Reinforcement Learning

Todd Sierens[1,2]

[1]Perimeter Institute for Theoretical Physics
[2]University of Waterloo

May 5, 2016

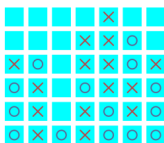# Outline

Supervised Learning vs Reinforcement Learning

- Supervised Learning
  - Have access to a large set of data with known desired results
  - Adjust model parameters to minimize an objective function
- Reinforcement Learning
  - Have access to an environment that can be modeled
  - Typically a reward function is used as a signal for how to adjust model parameters
  - Board games present a natural environment that is easily modeled and can provide a reward whenever a game finishes.
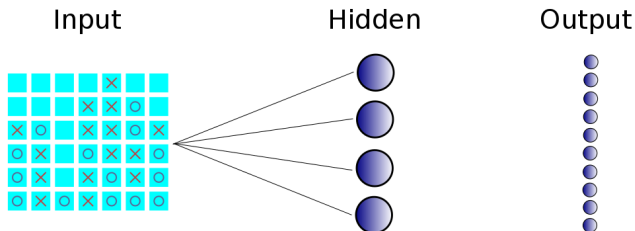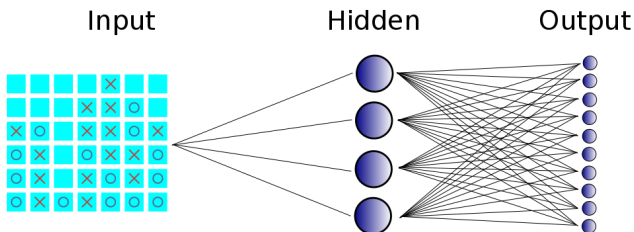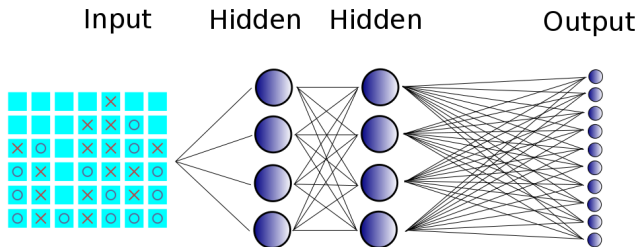
Simple Perceptron

Input

Hidden

Output

Machine Learning
○

Neural Networks
○●○
○
○○

Temporal Difference Learning
○○

Results
○○○○

Future Directions
○

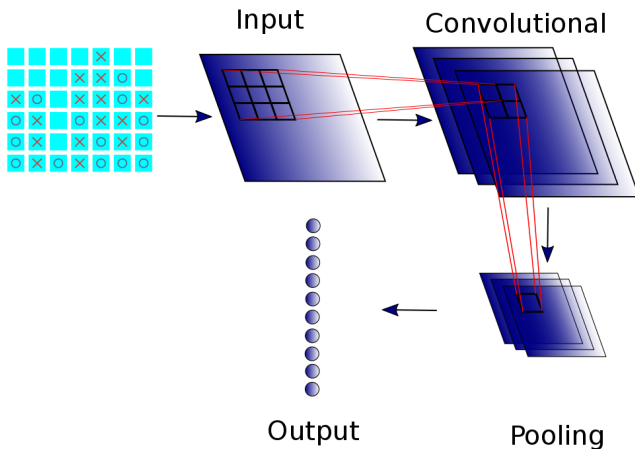# Simple Perceptron

Input   Hidden   Output

Simple Perceptron

Input   Hidden   Output

Multilayer Perceptron

Input   Hidden   Hidden   Output

Machine Learning
○

Neural Networks
○○○
○
●○

Temporal Difference Learning
○○

Results
○○○○

Future Directions
○

Convolutional Neural Network

Input

Convolutional

Output

Pooling

Machine Learning
○

Neural Networks
○○○
○
○●

Temporal Difference Learning
○○

Results
○○○○

Future Directions
○

Value Network

- A value network is a neural network that evaluates an environment and determines a value to associate to it
- In the case of board games, a value network can be used to determine the probability of winning from any given position
- Here is an example of a multi-layered neural network in action
- The network takes a Tic Tac Toe board as input, and through a succession of node activations the network outputs a prediction on who will win, in this case: the red player

Tic Tac Toe Position
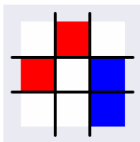
Network Node Activations
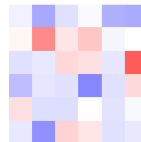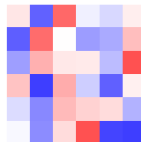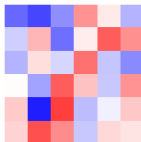
Predicted Winner
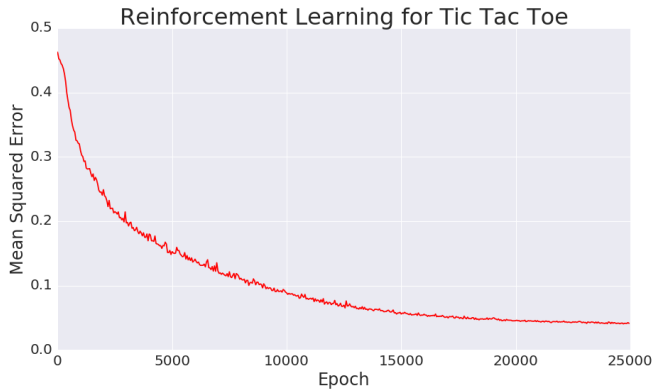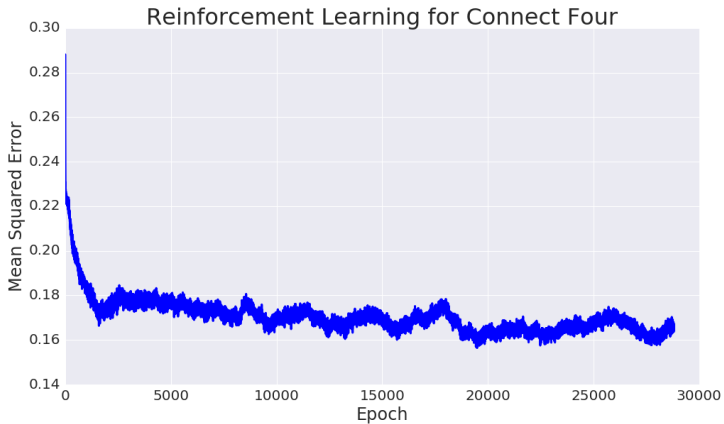
Temporal Difference Learning

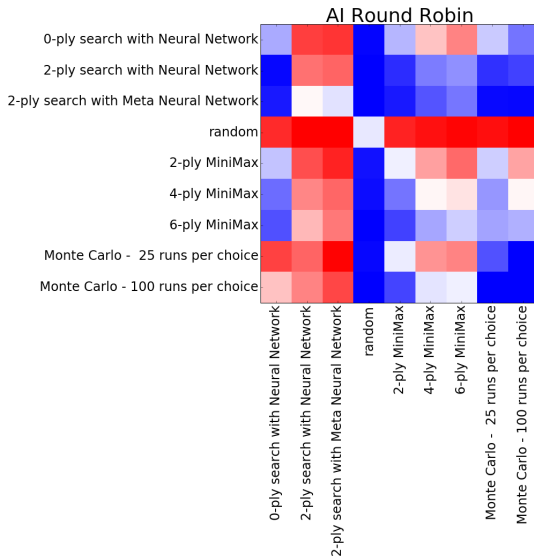- TD(λ) Equation

$$T_n = \quad (1 - \lambda) \quad V_n + \lambda \quad r$$

Temporal Difference Learning

- TD(λ) Equation

$$T_n = \sum_{n=0}^{N_0} (1 - \lambda)\lambda^n V_n + \lambda^{N_0} r$$

Reinforcement Learning for Tic Tac Toe

Reinforcement Learning for Connect Four

Machine Learning
○

Neural Networks
○○○
○
○○

Temporal Difference Learning
○○

Results
○○○●○

Future Directions
○

AI Round Robin

AI performance
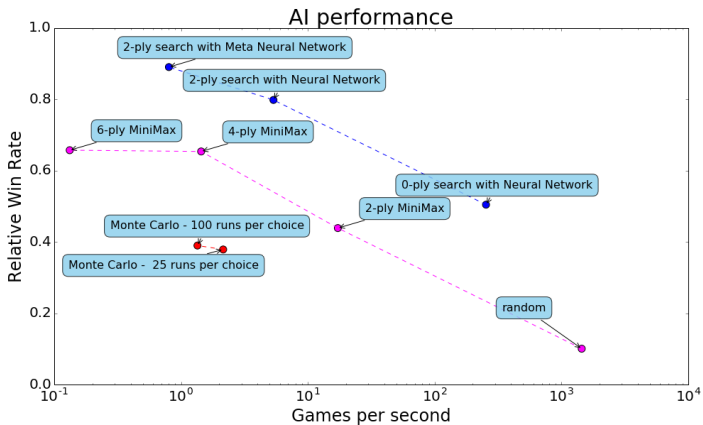
Future Directions

- Apply to more games
  - chess, hex, backgammon, go, video games, etc.
- Route finding app
- Traffic prediction analysis
- Control (robots, self-driving cars, etc.)
- Prediction (stock prices, sports betting, etc.)
- Acquire large data sets and compare with supervised learning