Self-Organizing map and anorectic data set

By: Theresa Helm

Overview

- Problem Description
- Anorectic Data Set
- Self-Organizing Map
- Implementation
- Results
- Conclusion
- Summary

Problem

- Using a mathematical algorithm to support medical research

Anorectic Data Set

- Sanitized Data Set

	Name	Label
1	weight	Body Weight
2	mens	Menstruation
3	fast	Restriction of food intake (fasting)
4	binge	Binge eating
5	vomit	Vomiting
6	purge	Purging
7	hyper	Hyperactivity
8	fami	Family relations
9	eman	Emancipation from family
10	frie	Friends
11	school	School/employment record
12	satt	Sexual attitude
13	sbeh	Sexual behavior
14	mood	Mental state (mood)
15	preo	Preoccupation with food and weight
16	body	Body perception
17	time	Time of interview
18	diag	Patient Diagnosis
19	tidi	Time/diagnosis interaction
20	number	Patient Number
21	diag2	Diagnosis
22	time2	

Self-Organizing Map (SOM)

- Artificial neural network
- Unsupervised learning
- Used for visualizing low-dimensional views of high-dimensional data

SOM continued

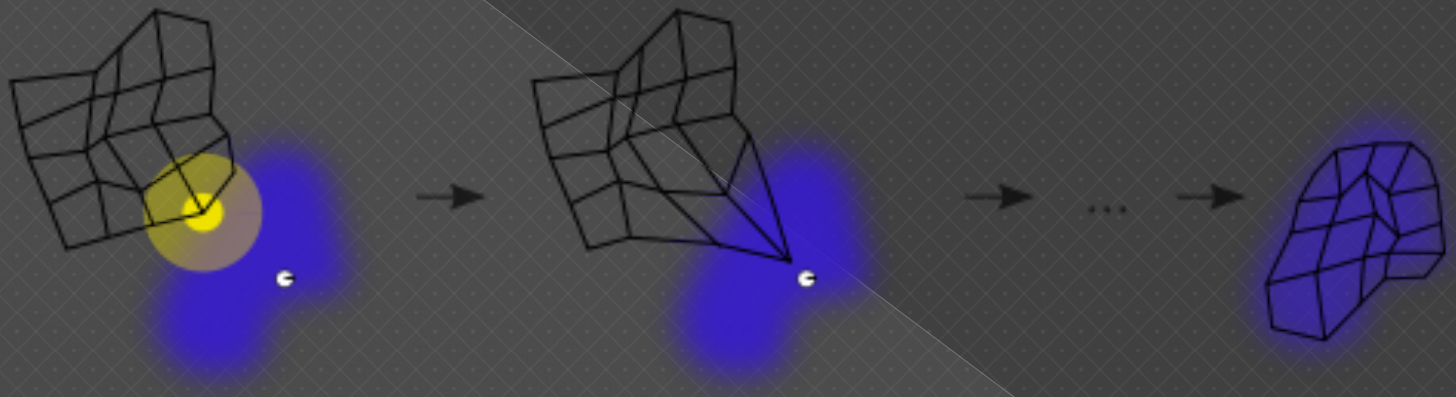
- Has two modes: training and mapping
- Consists of nodes

How SOM works

- Randomly initialized
- Receives the input vector
- Uses Euclidean Distance

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

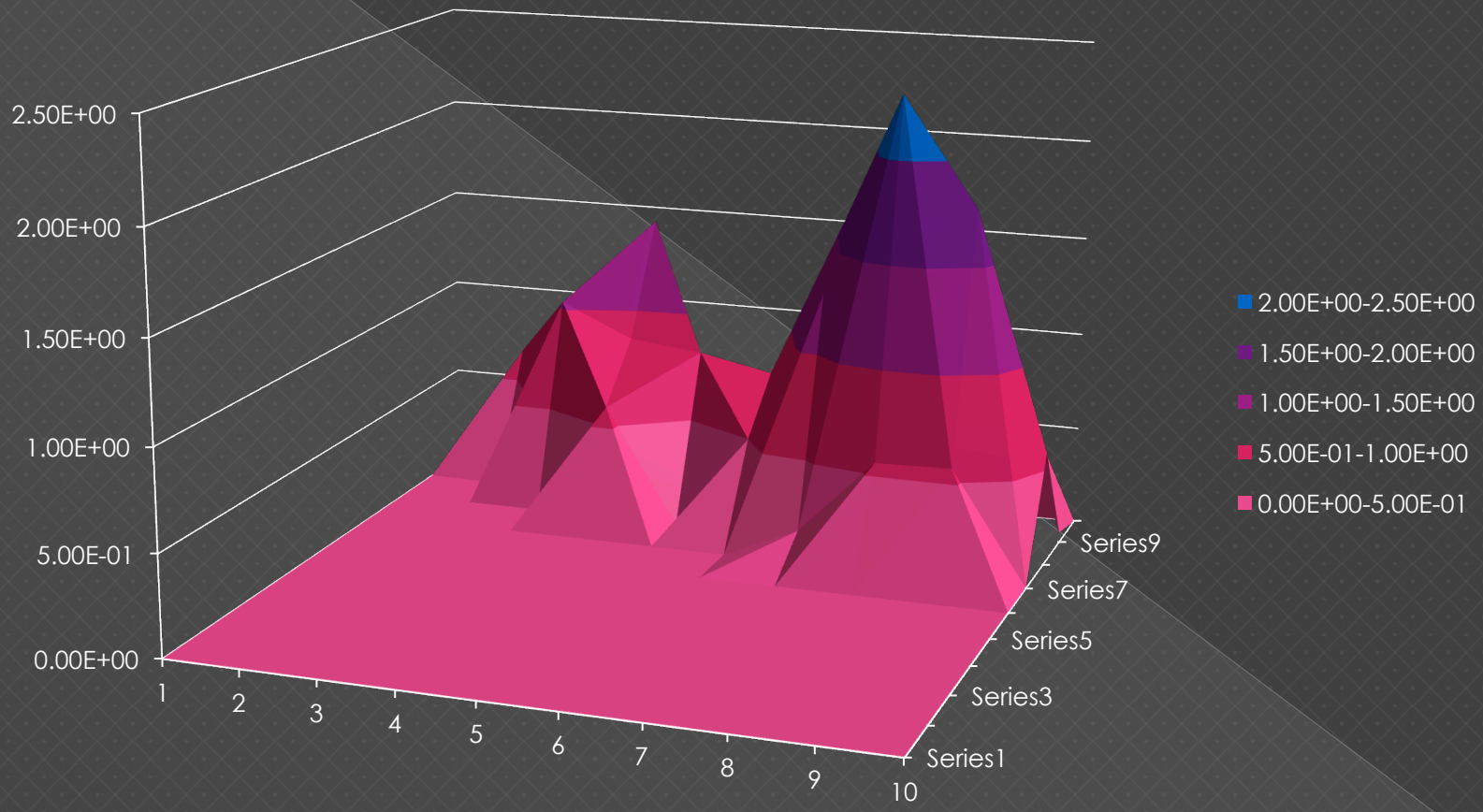
SOM Illustrated



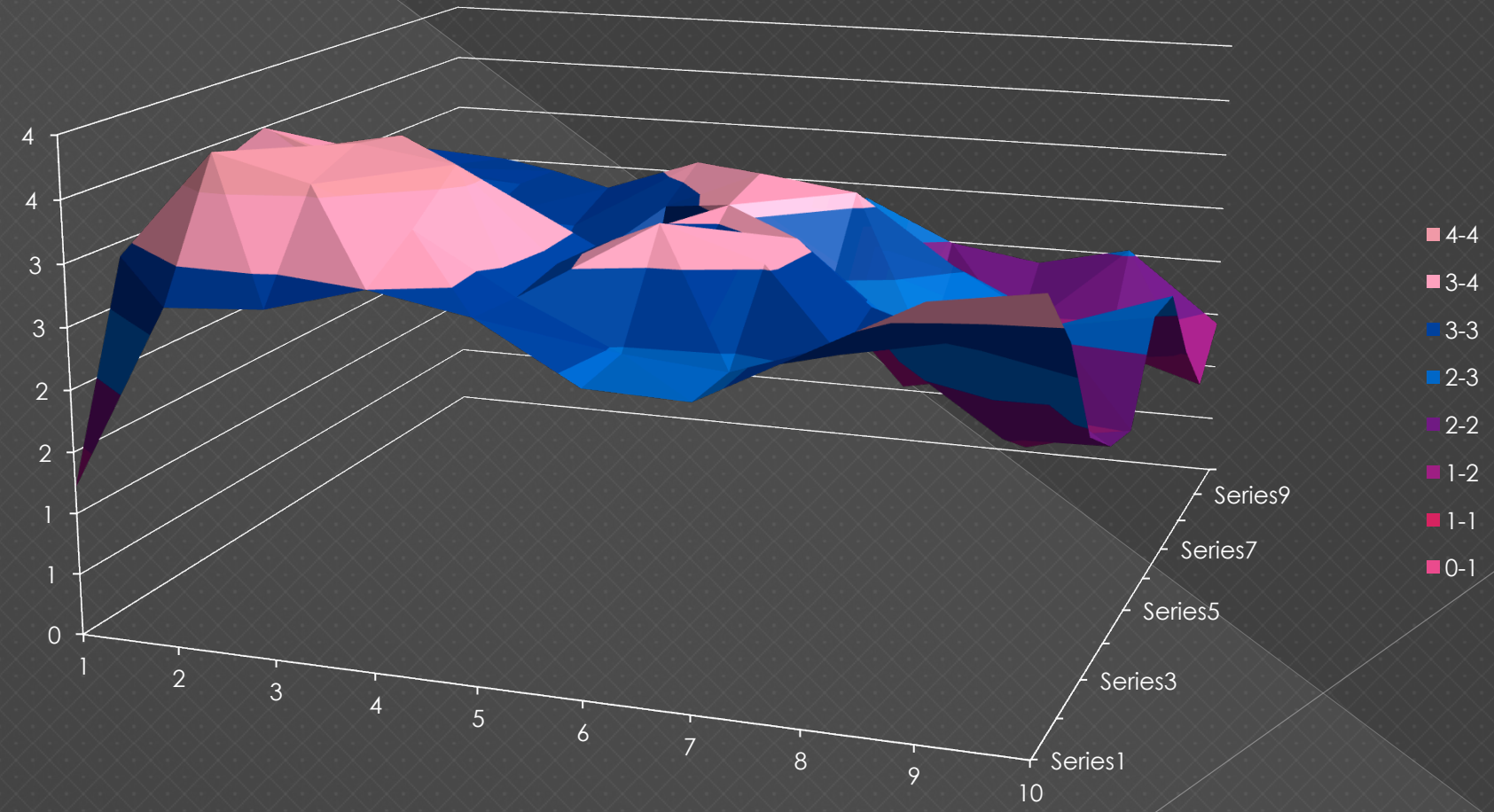
Implementation

- Ran the original anorectic data set with SOM
- Made surface graphs of each attribute
- Found that the original data set is noisy

Implementation



Implementation

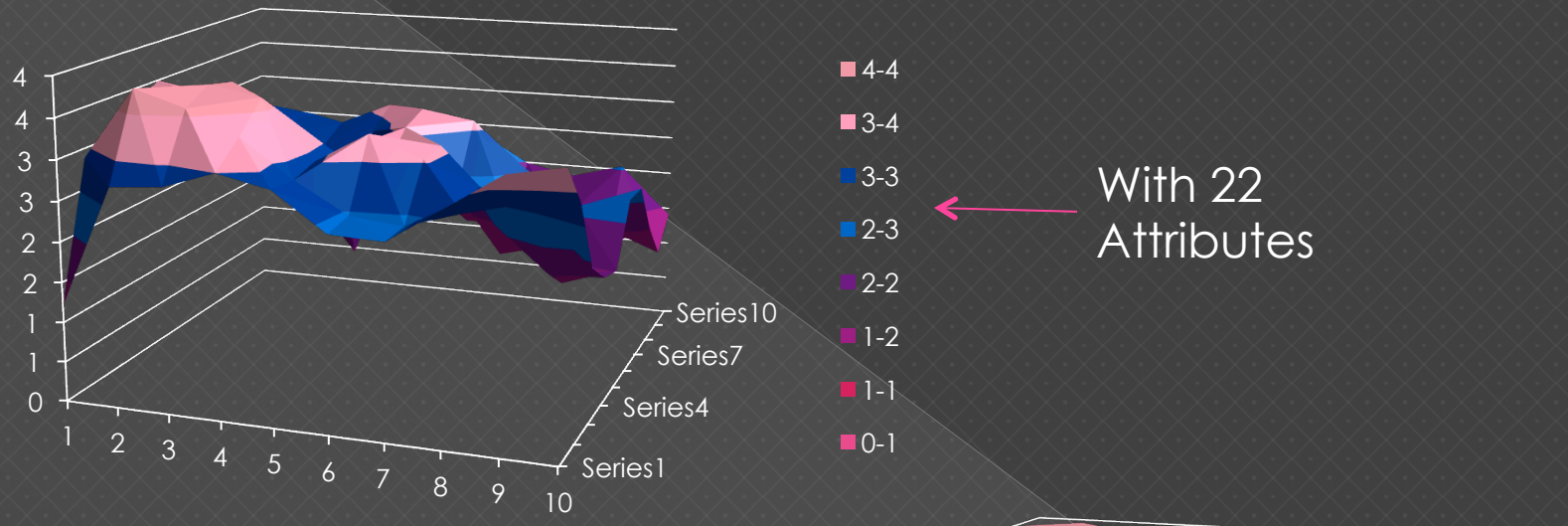


Implementation

- Took out the noisy data

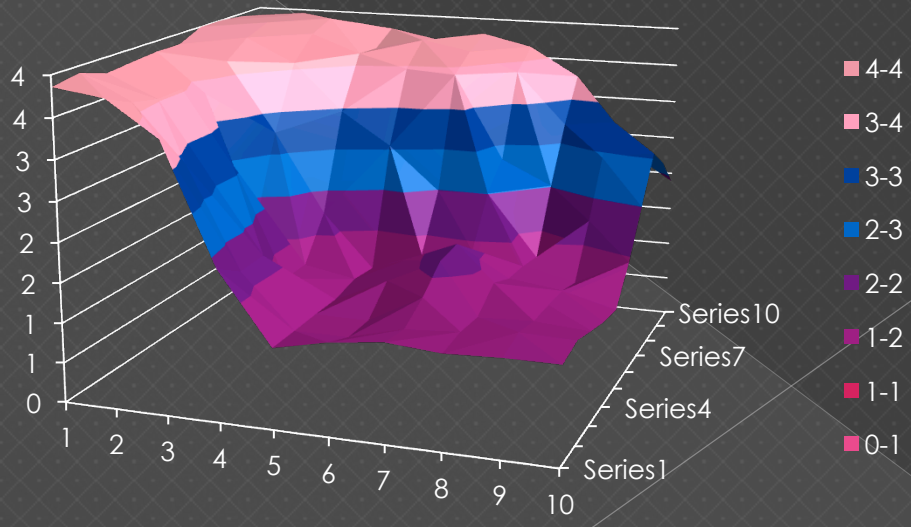
	Name	Label
1	weight	Body Weight
2	mens	Menstruation
3	fast	Restriction of food intake (fasting)
4	binge	Binge eating
5	vomit	Vomiting
6	purge	Purging
7	hyper	Hyperactivity
8	fami	Family relations
9	eman	Emancipation from family
10	frie	Friends
11	school	School/employment record
12	satt	Sexual attitude
13	sbeh	Sexual behavior
14	mood	Mental state (mood)
15	preo	Preoccupation with food and weight
16	body	Body perception
~~17~~	~~time~~	~~Time of interview~~
~~18~~	~~diag~~	~~Patient Diagnosis~~
~~19~~	~~tidi~~	~~Time/diagnosis interaction~~
~~20~~	~~number~~	~~Patient Number~~
~~21~~	~~diag2~~	~~Diagnosis~~
~~22~~	~~time2~~	

Results

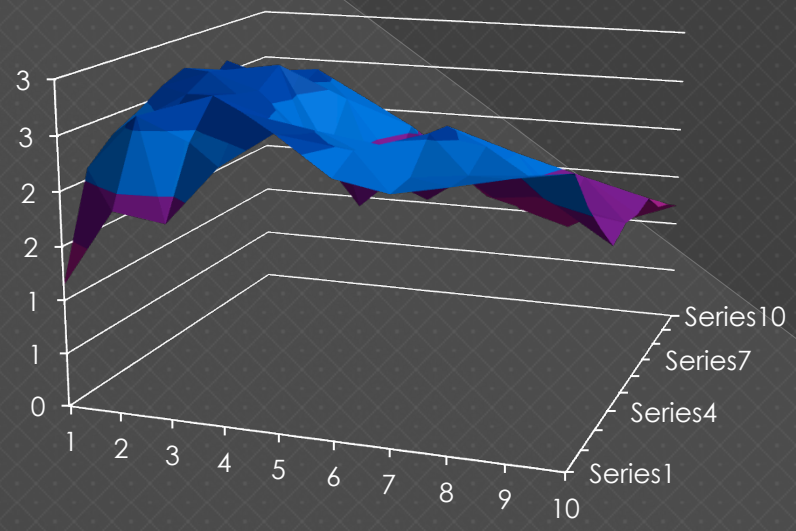


With 22 Attributes

With 16 attributes

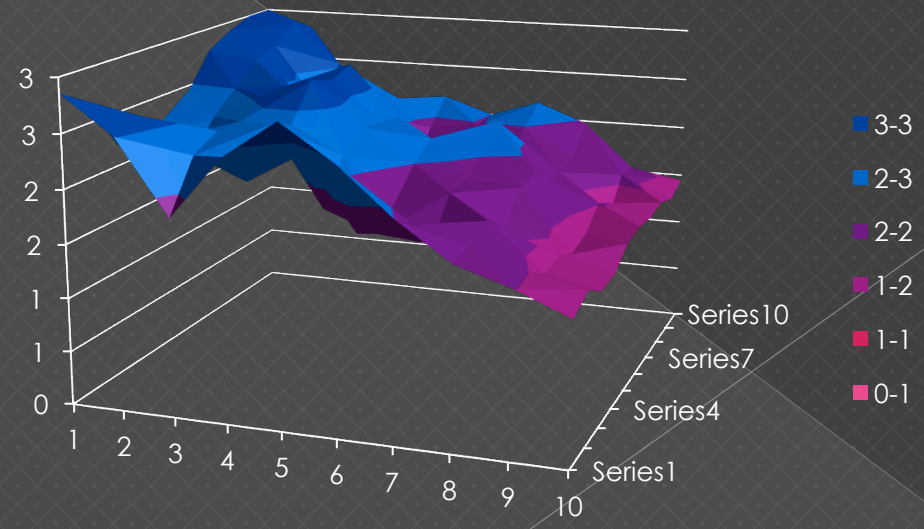


Results

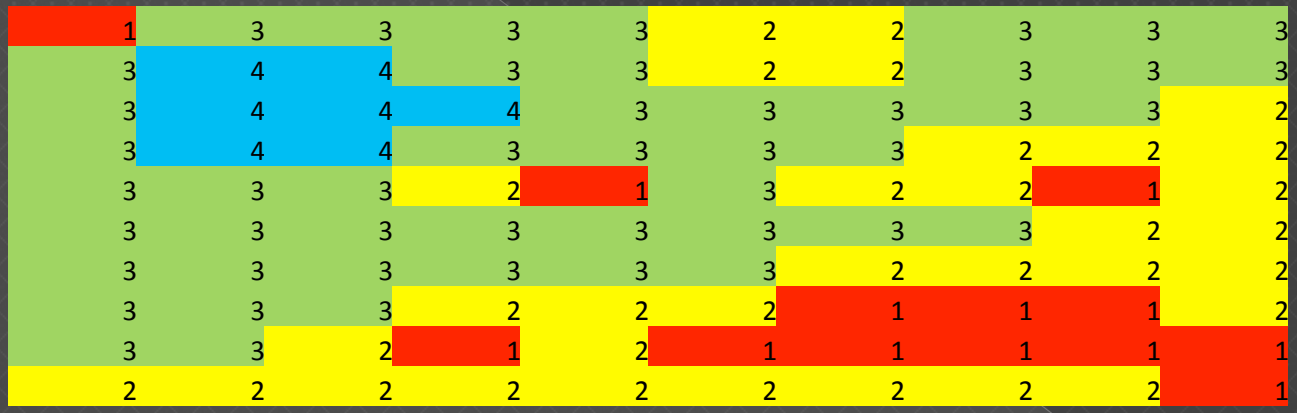


With 22 attributes

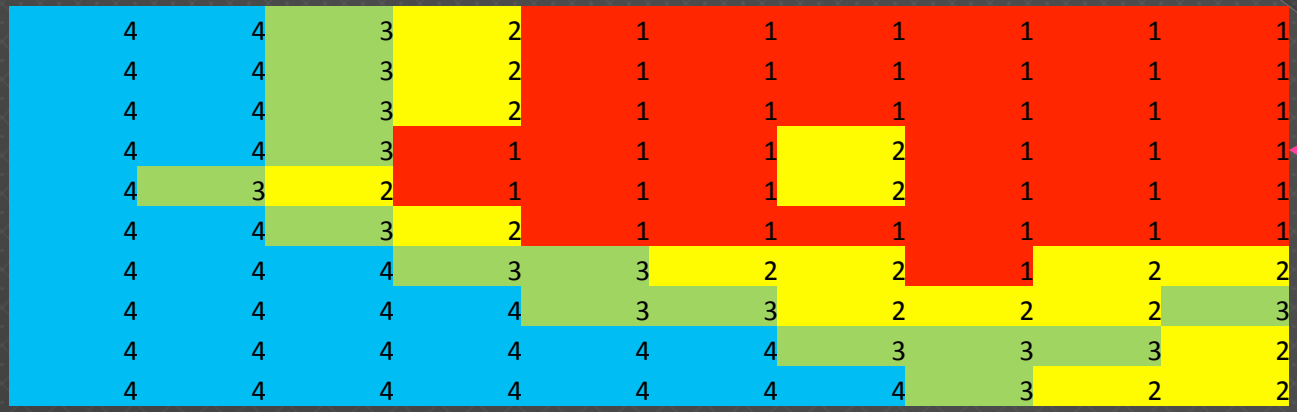
With 16 Attributes



The Map



With 22 attributes

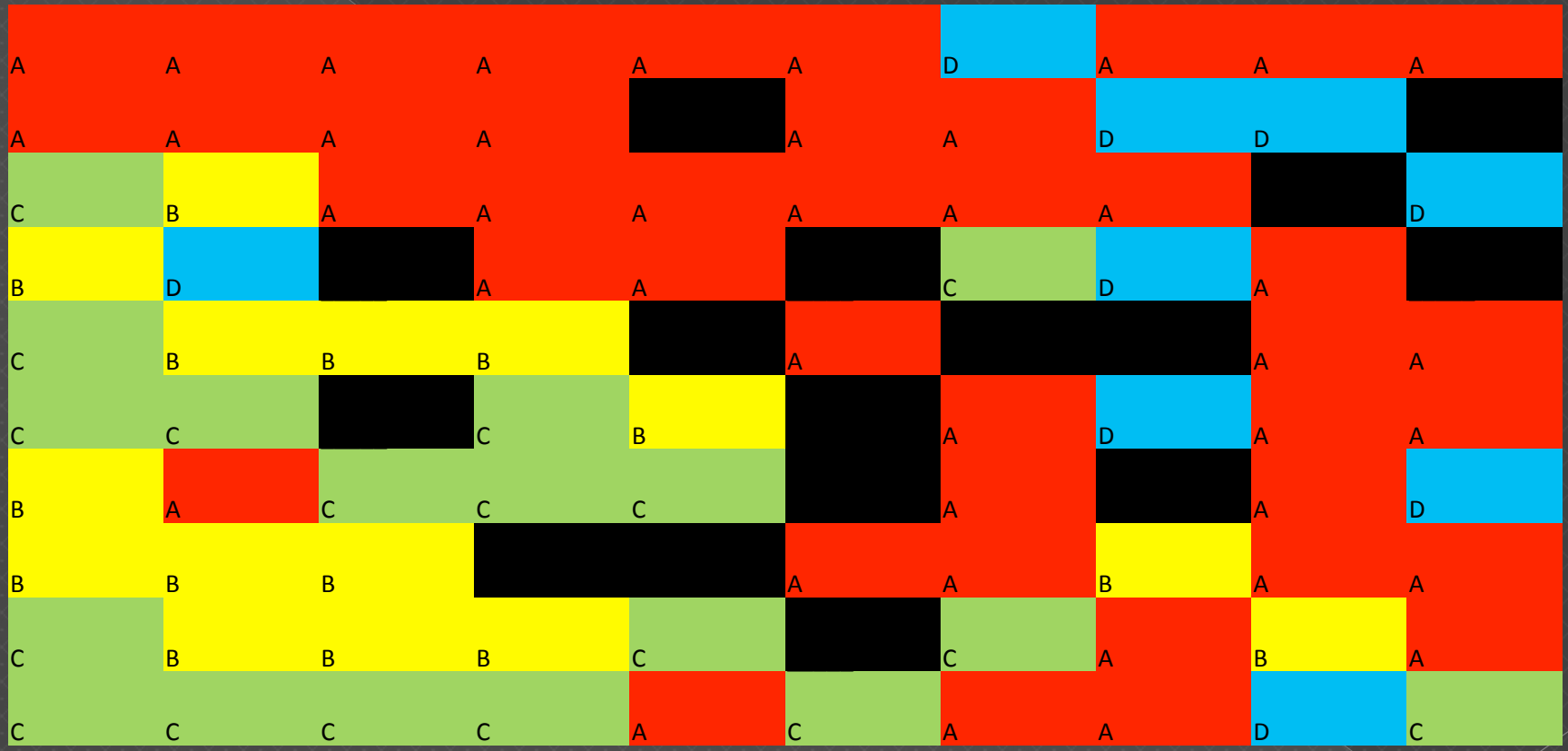


With 16 attributes

Labeling of the Map

- Anorectic data set has no labels
  - Anorectic
  - Border
  - Non-Anorectic
- Solution- Made my own labels

Labeling of the Map



A=1
B=2
C=3
D=4

Conclusions

- Not an accurate way to determine if a person is anorectic
- Use different attributes

Summary

- Problem Description
- Anorectic Data set
- Self-Organizing Map
- Implementation
- Results
- Conclusion

Questions

References

Artificial neural network. (2010, December 14). Retrieved December 14, 2010, from Wikipedia: http://en.wikipedia.org/wiki/Artificial_neural_network

MATH 3220 Data Mining Methods . (n.d.). Retrieved December 1, 2010, from John Aleshunas: http://mercury.webster.edu/aleshunas/MATH%203220/MATH%203220%20Home.htm

Self-organizing map. (2010, December 13). Retrieved December 14, 2010, from Wikipedia: http://en.wikipedia.org/wiki/Self-organizing_map

Unsupervised learning. (2010, December 9). Retrieved December 14, 2010, from Wikipedia: http://en.wikipedia.org/wiki/Unsupervised_learning