Reinforcement Learning

In-class tutorial: Worked examples
[DP, MC, basics of TD]

Subramanian Ramamoorthy
School of Informatics

17 January 2017
Plan for the Session

• Problems chosen to illustrate concepts covered in earlier lectures

• We will work out problems on the board and take questions to clarify concepts

• These slides provide the outline sketch of the questions to be covered
0. Interpretation of V and Q

Using the task of selecting a club to play the game of golf, discuss the meaning of V and Q.

What are:
- States
- Actions
- Rewards

What do you understand by the shape and numbers in this figure?
I. Interpretation of $V^\pi$ and $\pi$

- Cells = States
- NSEW actions resulting in movement by 1 cell
- Actions taking agent off grid have no effect but incur reward of -1
- All other actions result in a reward of 0
  - except those that move the agent out of the special states A and B.

Inspect and interpret $V^\pi$
I. Interpretation of $V^\pi$
I. Interpretation of $V^*$ and $\pi^*$

Calculate and show that Bellman’s equation holds for centre state – to understand nature of $V^*$
Finding the shortest path in a graph using optimal substructure; a straight line indicates a single edge; a wavy line indicates a shortest path between two vertices it connects (other nodes on these paths are not shown); bold line is the overall shortest path from start to goal. [From Wikipedia]

Understanding the recursion: If shortest path from LA to NY must include Chicago, then shortest path from LA to Chicago can be computed separately from last leg.
II. Value/Policy Iteration using Grid World

• Calculate initial steps of Policy Evaluation using a grid world example seen in our earlier lectures

\[ R = -1 \] on all transitions
$V^\pi$ and Greedy $\pi$ at $k = 2$
III. MC Value Evaluation

- Work out some steps of the MC value evaluation process for the 5-state Markov Chain example (for a random walker who goes one step to the left or right with equal probability)
IV. Understanding MC through modified random walk

• The transition probabilities for state C are as shown. For all other states, the transitions are based on a fair coin flip. The square is an absorbing terminal state with reward as shown.

Perform some initial steps of calculation of $V^\pi$ using first-visit MC.

Discuss MC with Exploring Starts, etc.

Exploring starts: Every state-action pair has a non-zero probability of being the starting pair.
Discuss SARSA and Q-learning procedures with respect to this example