Multi-Layer Perceptrons

Michael J. Watts

http://mike.watts.net.nz

Lecture Outline

- Perceptron revision
- Multi-Layer Perceptrons
- Terminology
- Advantages
- Problems

Perceptron Revision

- Two neuron layer networks
- Single layer of adjustable weights
- Feed forward network
- Cannot handle non-linearly-separable problems
- Supervised learning algorithm
- Mostly used for classification

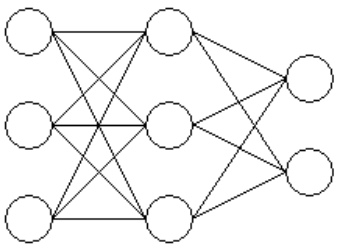
Multi-Layer Perceptrons

- Otherwise known as MLPs
- Adds an additional layer (or layers) of neurons to a perceptron
- Additional layer called hidden (or intermediate) layer
- Additional layer of adjustable connections

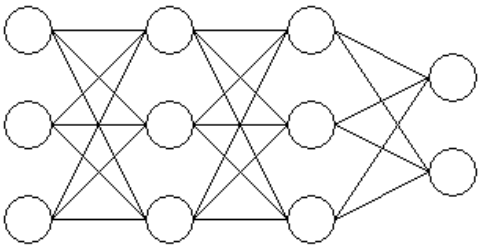
Multi-Layer Perceptrons

- Proposed mid-1960s, but learning algorithms not available until mid-1980s
- Continuous inputs
- Continuous outputs
- Continuous activation functions used
  - e.g. sigmoid, tanh
- Feed forward networks

Multi-Layer Perceptrons



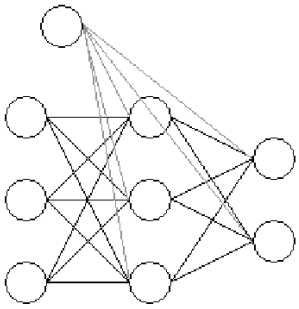
Multi-Layer Perceptrons



Multi-Layer Perceptrons

- MLPs can also have biases
- A bias is an additional, external, input
- Input to bias node is always one
- Bias performs the task of a threshold

Multi-Layer Perceptrons



Multi-Layer Perceptrons

- Able to model non-linearly separable functions
- Each hidden neuron is equivalent to a perceptron
- Each hidden neuron adds a hyperplane to the problem space
- Requires sufficient neurons to partition the input space appropriately

Terminology

- Some define MLP according to the number of connection layers
- Others define MLP according to the number of neuron layers
- Some don't count the input neuron layer
  - doesn't perform processing
- Some refer to the number of hidden neurons layers

Terminology

- Best to specify which
  - i.e. "three neuron layer MLP", or
  - "two connection layer MLP", or
  - "three neuron layer, one hidden layer MLP"

Advantages

- A MLP with one hidden layer of sufficient size can approximate any continuous function to any desired accuracy
  - Kolmogorov theorem
- MLP can learn conditional probabilities
- MLP are multivariate non-linear regression models

Advantages

- Learning models
  - several ways of training a MLP
  - backpropagation is the most common
- Universal function approximators
- Able to generalise to new data
  - can accurately identify / model previously unseen examples

Problems with MLP

- Choosing the number of hidden layers
  - how many are enough?
  - one *should* be sufficient, according to the Kolmogorov Theorem
  - two will always be sufficient

Problems with MLP

- Choosing the number of hidden nodes
  - how many are enough?
  - Usually, the number of connections in the network should be less than the number of training examples
- As number of connections approaches the number of training examples, generalisation decreases

Problems with MLP

- Representational capacity
- Catastrophic forgetting
- Occurs when a trained network is further trained on new data
- Network forgets what it learned about the old data
- Only knows about the new data

Problems with MLP

- Initialisation of weights
- Random initialisation can cause problems with training
  - start in a bad spot
- Can initialise using statistics of training data
  - PCA
  - Nearest neighbour

Problems with MLP

- Weight initialisation
  - can initialise using an evolutionary algorithm
  - initialisation from decision tress
  - initialisation from rules

Summary

- MLP overcome the linear separability problem of perceptrons
- Add an additional layer of neurons
- Additional layer of variable connections
- No agreement on terminology to use
  - be precise!

Summary

- Convergence is guaranteed over continuous functions
- Problems exist
  - but solutions also exist