Alin Tomescu, http://people.csail.mit.edu/~alinush
6.867 Machine learning | George Konidaris, PhD | Week 14, Thursday, December 5th, 2013| Lecture 25

Lecture 25: Reinforcement learning

Reinforcement learning is a machine learning paradigm concerned with the problem of learning how to interact with an environment to maximize reward.

Formulation: Agent interacts with an environment.

At each time t:

- Receives sensor signal s_t
- Executes action a_t
- Transition:
  - New sensor signal s_{t+1}
  - Reward r_t

Goal: find policy π that maximizes expected return (sum of discounted future rewards):

$$\max_{\pi} E\left\{R = \sum_{t=0}^{\infty} \gamma^t r_t\right\}$$

- Immediate rewards are valued more than future rewards (that's what γ takes care of)

You can apply this formulation to a broad set of scenarios

Markov decision process

Very useful formalism of the RL problem

$$\langle R, A, \gamma, R, T \rangle$$

- S: set of states
- A: set of actions
- γ: discount factor
- R: reward function:
- T: tranisiton function
- $T(s'|s, a)$ is the probability of landing in state s' given you were in state s and took action a