# Homework #3
# Defeat Policy Iteration

## Problem

### Description

Policy iteration (PI) is perhaps the most under-appreciated algorithm for solving MDPs. Although each iteration is expensive, it generally requires very few iterations to find an optimal policy. In this problem, you'll gain an appreciation for how hard it is to get policy iteration to break a sweat.

Currently, it is not known whether there is an MDP which requires more than a linear number of PI iterations in the number of states of the MDP. Your goal is to create a 30 state MDP that attains at least 15 iterations of PI before the algorithm terminates.

### Procedure

- Construct an MDP with at most 30 states and exactly 2 actions per state. You may assume the discount factor is 3/4. The MDP may have stochastic transitions.
- Use an editor or a simple program to create a json description of the target MDP that is parseable by the tester.
    - the json created should use double quotes instead of single quotes
    - the entire description must be less than 100,000 characters
- Validate your description
    - http://jsonlint.com/
    - http://www.charactercountonline.com/
- Test your MDP locally with the provided tester to ensure you meet the submission requirements.

## Note

If you are able to obtain 1000 iterations (the default maximum iterations for pymdptoolbox's policy iteration method) you most likely have encountered a bug that may or may not be replicable

when we test your MDP. We would like to see your MDP so we can determine what the bug is. Please add it to a private post to the instructors on Piazza. Then, update your MDP and try again...

## Example

The following is an example of the json definition of a simple MDP

```
{
  "gamma":0.75,
  "states": [
    {
      "id": 0,
      "actions": [
        {
          "id": 0,
          "transitions": [
            {
              "id": 0,
              "probability": 0.5,
              "reward": 0,
              "to": 0
            },
            {
              "id": 1,
              "probability": 0.5,
              "reward": 0,
              "to": 1
            }
          ]
        }
      ]
    },
    {
      "id": 1,
      "actions": [
        {
          "id": 0,
          "transitions": [
            {
              "id": 0,
              "probability": 1,
              "reward": 1,
              "to": 1
            }
          ]
        }
      ]
    }
  ]
}
```

## Resources

The concepts explored in this homework are covered by:

- Lectures
    - Lesson 1: Smoov & Curly's Bogus Journey
    - Lesson 5: AAA
- Readings
    - Littman (1996)(chapters 1-2)

Additionally, a tool to create and test your MDPs can be found here:
https://github.com/rldm/hw3-tester

# Submission Details

## The due date is indicated on the Canvas page for this assignment.

Make sure you have set your timezone in Canvas to ensure the deadline is accurate.

Your MDP definition must reach 15 iterations. No partial credit will be given for less than 15 iterations.

All valid solutions that achieve 15 iterations will receive full credit.

The top 10 solutions will get an additional **10 points of extra credit** for this assignment. You will be ranked by number of iterations (higher, better) and last submission datetime (smaller, better).

Rankings will be released periodically by the TAs.

To complete the assignment, submit the description of your MDP to:

**https://**rldm.herokuapp.com

Finally, those of you able to obtain more than 31 iterations are eligible for a **smiley-face sticker** 😃. Please send a self-addressed, stamped envelope to:

Box 1910
Computer Science Department
Brown University
115 Waterman St.
Providence, RI 02912