Project 3: Markov Decision Processes and Project 4 Proposal

- **Due: November 9. Project proposal should *not* be turned in late, or else my feedback to you on it will not be very useful.**
- **MDP Problems**: Fine to do by yourself or with a partner. More depth will be expected of a two-person project.
- **Project proposal**: Preferred if you do it in a two-person team, but not required.

1 Markov Decision Processes

This assignment is to formalize two domains as MDPs and then solve them using your own implementation of value iteration.

1.1 Domains

You can use these two domains, or you can make up one of your own (and do one of these). If you make up your own, be sure you describe it clearly in English.

Your write-up should provide a mathematical description of each MDP model (as well as the description in your code). Include a description of the results of running value iteration on your problems at the intuitive level: What does the policy do? Does it make sense?

1.1.1 Rental cars

Love, 1985[1]

A rental car agency serves two cities with one office in each. The agency has M cars i total. At the start of each day, the manager must decide how many cars to move from one office to the other to balance stock. Let $f_i(q)$ for $i \in \{1, 2\}$ denote the probability that the daily demand for cars to be picked up at city i and returned to city i equals q. Let $g_i(q)$ for $i \in \{1, 2\}$ denote the probability that the daily demand for 'one-way' rentals from city i to the other equals q. Assume that all rentals are for one day only, that it takes one day to move cars from city to city to balance stock, and if demand exceeds availability, customers go elsewhere. The economic parameters include a cost K per car for moving a car from city to city to balance stock, a revenue of R_1 per car for rentals returned to the rental location and R_2 per car for one-way rentals.

[1] This and other problems with citations are taken from *Markov Decision Processes: Discrete Stochastic Dynamic Programming* by Martin L. Puterman, Wiley, 1994

a. Formulate this system as an MDP. Solve it for some maximum value of q.

b. Suppose, in addition to relocating cars to balance stock, the manager may reserve a certain number of cars for one-way rental only. Formulate this system as an MDP.

1.1.2 Lazy adaptable lion

Mangel and Clark, 1988 An adult female lion requires about 6 kg of meat per day and has a gut capacity of about 30 kg; which means that, with a full gut, it can go six days without eating. Zebras have an average edible biomass of 164 kg, large enough to meet the food demands of many lions. If a female selects to hunt in a group, the probability of a successful kill per chase has been observed to increase with group size to a point, and each chase consumes 0.5 kg of energy. Suppose that the probabilities of a successful hunt are given by $p(1) = 0.15, p(2) = 0.33, p(3) = 0.37, p(4) = 0.40, p(5) = 0.42, p(\geqslant 6) = 0.43$ where $p(n)$ represents the kill probability per chase for lions hunting in a group of size n.

Formulate this as an MDP in which the lion decides whether to hunt and, if so, in what size group. Assume one hunt per day and that if the hunt is successful, lions in the hunting group share the meat equally.

a. Formulate and solve for the case in which lion's objective is to maximize its probability of survival over 20 days, with no discounting.

b. Formulate and solve a discounted model. Select a value for γ that has an expected horizon of 20.

1.2 Value iteration

Using any language you like, implement the value iteration algorithm. Be sure it takes a reasonably generic description of an MDP as input: it shouldn't be tailored at all to the particular problem domains you selected.

Include the code for your value iteration solver and for the description of the problem domains.

2 Project Proposal

Propose a final project for the course! You can concentrate on any area we have discussed, including classical planning, MDPs, reinforcement learning, or partially observable domains (if you're interested in this, you may have to read ahead a bit in advance of the lectures).

Your project should consist of:

- Formulation of at least one planning domain. If it's just one domain, then that domain should have some interesting variability among its instances (e.g., size, distribution of transition probabilities, degree of observability, etc.).
- Implementation of one solution method.
- Either implementation of another solution method or use of existing software implementing another method.
- A comparative study of the methods in the domain(s), including quantitative results and a qualitative analysis of those results.
- A paper and a 20-minute talk describing the results.

Your proposal should be at least one page, and sketch the proposed domains and algorithms. It should:

- Provide three milestones (in which the last one is the finished project) with associated dates.
- Include citations if you want to implement a method or use a problem described somewhere other than in the textbook.
- For a multi-person project, include a rough task breakdown, including who will be primarily responsible for which parts.

I'm happy to talk with anyone who would like suggestions of papers to read, methods to try, interesting domains, etc.