Last Time

MICHIGAN ENGINEERING
UNIVERSITY of MICHIGAN ■ COLLEGE of ENGINEERING

- Special structured optimal policies

Midterm

- Midterm Exam Reminder:
  - Take home exam:
    - Starts Oct 23, 10:30am
    - Ends Oct 26, 10:30am
  - Worth 30% of final grade
  - You can use lecture notes, Dreyfus and Law, and Puterman (no other sources)

Midterm

- Midterm Exam:
  - Commonly asked questions:
    - Are there office hours during the exam? No
    - How many questions? Probably 4-5
    - How long will it take? This depends on your preparation for the exam
    - Are the questions applied or theoretical? Both
    - How hard are the questions? Similar level of difficulty to assignments
    - Can I use Matlab? Yes
    - Will you grade all the questions? Yes

Midterm: Sample Front Page

MICHIGAN ENGINEERING
UNIVERSITY of MICHIGAN ■ COLLEGE of ENGINEERING

IOE512 Dynamic Programming– Midterm Exam

Due Thursday October 26, 10:30am

Exam Requirements: This test is to be completed separately by each student. You may consult the class lectures and notes, chapters 1-3 of Dreyfus and Law, and chapters 1-4 of Puterman. **You may not consult any other resources and you may not discuss the questions or answers to any of these questions with any person**.

Sign below after you have completed the exam.

Honor Pledge: "I have read the above statement regarding requirements for this test. I have neither received nor given aid on this test, nor have I concealed any violations of the Honor Code"

Signed: _____

Print Name: _____

Instructions:

This test includes 5 Questions. Read each question carefully before answering. When answering, show all of your work. Points are allocated to each part of the solution process, and partial points will be given for partial solutions. Provide clear and concise answers.

Mini Project

Work in groups of 3 to develop and describe a novel dynamic programming method or application not covered in class

- Present project in-class (10 minute presentation)
- Presentations will be on the second and third last classes

Send me list of team members, project title, and a paragraph summarizing your topic no later than 10:30am Nov 8.

It will be graded as an In-Class Assignment.

Mini Project Examples

MICHIGAN ENGINEERING
UNIVERSITY of MICHIGAN ■ COLLEGE of ENGINEERING

- DP for hybrid electric vehicle controllers
- Optimal timing of treatment for HIV
- Spell checking algorithms for editors
- Approximate stochastic DP approaches for vehicle routing
- DP models for optimizing energy storage
- Temporal difference learning

Mini Project

MICHIGAN ENGINEERING
UNIVERSITY of MICHIGAN ■ COLLEGE of ENGINEERING

Grade based on class survey which will count for **1/3 of participation component** (5% of final grade overall);
Survey question:

"This presentation was clear, interesting, and informative: 1, 2, 3, 4, 5".

Resources for presentations:

http://www.nature.com/scitable/ebooks/english-communication-for-scientists-14053993/giving-oral-presentations-14239332

http://www.pcworld.idg.com.au/slideshow/366369/world-worst-powerpoint-presentations/

Excellent YouTube Video: https://www.youtube.com/watch?v=meBXuTIPJQk

Example: Maintenance Problem

A tool **deteriorates** stochastically with states $S = \{0,1,2,\ldots\}$.The decision maker chooses from actions $\mathrm{A} = \{\mathrm{D}, \mathrm{R}\}$, to replace the tool (R) or defer replacement (D). The tool deteriorates by i states with probability $\mathrm{p(i)}$. **Action R has the potential to returns the tool to the ideal state 0**.

Transition probabilities:

$$\mathrm{p_t}(j|s,D) = \begin{cases} 0, & j < s \\ p(j-s), & j \geq s \end{cases}$$

and $\mathrm{p_t}(j|s,R) = p(j), j \geq 0$. Rewards:

$$\mathrm{r_t}(s,a) = \begin{cases} W - h(s), & a = D \\ W - K, & a = R \end{cases}$$

Where W is a fixed reward for each epoch, K is a fixed cost of replacement, and h(s) is a state dependent nonnegative maintenance cost. The salvage value at the end of period N is $\mathrm{R_N}(s)$

Exercise: Using Theorem 4.7.5 provide conditions under which there exists an optimal policy that is monotone.

Exploiting Monotonicity

To use this algorithm, states $s = 1, \ldots, M$, are ordered such that optimal actions, $a^*(s)$, are non-decreasing in s.

Complete set of possible actions

Algorithm (Monotone Backward Induction):

1. $v_N(s_N) = R_N(s_N), \quad for\ all\ S_N$, Set $\mathrm{t} = \mathrm{N} - 1$, $Set\ A_1 = A$

2. $Evaluate\ v_t(s_t)\ for\ all\ s_t = 1\ to\ M\ as$:

$$v_t(s_t) = \max_{\mathrm{a} \in A_{s_t}} \{\mathrm{r_t(s_t}, a) + \lambda \sum_{j \in S} p_t(j|s_t, a) v_{t+1}(j)\}$$

Set of optimal actions at stage t

$$A^*_{s_t} = \arg\max_{\mathrm{a} \in A_{s_t}} \{\mathrm{r_t(s_t}, a) + \lambda \sum_{j \in S} p_t(j|s_t, a) v_{t+1}(j)\}$$

$$a^*_{s_t} = \max\{a \in \mathrm{A^*_{s_t}}\}$$

New set of possible actions at state $s_t + 1$

$$A_{s_t+1} = \{a \in A | a \geq a^*_{s_t}\}$$

3. If $\mathrm{t} = 1$ then stop; otherwise $\mathrm{t} = \mathrm{t} - 1$, and return to step 2.

In-Class Assignment

Assume the manufacturing tool has states $S = \{G, B, F\}$, i.e., good (G), bad (B), failed (F) .The decision maker chooses from actions $A = \{D, R\}$, to replace the tool (R) or defer replacement (D). There are three stages t=1,2,3.

Transition probability matrix:

$$P = \begin{bmatrix} 0.6 & 0.3 & 0.1 \\ 0 & 0.4 & 0.6 \\ 0 & 0 & 1.0 \end{bmatrix}$$

and repair returns the state to G w.p. 1.

Rewards:

$$r_t(s, a) = \begin{cases} 1 - h(s), & a = D \\ 0, & a = R \end{cases}$$

Where W=1, K=1, $h = (0,1,2)$ and the salvage is $R_3 = (2, 1, 0)$.

Exercise: Solve this problem using monotone backward induction.