COMPSCI 370	**Instructor:** Ronald Parr
Homework 3	**Due:** Thursday March 30, 2023

1 Value of Information Calculation (10 points)

Consider the traffic information question from the slides. In the slides, we assume that we are told the true value of the traffic variable T. In greater generality, however, we may only get a traffic report R which is correlated with T. In this context, we can think of the case in the slides as corresponding to:

$$\begin{aligned} P(r|t) &= 1.0 \\ P(\bar{r}|\bar{t}) &= 1.0 \end{aligned}$$

Assume $p(r|t) = 1.0$, but let $P(\bar{r}|\bar{t})$ be some free variable $0 \leq x \leq 1$. At what value of x is the VOI minimized, and what is the VOI at this point?

2 Value Iteration (10 points)

Consider the following problem: It is currently year 0. You have a choice between a agreeing to a one-time payment of 300 dollars at year 1, or a recurring payment of 30 dollars every year in perpetuity, also starting in year 1. You can think of the problem as having three states:

- S_0 which is the initial state in which you make the choice.
- S_1 is a terminal state corresponding to the choice for the one time payment of 300 dollars. Its value is, therefore, fixed at 300. There are no action choices in this state.
- S_2 corresponds to the choice for a 30 dollar per year payment in perpetuity. This state has a probability 1.0 transition to itself. There are not action choices in this state.

Assume the initial value function is $V_0 = 0$, i.e., zero for all states. For a discount factor of $\gamma = 0.95$, compute the first three iterations of value iteration, i.e., $V_1 \ldots V_3$.

Finally, you should be able to figure out what this will converge to, so work out V_∞, and explain how you got it.

3 Policy Iteration (10 points)

Working with the MDP from the previous question we'll now see show that policy iteration gives us the same answer. To demonstrate your understanding of policy iteration, do the following steps, showing your work:

a) Start with the policy that picks the one time papyment (call this π_1) and compute the exact value of that policy by solving the system of equations.

b) Compute the greedy policy for the value function from part a, and show that this picks the repeated payment. Call this policy π_2.

c) Solve the system of equations to compute value of π_2.

d) Demonstrate that you have converged by computing the greedy policy for V^{π_2} and verifying that it is unchanged from π_2.

4 Fun with discount factors (5 points)

Using the same example from the previous question, solve for discount factor $\gamma > 0$, where you would be indifferent between π_1 and π_2.