COMPSCI 570 **Instructor:** Ronald Parr

Homework 5 **Due:** Friday December 4, 2025

1 Value of Information Calculation (40 points)

For this problem, we will assume that we can measure utility in units of *utils* and that we have somehow figured out how to convert things like health and money into utils. Assume that there are two mutually exclusive diseases, $\{d_1, d_2\}$, and two medications $\{m_1, m_2\}$. We are considering how to treat a patient who has a 50% probability of having each of these diseases. Possible outcomes of treating the patient are h_1, which corresponds to a healthy patient, or h_2 which corresponds to a significant worsening of the patient's condition. $U(h_1) = 100$. $U(h_2) = -1000$.

The cost of medication m_1 is 50 utils. m_1 is 100% effective on disease 2, i.e., $P(h_1|m_1, d_2) = 1.0$. However, it is only 50% effective on disease 1: $P(h_1|m_1, d_1) = P(h_2|m_1, d_1) = 0.5$.

The cost of medication m_2 is 500 utils, but it is 100% effective against both m_2 and m_1: $P(h_1|m_2, d_1) = P(h_1|m_2, d_2) = 1.0$.

Optimal Strategy with No Additional Information (10 points)

Considering just a single stage of treatment (just one chance to administer medication), compute the optimal medication to give the patient. Your answer should take into account the utility of the outcomes, as well as the cost of the medications.

Value of Perfect Information (15 points)

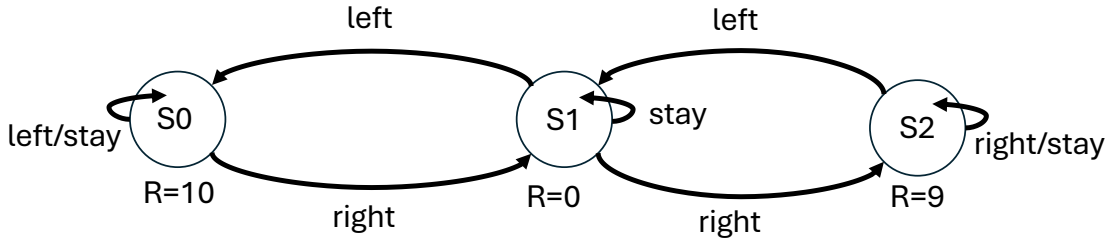
Suppose there is a test R with two outcomes r_1 and r_2. This test is 100% effective in revealing the true disease the patient has, so $P(r_1|d_1) = P(r_2|d_2) = 1$. What is the value of information for this test, i.e., the maximum amount you'd be willing to pay for the test in utils before the expected utility becomes negative.

Imperfect Information (15 points)

Now consider the case where the test is not perfectly reliable, so $P(r_1|d_1) = P(r_2|d_2) = X$, $X \geq 0.5$. Solve for the largest value of X where the value of knowing R becomes 0.

2 MDPs (30 points)

Consider the 3-state MDP shown in the image below.



Each state has 3 possible actions, left, right, and stay. All actions are deterministic, so arcs are not annotated with probabilities. We will refer to polices using a sequence of three letters. For example, the SRL, would correspond to taking stay action in state 0, the right action in state 1, and the left action in state 2.

a) Using a discount factor of $\gamma = 0.95$, and an initial value function that is all zeros, show three steps of value iteration. (More specifically: Start with $V_0(S0) = V_0(S1) = V_0(S2) = 0$, and show V_1, V_2, and V_3. (9 points)

b) Solve for the exact value of the policy SLL with discount factor $\gamma = 0.95$. (6 points)

c) Observe that left and stay are basically the same action in $S0$, and right and stay are basically the same action in $S2$. Assuming that ties are always broken in favor of the stay action, prove the SLL is the optimal policy for $\gamma = 0.95$. Hint: You can use the termination condition for policy iteration. (8 points)

d) It's easy to see that for any discount factor, the optimal action in $S0$ will always be S (equivalent to L) and that the optimal policy in $S1$ will always be L. Solve for the value of γ where the optimal policy for $S2$ becomes L. (7 points)

3 MDP Theory (30 points)

Consider an MDP, with rewards that are defined over states only, and with all rewards non-negative. Suppose the optimal policy under discount $0 < \gamma_1 < 1$ is π_1^* with value function V_1^*. Now consider the optimal policy under $\gamma_1 < \gamma_2 < 1$, and call that π_2^* with value function V_2^*.

Prove $\forall_s : V_1^*(s) \leq V_2^*(s)$.