

Reinforcement Learning

Todd Sierens^{1,2}

¹Perimeter Institute for Theoretical Physics

²University of Waterloo

May 5, 2016

Outline

- 1 Machine Learning
- 2 Neural Networks
 - Simple Perceptron
 - Multilayer Perceptron
 - Convolutional Neural Network
- 3 Temporal Difference Learning
- 4 Results
- 5 The End

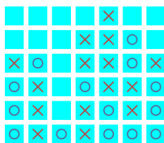
Supervised Learning vs Reinforcement Learning

- 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.

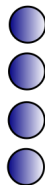


Simple Perceptron

Input



Hidden

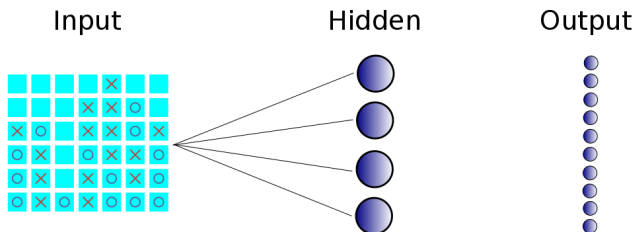


Output



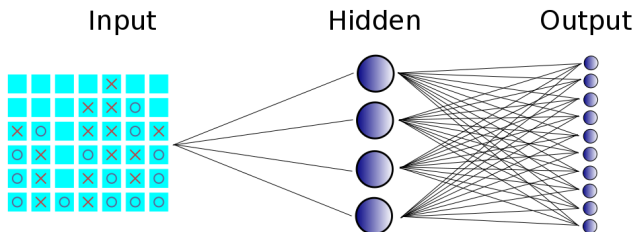


Simple Perceptron



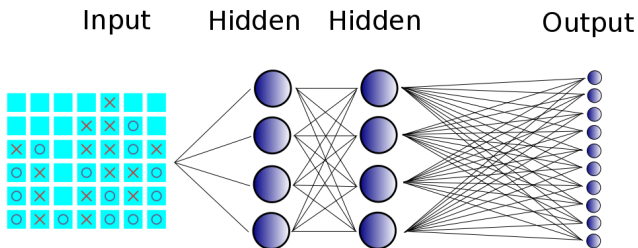


Simple Perceptron



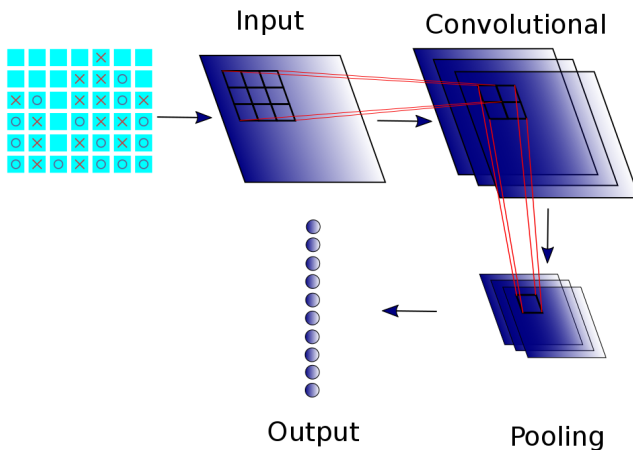


Multilayer Perceptron





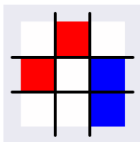
Convolutional Neural Network



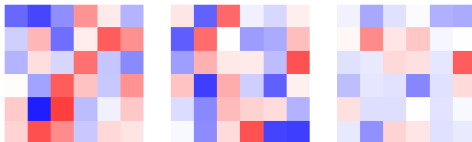
Value Network

- 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

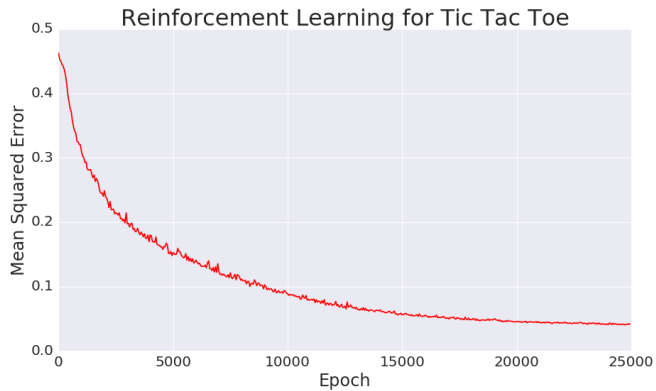
$$T_n = (1 - \lambda) V_n + \lambda r$$

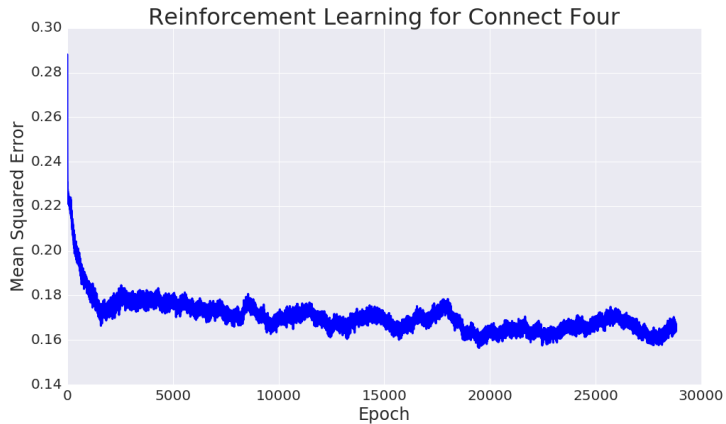


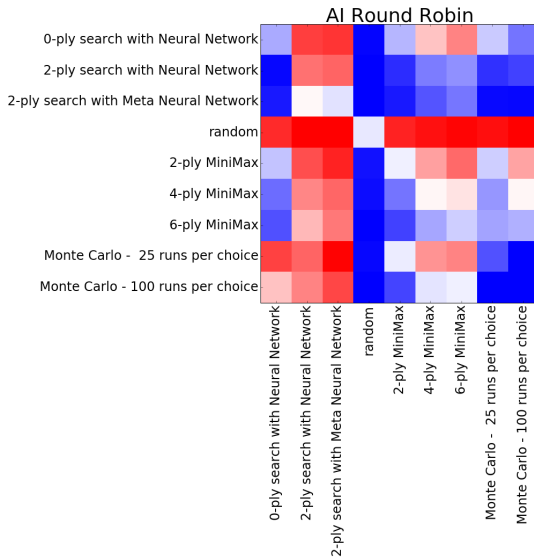
Temporal Difference Learning

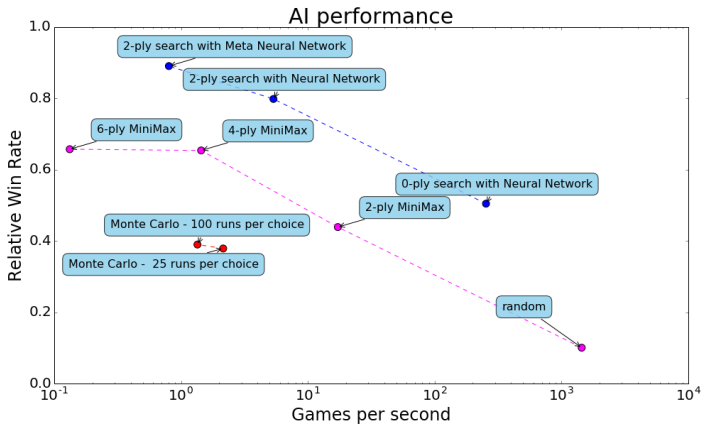
■ TD(λ) Equation

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









The End

The End