Self-Organizing Maps

Kyle Thayer
Organizing Marbles
Self-Organizing Maps
Algorithm (Definitions)

• Distance in Data Space
• Best-Matching Unit (BMU)
  – Node that is closest to a given input vector.
• Neighborhood
Algorithm (Initialization)

• Random
  – From data range
  – From data set

• Linear
Original Iterative SOM

- One random data point (x) per iteration
  1) Node n = BMU(x)
  2) Shift weights of n and neighborhood of n toward weights of x.

- Neighborhood size and shift amount decrease over time.
Properties and issues

• Preserves topology
• Data evenly distributed.
• Exceptions:
  – Edges pulled in
  – Nodes between clusters in the data (low density)
  – Data that can’t map to 2D space.
Batch SOM

- Many random points per iteration
  1) Place the points on their BMU node in the SOM
  2) Every node in the SOM’s new weight is the average of all data points that landed in its neighborhood.

- Neighborhood shrinks over time.
- Note: Neighborhood of 0 is k-means.
Feature Maps

• U-Matrix
• Hit Histogram (Density Map)
SOM Accuracy

- Avg. distance from data point to BMU
- Topology preservation
Clustering

- U-Matrix
- K-Means
- Hit Histogram
- Visualizing Clusters
Extensions

- Different node arrangements
- Hierarchical SOM
- Dynamic node creation
Applications

Fig. 3.2. In this exemplary application, each processing element in the hexagonal grid holds a model of a short-time spectrum of natural speech (Finnish). Notice that neighboring models are mutually similar.
More info

• My program is at www.ifcsoft.com
• Book: Self-Organizing Maps by T. Kohonen
• Other sites:
  – http://davis.wpi.edu/~matt/courses/soms/