Crime & Violence Statistics of the United Sates
As Defined by the Self Organizing Map (SOM)

Kevin Brown
Data Mining Methods
-Webster University-
2006

Overview

- Introduction to SOM
- Explanation of Euclidian Distance
- Dataset Overview
- Describe Conducted Research
- Explain Hypothesized Results
- Explain Actual Results
- Summary
- View Sources
- Q and A

Self-Organizing Map (SOM)

- Is a Subtype of Artificial Neural Networks
- Developed by Teuvo Kohonen at Helsinki University of Technology
- Useful for visualizing low-dimensional displays of high-dimensional data
- The SOM is mainly used for dimensionality reduction

Euclidean Distance

- When a training sample is given to the network, its Euclidean distance to all weight vectors is computed

$$\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \cdots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2}$$

Dataset Overview

- 50 U.S. States Including D.C. (51 total)
- Both Violent and Property Crimes included
- Income and Population also factors
- Fairly Large and complex dataset

Conducted Research

- Used SOM executables derived from Teuvo Kohonen model
- Created data and batch files to match SOM executables necessary format
- Depicted results in aesthetically friendly map display
- Formed a method to efficiently evaluate map outputs and clusters
- Depicted observations in a similar map display, labeling all clusters and patterns
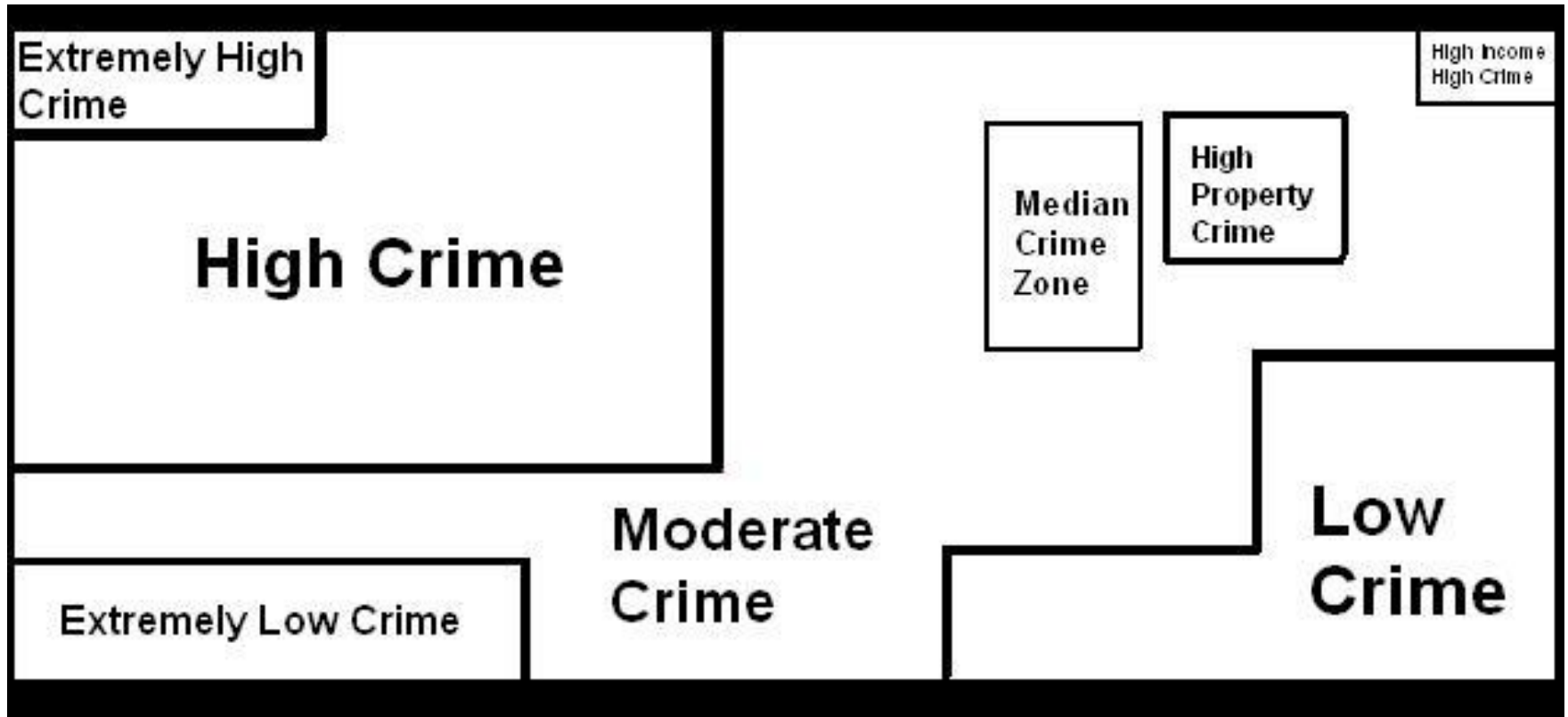
Hypothesized Results

- Groups would map similar to geographic topography of the US

- Established Regions of the U.S. (i.e. mid-west, west coast, east coast, southwest) would map together

- SOM would be able to map the complex data set

Actual Results
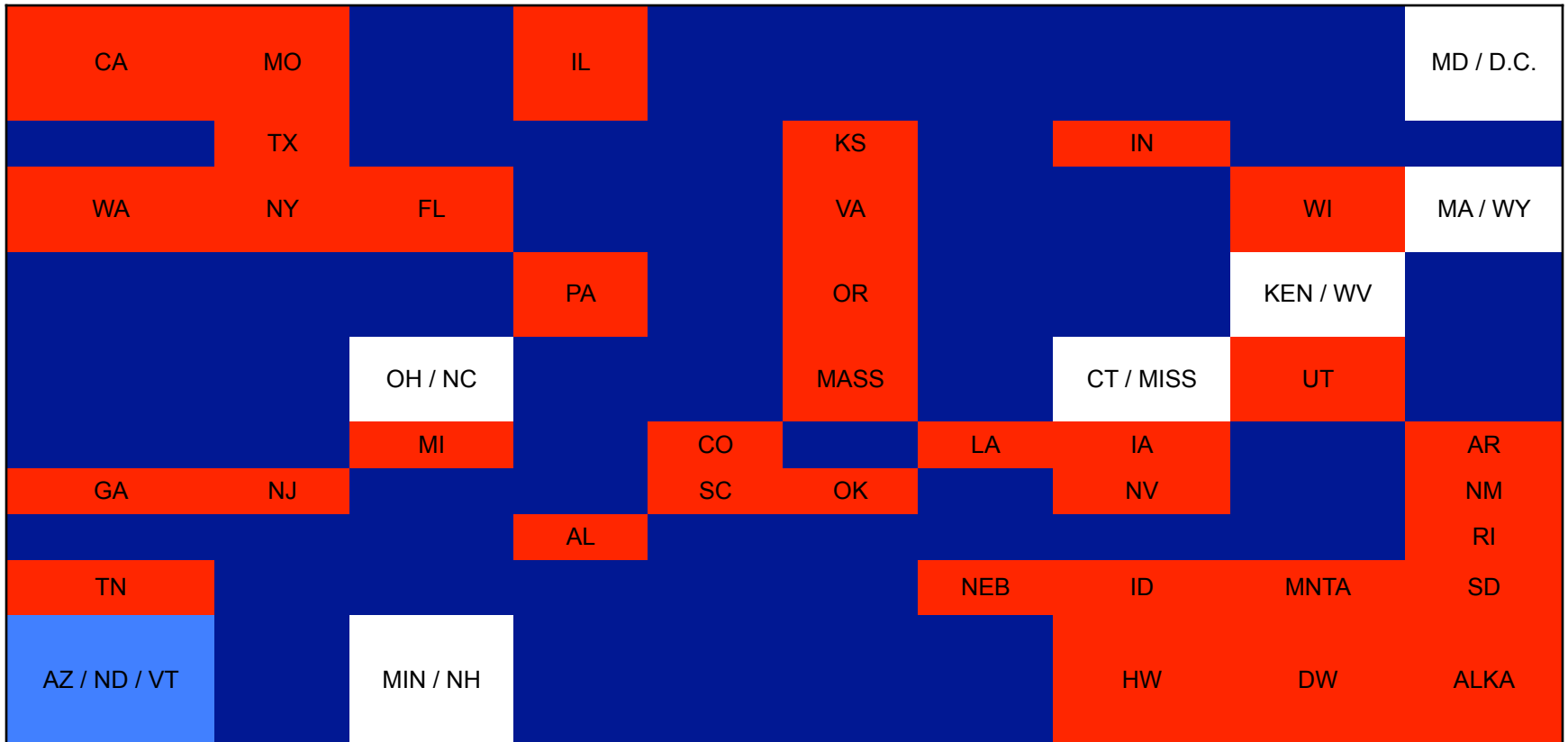
Results after Analysis

Analysis Technique

- Maximum, minimum, and mean value calculated for each attribute

- Grouped data sets together as dipected on map for easy comparison

- Noted observations of similarities and differences within each cluster mapped

Conclusion

- SOM effectively handled complex data by displaying a simple 2D map
- Results differed from hypothesized results
- SOM mapped clusters in a coherent manner

Summary

- SOM is a artificial neural network used to visually create a low-level display of high-level data

- SOM successfully displayed my high-level data set in an easy to understand manner

- SOM using training algorithms and Euclidian distance to map it's nodes

- SOM is an efficient means to easily display clusters.

Sources

- **1 Self Organizing MAP SOM (tools):**
  mercury.webster.edu/aleshunas/
- **2 Crime and Violence Data Sets:**
  http://bjsdata.ojp.usdoj.gov/dataonline/Search/Crime/State/statebystatelist.cfm
- **3 Self Organizing Map SOM (information):**
  http://en.wikipedia.org/wiki/Self-organizing_map
- **4 Population and Income Data Sets:**
  http://quickfacts.census.gov/qfd/index.html
- **5 Euclidean Distance:**
  http://en.wikipedia.org/wiki/Euclidean_distance