AINT351 - Reinforcement Learning and Dynamic programming

Date: 17-10-16

RL

- Learnign what to do when
- What
  - What actions
- When
  - State
- Feedback
  - Reward
- Used for a lot
  - Decision making

Representing a problem

- Finite state machine
- **S** - Set of states
- **A** - Set of actions
- **T** - Transition function
- **R** - Reward function
- Agent changing the environment through an action
  - Feedback is reward and state whilst it goes back to the agent to make another decision

Probabilistic actions

- Not always certain that an action will lead to a result
- 90% chance of going north if you move north
  - Due to noise, other variables

Transition function

- Taken an action given a state
  - The probability of landing in another state

Policy

- The probability of choosing a given action based on a given state
- Decision learnt within the agent
- Commonly based on a value function and an action selection mechanism
- Value
  - ascribes values to different states, then selects state with highest value
  - Greedy action, selections always with highest value state

State-value function

- How good is a policy
- Evaluated in term of expected reward
- Instead of looking at all steps
  - Look at one big estimate
  - As we repeat it converges to make information about unknown states