COMPSCI 370	**Instructor:** Ronald Parr
Homework 4	**Due:** Wednesday April 7, 2021

10 Value of Information Calculation (7 points)

For this question, we will consider the problem of buying an airline ticket to travel to an event this summer. Everything will be measured in units of utility so we don't need to worry about converting money to utility. Assume that the ticket will cost 300 units of utility, but that attending the event will provide you will 400 units of utility for a net benefit of 100 units. The utility of staying home and doing nothing is 0.

The reason why decision theory is involved is that we don't know if the event will actually happen because of the pandemic. The airline ticket is not refundable, and there's no other reason to visit the city in which the event is held, so the net utility in the scenario where you buy the ticket and the event is canceled is -300.

a) Suppose there is a $\frac{1}{6}$ chance that the event will be canceled. Show the work to compute the optimal decision (buy or no-buy) and the utility of this decision. (3 points)

b) Suppose somebody can predict with 100% accuracy whether the event will be held. Show the work to compute the value of this information to you. (4 points)

11 MDP Value Determination (2 points)

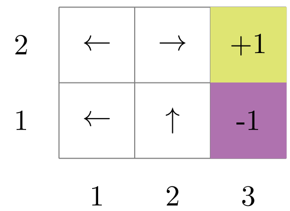

Figure 1: 3×2 Gridworld

Consider a simplified version of the gridworld example in the lecture slides, which is depicted in Figure 1. The robot can move in four directions: up, down, right and left. It moves in the intended direction with probability 0.8 and goes sideways with probability 0.1. If the direction is blocked by the wall, then it stays in the same cell. The rewards are zero except for the terminal states. The terminal sates have their values fixed at the values shown.

(a) Construct the transition matrix $\mathbf{P}_\pi$ for the policy in Figure 1 by filling in a table like the one below. Note that we have omitted entries that are zero. (1 point)

From	To	Probability
$\langle 1, 1 \rangle$	$\langle 1, 1 \rangle$	
$\langle 1, 1 \rangle$	$\langle 1, 2 \rangle$	
$\langle 1, 2 \rangle$	$\langle 1, 2 \rangle$	
$\langle 1, 2 \rangle$	$\langle 1, 1 \rangle$	
$\langle 2, 1 \rangle$	$\langle 1, 1 \rangle$	
$\langle 2, 1 \rangle$	$\langle 3, 1 \rangle$	
$\langle 2, 1 \rangle$	$\langle 2, 2 \rangle$	
$\langle 2, 2 \rangle$	$\langle 2, 2 \rangle$	
$\langle 2, 2 \rangle$	$\langle 2, 1 \rangle$	
$\langle 2, 2 \rangle$	$\langle 3, 2 \rangle$	

(b) Solve the linear system: $\mathbf{V}^\pi = \gamma \mathbf{P}_\pi \mathbf{V}^\pi + \mathbf{r}$ for $\gamma = 0.99$. If you solve the problem by back-substitution, you should show your work. If you solve the problem by inverting a matrix, you should show the matrix that you inverted, but you don't need to invert the matrix yourself - you can give it to a solver. Write down your answers in a table like the one shown below. (1 point)

State	Value
$\langle 1, 1 \rangle$	
$\langle 1, 2 \rangle$	
$\langle 2, 1 \rangle$	
$\langle 2, 2 \rangle$	

12 Policy Iteration (4 points)

Assume that we are doing policy iteration starting with the assumptions and value function you computed in the previous section.

(a) Write down the greedy policy given this value function, using U, D, R or L to indicate the action using a a table like the one below (1 point):

State	Action
$\langle 1, 1 \rangle$	
$\langle 1, 2 \rangle$	
$\langle 2, 1 \rangle$	
$\langle 2, 2 \rangle$	

(b) Write down the transition probabilities for this policy using a table like the one below (1 point):

From	To	Probability
$\langle 1, 1\rangle$	$\langle 1, 1\rangle$	
$\langle 1, 1\rangle$	$\langle 1, 2\rangle$	
$\langle 1, 1\rangle$	$\langle 2, 1\rangle$	
$\langle 1, 2\rangle$	$\langle 1, 2\rangle$	
$\langle 1, 2\rangle$	$\langle 1, 1\rangle$	
$\langle 1, 2\rangle$	$\langle 2, 2\rangle$	
$\langle 2, 1\rangle$	$\langle 2, 1\rangle$	
$\langle 2, 1\rangle$	$\langle 1, 1\rangle$	
$\langle 2, 1\rangle$	$\langle 3, 1\rangle$	
$\langle 2, 1\rangle$	$\langle 2, 2\rangle$	
$\langle 2, 2\rangle$	$\langle 2, 2\rangle$	
$\langle 2, 2\rangle$	$\langle 2, 1\rangle$	
$\langle 2, 2\rangle$	$\langle 1, 2\rangle$	
$\langle 2, 2\rangle$	$\langle 3, 2\rangle$	

(c) Solve for the value function for this policy. If you solve the problem by back-substitution, you should show your work. If you solve the problem by inverting a matrix, you should show the matrix that you inverted, but you don't need to invert the matrix yourself - you can give it to a solver. Write down the values using a table like the one below (1 point).

State	Value
$\langle 1, 1 \rangle$	
$\langle 1, 2 \rangle$	
$\langle 2, 1 \rangle$	
$\langle 2, 2 \rangle$	

(d) Is this policy optimal? Provide a mathematical justification for your answer (1 point).