COMPSCI 370 **Instructor:** Ronald Parr

Homework 4 **Due:** Thursday March 31, 2022

10 Value of Information Calculation (5 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:

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

Assume $p(r|t) = 1.0$, but let $P(\overline{r}|\overline{t})$ 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?

11 MDP Value Determination (3 points)

Recall that value determination is the problem of computing the value of a fixed policy π, and that this can be done iteratively with the iteration:

$$V_{i+1}^{\pi} \leftarrow R + \gamma P^{\pi} V_i^{\pi}$$

Where the each V is a column vector, R is a column vector and P^{π} is the transition matrix for policy π. In this question we will show that under certain assumptions V_i is monotonically non-decreasing in i.

Assume that for some j, $V_{j+1} \geq V_j$, which means that for all states s, $V_{j+1}(s) \geq V_j(s)$. Prove that for any $k \geq j$, $V_{k+1} \geq V_k$. Hint: It's helpful to do induction on k, and use the fact that P^{π} is a row-stochastic matrix, i.e., it has rows that sum to 1.

12 Policy Iteration (4 points)

In the handwritten MDP example from class, we showed how to find the optimal policy for a trivial MDP by value iteration. Now we'll 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 action a_1 (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 action $a2$. 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.

13 Fun with discount factors (2 points)

Using the same example from the previous question, solve for the discount factor γ, where you would be indifferent between π_1 and π_2.