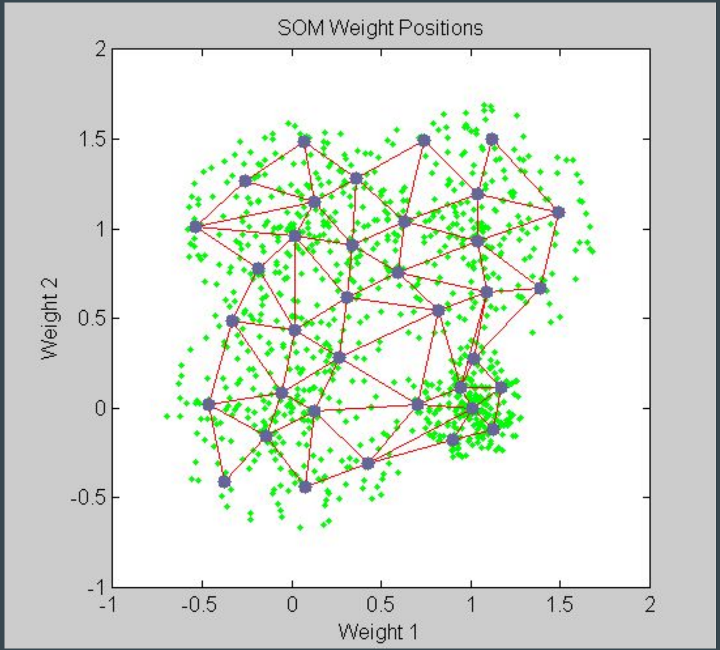
- The process of training occurs the following way:
  - An input is presented to the input layer of the neural network
  - Each neuron calculates the distance from the input to its weight vector
  - The neuron that possesses the smallest distance will be updated such that the weights are made more similar to the input profile
    - Depending on the training rule, weight of surrounding neurons may be updated as well
  - Each input is presented to the network in this way
  - After each input is presented, the learning rate and/or neighborhood distance is decreased and this process is repeated

SOFM Output

- The SOFM provides three benefits
    - Each neuron of the SOFM will represent the set of common features extracted from the input patterns by the neuron
    - The SOFM yields a set of clusters, with all inputs activating the same unit being clustered together
    - The relationship between the neurons activated by specific genes will be closely related to the relationship between the genes

SOFM Use

- SOFMs can be used as a clustering or visualization tool
- One could simply plot the clusters generated from the SOFM
- The prototypes (weights of the units) could be plotted in the input space, with the topology of the network showing the links between the units



[3]

K Means Results

- The K-Means clustering algorithm was implemented from scratch to allow for different distance metrics to be used
- A function to produce meaningful output for supervised learning sets, with any amount of clusters, was crafted
    - This technique could be applied to determine how well classification can even be implemented

K Means Results

Custom Implementation

Euc.

```
between_ss / total_ss =  62.23644 %

Cluster Results:
     ALL        AML
[1,]   1 0.09090909
[2,]   0 0.90909091
```

Man.

```
between_ss / total_ss =  60.78344 %

Cluster Results:
           ALL AML
[1,] 0.4074074   1
[2,] 0.5925926   0
```

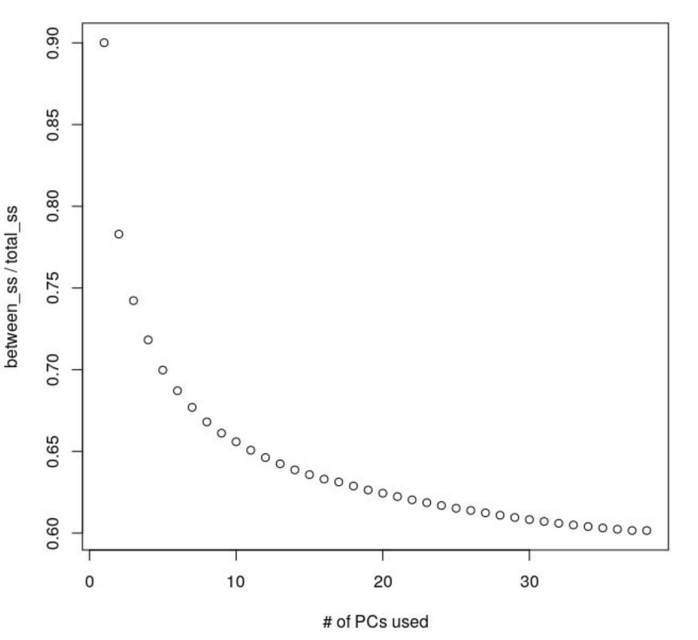
Library Function

Euc.

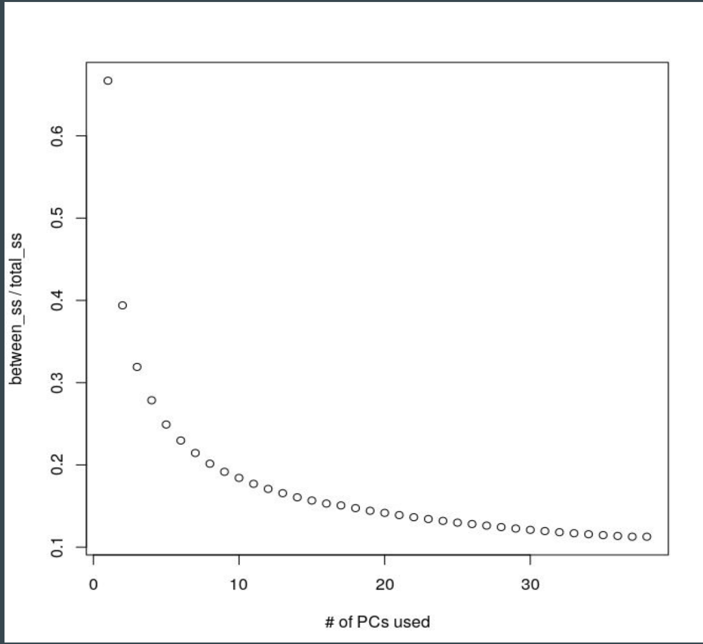
```
between_ss / total_ss =  11.30524 %

Cluster Results:
           ALL       AML
[1,] 0.8518519 0.2727273
[2,] 0.1481481 0.7272727
```

K Means of Principle Components (Euc)

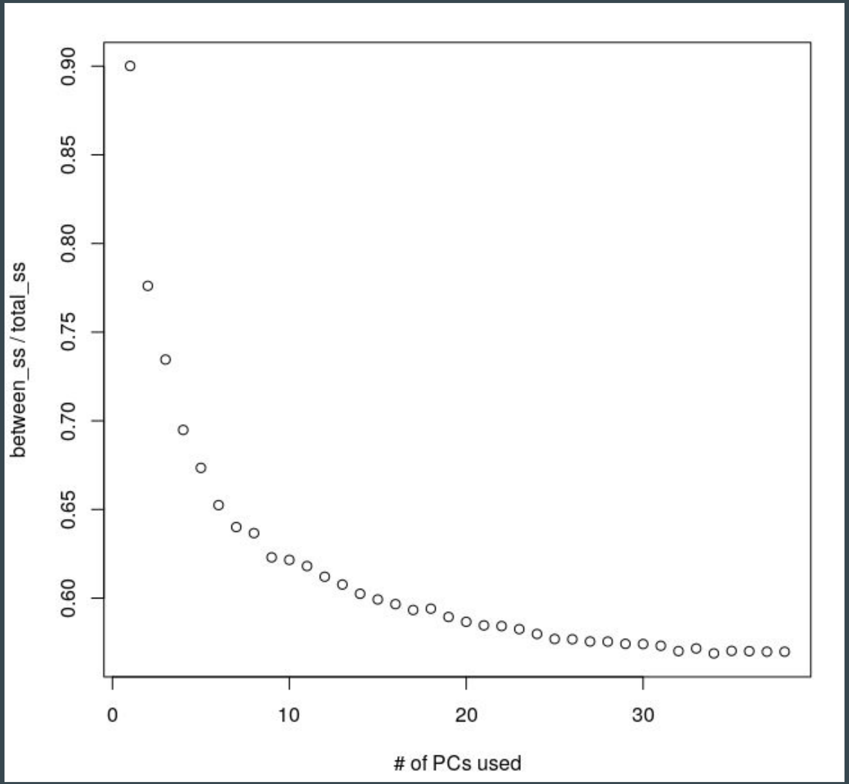
Custom Implementation



Library Function



K Means of Principle Components (Man)



K Medoids Results

Euclidean

Cluster Results:

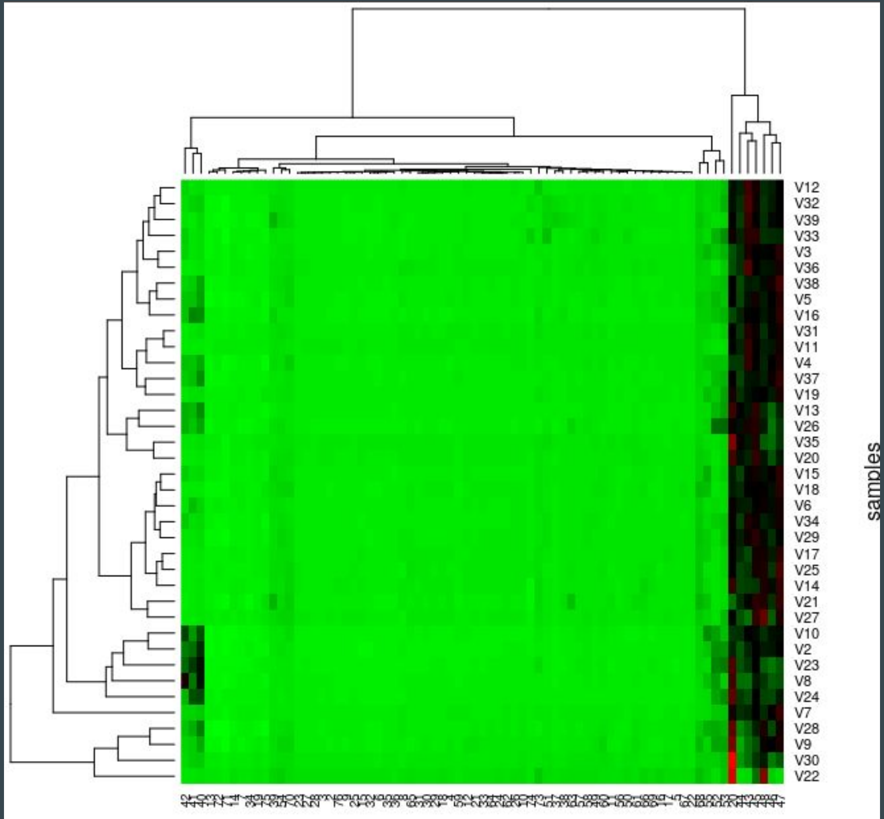
	ALL	AML
[1,]	0.8888889	0.1818182
[2,]	0.1111111	0.8181818

Manhattan

Cluster Results:

	ALL	AML
[1,]	0.92592593	0.1818182
[2,]	0.07407407	0.8181818

Hierarchical Clustering



Rerenerences

[1] Drăghici Sorin. Statistics and Data Analysis for Microarrays: Using R and Bioconductor. Chapman and Hall, 2012.

[2] “Kohonen Feature Map.” Kohonen Feature Map - Neural Networks with Java, www.nnwj.de/kohonen-feature-map.html.

[3] “Cluster with Self-Organizing Map Neural Network.” Cluster with Self-Organizing Map Neural Network - MATLAB & Simulink, www.mathworks.com/help/deeplearning/ug/cluster-with-self-organizing-map-neural-ne
twork.html;jsessionid=38782e43df7fa2b45736dd152319.