

Artificial Neural Networks Markov Decision Processes

Shiwali Mohan

Computer Science and Engineering
University of Michigan

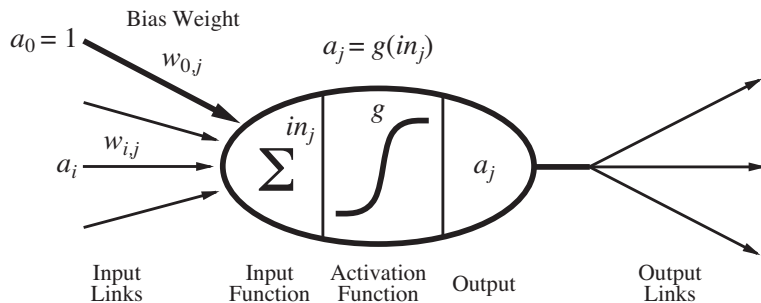
May 25, 2012

Outline

① Artificial Neural Networks

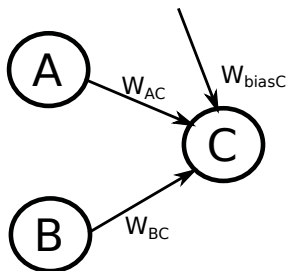
② MDP and Learning

Perceptron Review



Perceptron Design

- Implement $X \vee Y$ using a single perceptron unit. Assume bias input -1.



W_{AC}	1
W_{BC}	1
$W_{\text{bias},C}$	0.5

Perceptron Network Design

- Implement $X \oplus Y$ using a perceptron network.

W_{AC}	1
W_{AD}	-1
W_{BC}	-1
W_{BD}	1
$W_{\text{bias},C}$	0.5
W_{CE}	1
$W_{\text{bias},D}$	0.5
W_{DE}	1
$W_{\text{bias},E}$	0.5

Perceptron Network Design

- Suppose that after training for a long time, the weights of a 3 input (A, B, C) perceptron converge to the values $W_a = 1$, $W_b = 0.5$, $W_c = 0.5$, $W_0 = -0.75$ (bias input is 1). Approximately what Boolean function has the perceptron learned? (Express in logical form as a function of the inputs A, B, and C. You can assume that the inputs take on values of either 0 (false) or 1 (true)). Use step function for activation.

Perceptron Network Design

- Suppose that after training for a long time, the weights of a 3 input (A, B, C) perceptron converge to the values $W_a = 1$, $W_b = 0.5$, $W_c = 0.5$, $W_0 = -0.75$ (bias input is 1). Approximately what Boolean function has the perceptron learned? (Express in logical form as a function of the inputs A, B, and C. You can assume that the inputs take on values of either 0 (false) or 1 (true)). Use step function for activation.

	A	B	C	W0 (-0.75)	Sum(wx)	G(Sum(wx))
	0	0	0	1	-0.75	0
	0	0	1	1	-0.25	0
	0	1	0	1	-0.25	0
•	0	1	1	1	0.25	1
	1	0	0	1	0.25	1
	1	0	1	1	0.75	1
	1	1	0	1	0.75	1
	1	1	1	1	1.25	1

$$A \vee B \wedge C$$

MDP review

- What kinds of problems can we reason about with MDPs?

MDP review

- What kinds of problems can we reason about with MDPs?
 - MDPs are used for making decisions about how to act in a world. The basic setup is that we have an agent who gets some information about the world, takes an action, gets a reward. The goal is to learn an action plan that will maximize reward.

MDP review

- What kinds of problems can we reason about with MDPs?
 - MDPs are used for making decisions about how to act in a world. The basic setup is that we have an agent who gets some information about the world, takes an action, gets a reward. The goal is to learn an action plan that will maximize reward.
- What is the Markov assumption?

MDP review

- What kinds of problems can we reason about with MDPs?
 - MDPs are used for making decisions about how to act in a world. The basic setup is that we have an agent who gets some information about the world, takes an action, gets a reward. The goal is to learn an action plan that will maximize reward.
- What is the Markov assumption?
 - The future is independent of the past, given the present.

MDP review

- What kinds of problems can we reason about with MDPs?
 - MDPs are used for making decisions about how to act in a world. The basic setup is that we have an agent who gets some information about the world, takes an action, gets a reward. The goal is to learn an action plan that will maximize reward.
- What is the Markov assumption?
 - The future is independent of the past, given the present.
- How is an MDP defined?

MDP review

- What kinds of problems can we reason about with MDPs?
 - MDPs are used for making decisions about how to act in a world. The basic setup is that we have an agent who gets some information about the world, takes an action, gets a reward. The goal is to learn an action plan that will maximize reward.
- What is the Markov assumption?
 - The future is independent of the past, given the present.
- How is an MDP defined?
 - States, actions, transition models, rewards

MDP review

- What kinds of problems can we reason about with MDPs?
 - MDPs are used for making decisions about how to act in a world. The basic setup is that we have an agent who gets some information about the world, takes an action, gets a reward. The goal is to learn an action plan that will maximize reward.
- What is the Markov assumption?
 - The future is independent of the past, given the present.
- How is an MDP defined?
 - States, actions, transition models, rewards
- Algorithm for solving MDP.
 - Value iteration.

Value Iteration Example

	1	2	3	4	5
A	\$				
B					
C					

- **Actions:** forward, back, left, right, idle, pick-up
- **Rewards:** pick-up treasure: 100, move: -5, idle: none
- **Transition:** 0.8 move intended, 0.1 move to left, 0.1 move to right

- Formulate the problem as an MDP
- Compute utilities and optimal policy in the first iteration.
- In second iteration.
- Sketch an optimal policy.

Reinforcement Learning

- What if rewards were not known?
- Review TD/Q learning.