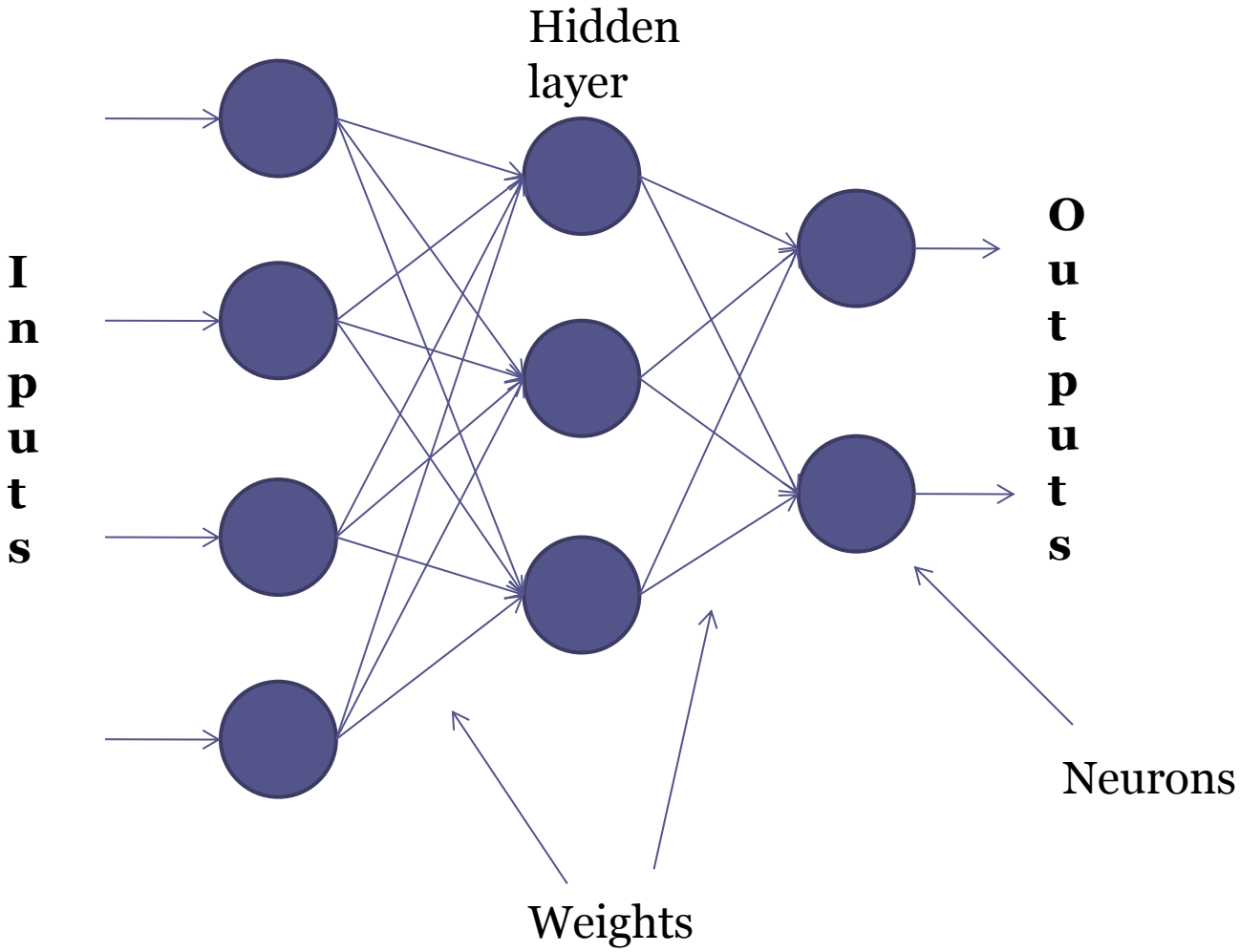
Using Artificial Intelligence and Machine Learning Techniques. Some Preliminary Ideas.

Presentation to CWiPP
1/8/2013
ICOSS
Mark Tomlinson

Artificial Intelligence Models

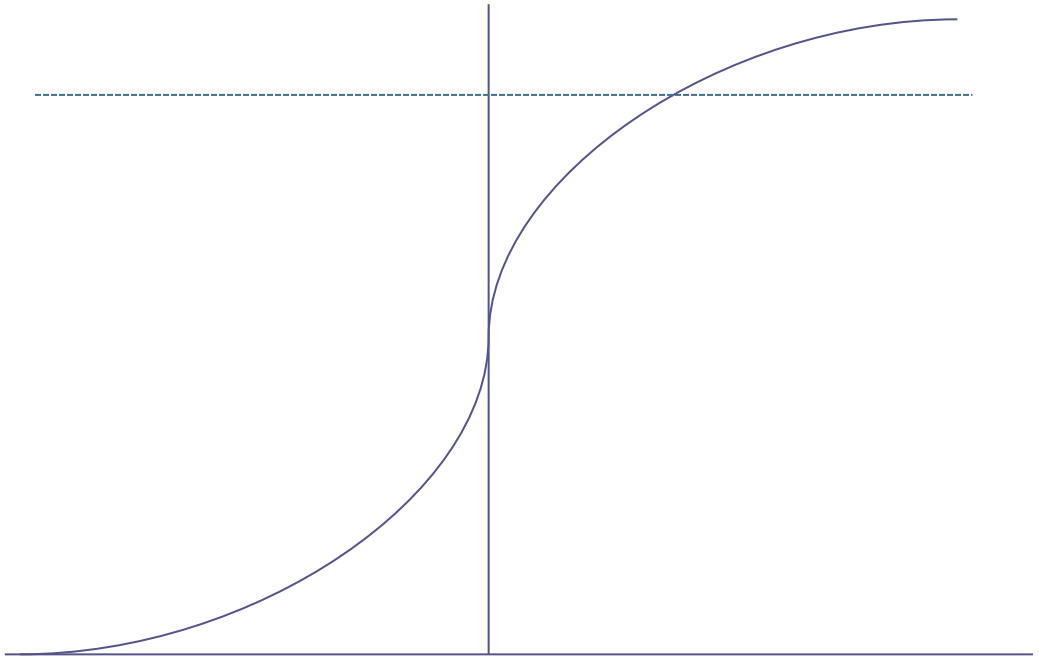
- Very experimental, but timely?
  - Big data
  - Blurring of boundaries between quant/qual
- Neural networks
  - Gøsta's Brain project – welfare regime theory
- Self Organising Maps (SOMs)
  - Initial experiments on child self-esteem
- Genetic algorithms

A simple neural network (perceptron)



Neuron fires when the sigmoid function reaches a threshold based on the sum of the weighted inputs

$y=1/(1+e^{-x})$



Neural networks to simulate Gøsta's brain

- A network with sufficient hidden nodes can be trained to be 100% accurate over a particular dataset
- Train network to predict Esping-Andersen typology perfectly in 1980 using macro policy data (Scruggs's decommodification data)
- Use data from different years to predict regimes using the same network
- Check for stability

Results using Scruggs decommodification data with a threshold of 0.9

5000 iterations on 1980 data

LIB=Liberal
CON=Conservative
SOC=Social Democrat

Country	1971	1980	1990	2000
AU	LIB	LIB	LIB	LIB
AT	CON	CON	CON	CON
BE	CON	CON	CON	CON
CA	LIB	LIB	LIB	LIB (0.872)
DK	SOC	SOC	SOC	SOC (0.864)
FI	SOC	SOC	SOC	SOC (0.640)
FR	CON	CON	CON (0.820)	LIB
DE	CON	CON	CON	CON
IR	LIB	LIB	LIB	LIB
IT	CON	CON	CON (0.891)	CON
JP	LIB	LIB	LIB	LIB
NL	SOC	SOC	SOC	SOC
NO	SOC	SOC	SOC	SOC
NZ	LIB	LIB	LIB	LIB
SW	SOC	SOC	SOC	SOC (0.806) LIB (0.450)
CH	CON (0.888)	LIB	LIB	LIB
UK	LIB	LIB	LIB	LIB
US	LIB	LIB	LIB	LIB

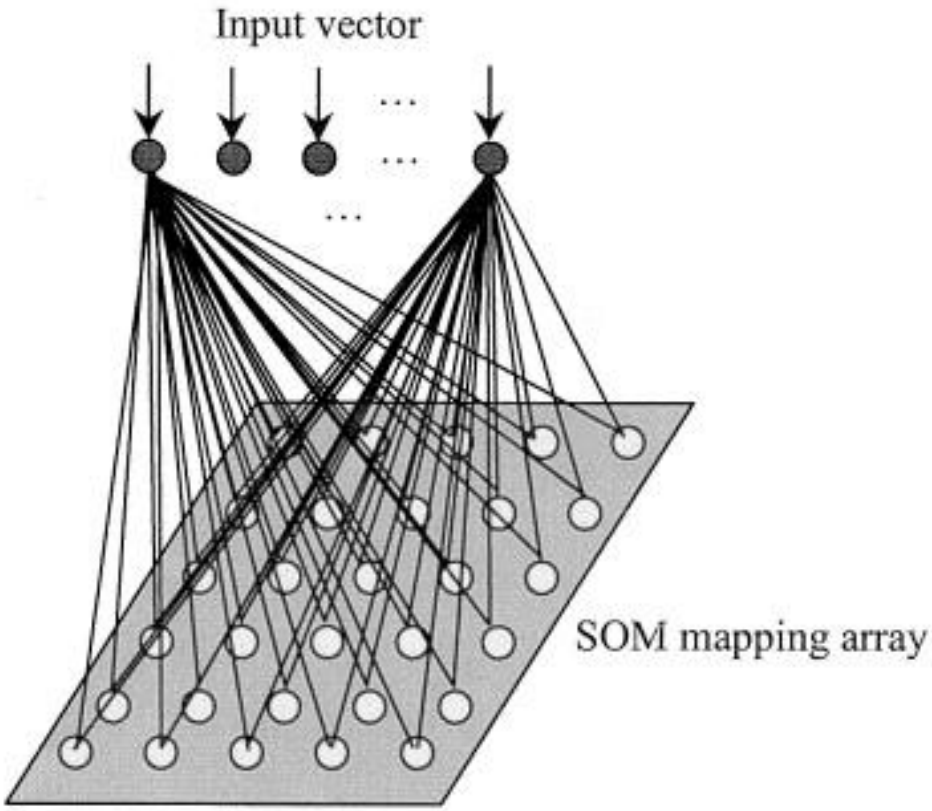
Other possible directions

- How to interpret weights in the network
  - This is very difficult
- Statistical properties of the weights
  - Develop pseudo-T statistics
- Significance of different weights and their interpretation
- Use of networks as a method of classifying social science data based on recategorising new data
  - Children at risk of future problems
- Use of neural networks as an alternative to regression

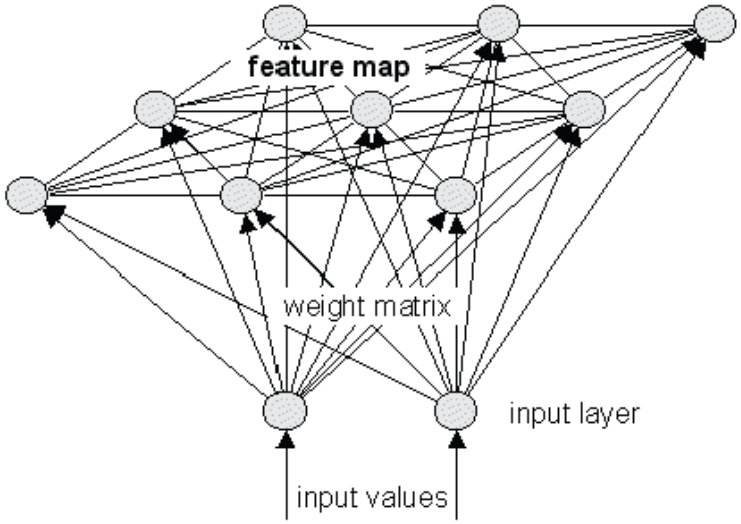
Self Organising Maps

- Kohonen network
- Not a neural network in the sense above
- Used to cluster large amounts of data
- Dependent variables can be incorporated that use the clustering to determine the outcomes simultaneously

A vector (set of data points) is mapped onto a lattice via a network of weights

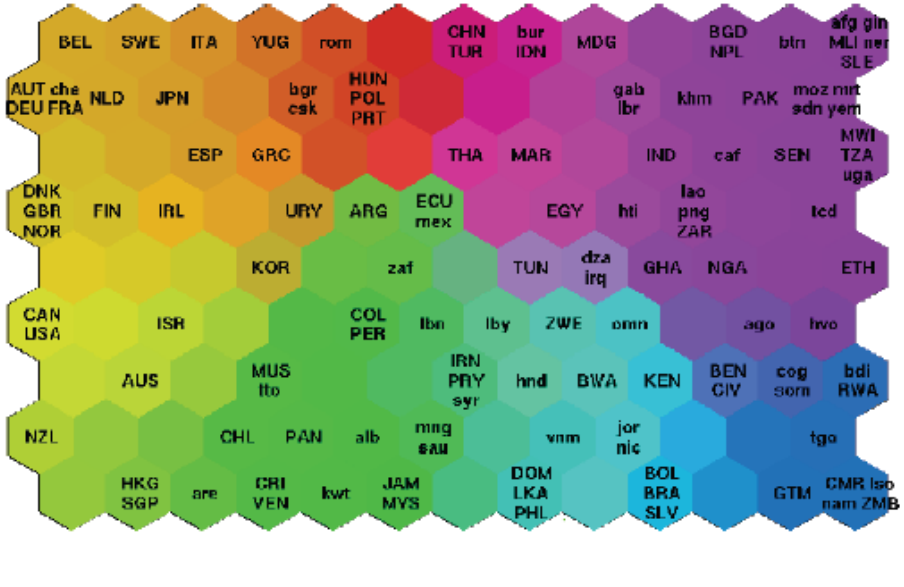


Already been used to map quality of life data



(39 indicators describing various quality of life factors, such as state of health, nutrition, educational services, etc)

Matlab can produce these charts

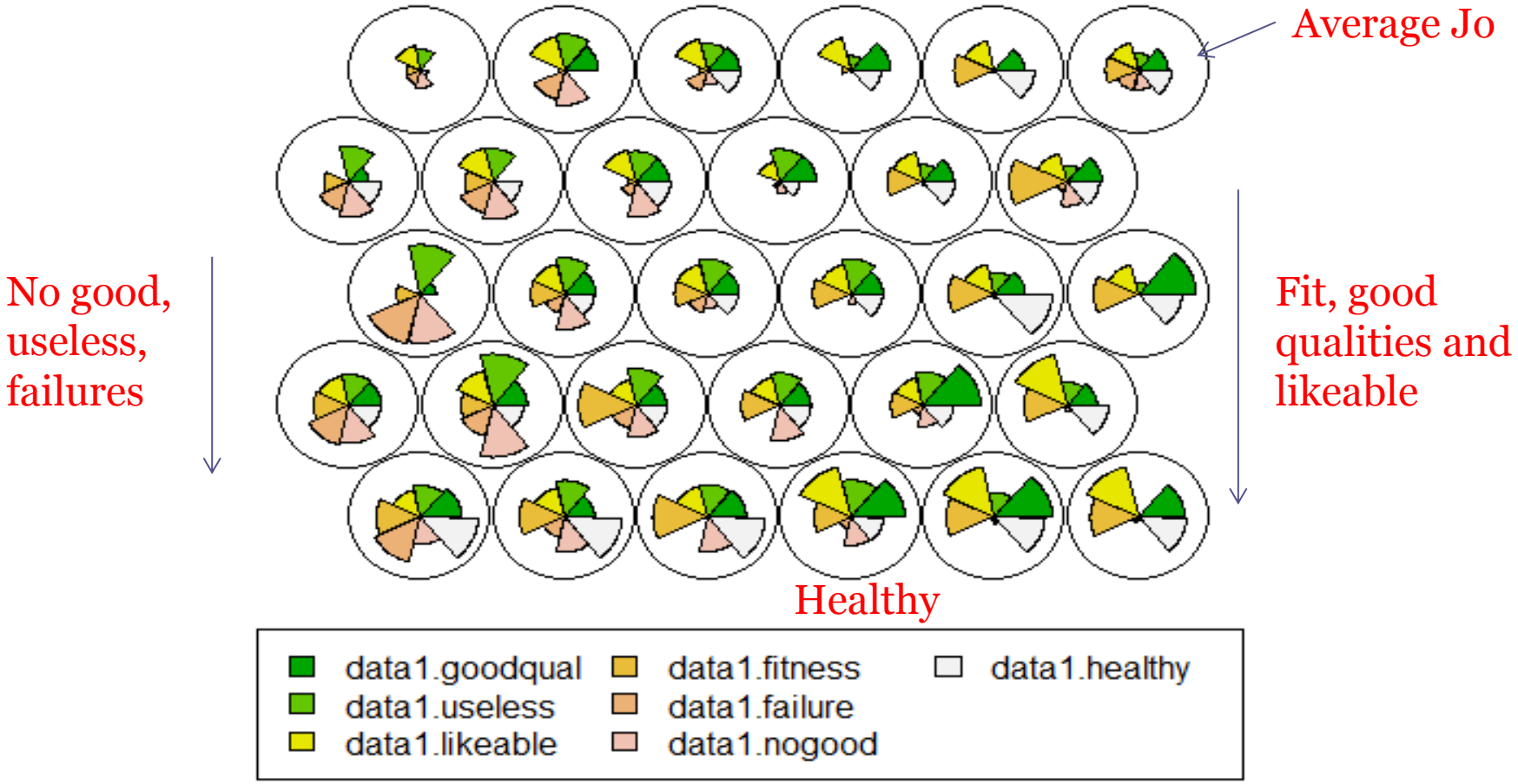


S. Kaski , T. Kohonen 1996

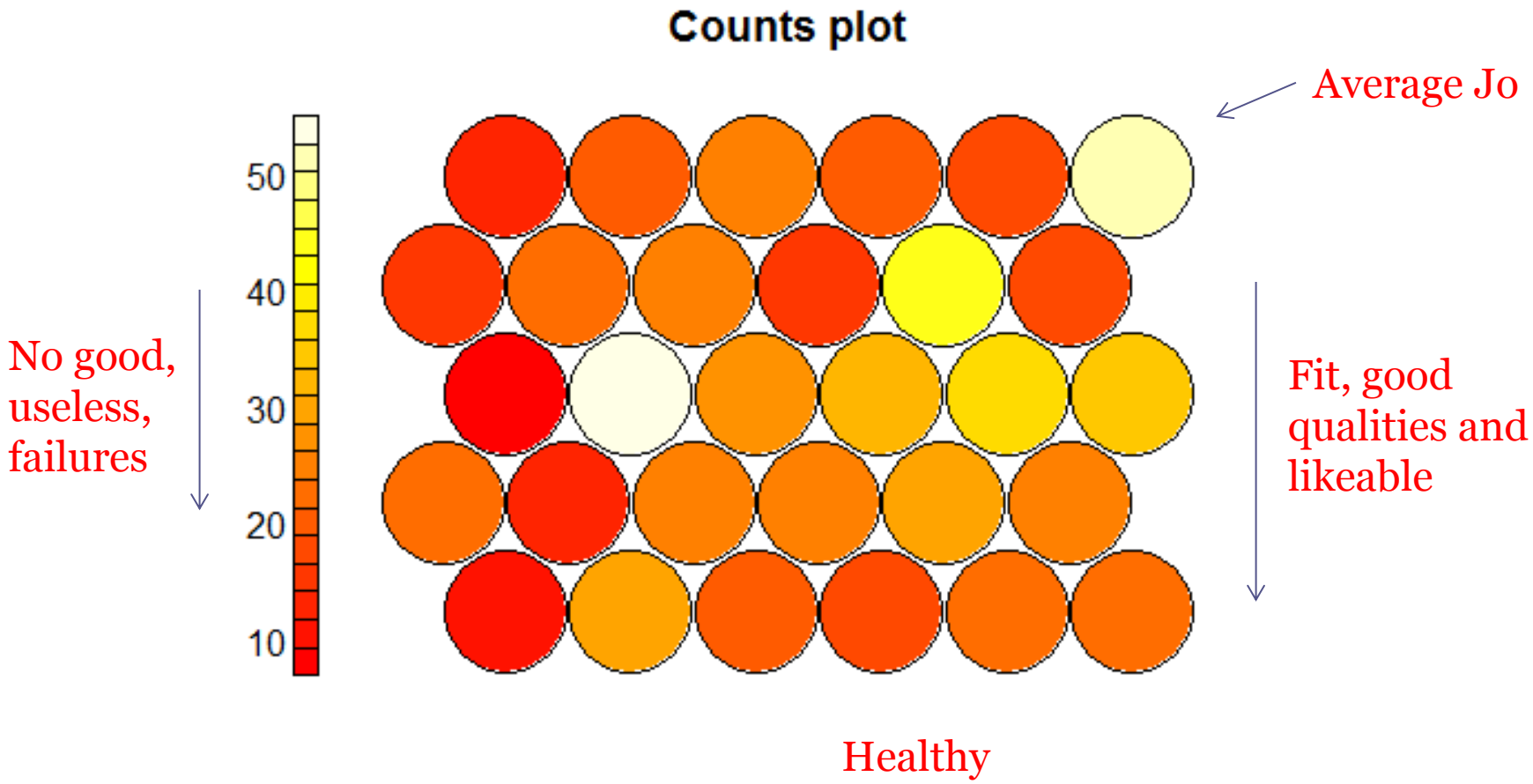
British Youth Panel ‘self-esteem’ data

- BYP around 1000 teenagers wave D (1994)
- Part of the BHPS and collected annually
- Data on self-esteem (4 point scales)
  - I have good qualities
  - I feel useless at times
  - I’m a likeable person
  - Enjoy exercise to keep fit
  - Feel a failure
  - Feel I am no good
  - Health is good

SOM of self-esteem using BYP 1994 (using R)



Count plots using R



Genetic Algorithms (GAs)

- Used to determine optimal fitness based on a 'genetic' coding
- A population is set up based on characteristics of agents coded as chromosomes with a collection of 'genes'
  - ABCFDSCFDCFDRT etc…
- A fitness function is determined based on the genetic makeup of the chromosome (could be derived from NNs or regression equations)
- The chromosomes mate with each other at random using a set of routines where sections of genetic material are exchanged and mutations can occur. Choice of chromosome is based on fitness (higher fitness = higher probability of selection)
- Thus the population expands
- The fitness function is rerun and the weaker chromosomes are eliminated
- Eventually an optimal population appears that is stable
- This represents the ideal set of agents

Example

- Could develop a model of well-being (i.e. fitness) based on the interaction of several characteristics using real data
  - (perhaps using a neural network architecture based on several interactions via a hidden layer)
- Code these characteristics into a chromosome pool based on the data
- Run the GA on the data
- Compare the fitness of the real population against the ideal one

Conclusions

- This is just scratching the surface of the methods available
- It is not entirely clear whether these methods have advantages over more traditional statistical methods
  - However there is promising scope:
    - NNs are more efficient at prediction than regression, but the weights are difficult to interpret
    - NNs can deal with interactions better than standard regression
    - SOMs have been proven to be efficient at determining clusters with very large data arrays, but in scientific applications such as chemistry and biology, and also bibliometrics and library science
    - GAs are very efficient at solving certain problems (such as the travelling salesman problem)
- Software is generally designed for use by, for example, bioinformatics and computer scientist types and not always so accessible to social scientists (R, Matlab, Weka)