Reinforcement Learning

Some Remarks reg. Material for Exam

Subramanian Ramamoorthy
School of Informatics

7 April, 2017

Structure of the Exam

- Style is similar to past years' exams
  - 1 compulsory question for everyone (25 marks)
  - 1 additional question to be answered – choose from 2 alternatives (25 marks)
  - Final marks are scaled to be worth 80% of course mark
- Each question has many subparts, could be combination of:
  - Bookwork, e.g., what is actor-critic architecture?
  - Modelling questions, e.g., given word problem, can you write down an MDP description?
  - Technical questions regarding some aspect of a famous algorithm

What is Expected?

- Concise answers addressing only the specific question being asked
- You will not get extra credit for long answers, or for detailed description of neighboring topics that are not the subject of the question
- A majority of the questions will only require
  - A few sentences explaining your reasoning
  - A few equations or other symbolic descriptions
  - A figure or two where needed
- These exams will not typically require long derivations, instead they will try to test **understanding** of the concepts

Coverage of Material

- Focus on the core concepts and make sure you understand **why** things are the way they are
  - Think about when assumptions get violated and what that might imply
  - Try to construct toy examples as thought experiments for yourself

- You do need to remember core formulas, e.g., definition of value, structure of TD updates or setup of POMDP
  - But emphasis *will not* be on rewarding your ability to memorise formulas and derivations

Highlights of Concepts Covered

- Multi-armed bandits
  - Nature of this decision problem
  - Simplest concepts of value and explore-exploit tradeoff
  - Many ways of making that choice
    - Sample averages
    - Regret
    - Confidence interval based reasoning
- Markov Chains and MDPs through that
  - Focus was on understanding nature of the decision problem and solution concepts before looking at learning
  - Do you understand key elements and properties of a Markov chain?

Highlights of Concepts Covered

- Markov Chains and MDPs through that, contd.
  - MDP is basically a "controlled" Markov Chain, so we can apply this understanding to think about policy choice
- Dynamic Programming
  - Value function and Bellman's equation, in various forms
  - Using this, we looked at
    - Policy evaluation, estimating value given a policy
    - Policy improvement
    - Combining the two into Policy iteration
    - Value iteration as an alternative procedure
  - Revisit text book examples to check your understanding

Highlights of Concepts Covered

- Monte Carlo methods
  - Sample-based rather than model-based computation of value
    - Similar concept of value but computation is quite differently done
  - Again, a procedure for policy evaluation and improvement
  - Notion of on-policy vs. off-policy learning: do you get the core concept (we will not need the detailed derivations)
- Temporal Difference Learning
  - Improves on MC by doing updates as episodes proceed
  - TD(0), SARSA and Q-learning: do you understand the update rule and the differences between them?
  - Again, spend some time looking over text book examples to understand the behaviour of these methods

Highlights of Concepts Covered

- Function approximation
  - Features vs. states
  - What is objective of learning procedure?
  - Structure of linear FA and gradient computation
  - Various ways of setting up features, coding alternatives

Beyond this point, we switch gears and go to material not in S+B book

Highlights of Concepts Covered

- Abstraction and hierarchy
  - Initial general remarks to set the scene: you do not need to know details of these
  - SMDP: do you understand the core formulation, i.e., how to define Q/V in terms of this model
  - Elevator example to give a feel for the model: you don't need to remember any detail of this example but make sure you understand how SMDP is used there
  - Options: what are the core elements and do you understand their use in the rooms example?
  - A simple procedure for extracting options from data: what are the core elements?

Highlights of Concepts Covered

- POMDP
  - What is it and how does it differ from a regular MDP?
  - What is a belief state and how do we define Value over it?
  - Core computational steps for performing policy evaluation, by going through one step of observation and state transition
  - Do you understand what PBVI does? How and why does this help beyond basic POMDP iteration?
- IRL
  - Basic idea of behavioural cloning
  - Ng+Russell formulation of IRL as LP
  - Combining this LP with FA
  - You do not need to be able to reproduce full derivation, but you should know the problem formulation

Highlights of Concepts Covered

- Exploration and Controlled Sensing
  - Basic ideas of Bayesian updates given measurements: not unlike what happens in POMDP observation update
  - Value of Information
  - Information gain and use in exploration
  - You need to understand the core definitions and concepts, but you will not be quizzed on specific examples, etc.
- MARL
  - Core concepts: game, strategy, equilibrium
    - Do you understand these and can you precisely define them?
    - What is a solution concept and how does this work in computation?
  - Stochastic game and applying TD style learning within them