In this assignment, you solve problems for Markov Decision Processes (MDP) - Average. You need to formulate the mathematical model and solve it with Python and Gurobi.

1. **Submission Instructions**

Submit a PDF file describing a Markov Decision Processes (MDP) model and reporting the solution to the problem instance. Also, submit a program (a Python script or a Jupyter notebook) using Gurobi to solve the problem instance. In the MDP formulation, clearly define the stages, states, and actions and formulate the Bellman equations. In the Linear Programming (LP) formulation, clearly define decision variables and state the objective function and constraints. For the problem instance, report the values of the objective and solution.

2. **Problem**

A patient is suffering from a chronic disease that is non-lethal but incurable. Each stage not only impacts the patient's quality of life but also incurs increasing costs for symptom relief per cycle: $100 for mild (M), $300 for intermediate (I), and $500 for severe (S). To manage this condition, two treatment options are available: basic medication (B) and advanced medication (A). Basic one costs $100 per cycle, while the advanced costs $400 but is more effective.

Treatment Effects:

- **Mild Stage:**
  Basic Medication: 80% chance of remaining mild, 20% chance of worsening to intermediate.
  Advanced Medication: 90% chance of remaining mild, 10% chance of worsening to intermediate.

- **Intermediate Stage:**
  Basic Medication: 60% chance of staying intermediate, 40% chance of worsening to severe.
  Advanced Medication: 70% chance of staying intermediate, 20% chance of improving to mild, 10% chance of worsening to severe.

- **Severe Stage:**
  Basic Medication: 100% chance of remaining severe.
  Advanced Medication: 80% chance of staying severe, 15% chance of improving to intermediate, and 5% chance of improving to mild.

Question 1: draw the State Transition Diagram for MDP, formulate an Average Reward MDP as LP and then solve it using Python and Gurobi.

Question 2: formulate the Bellman equations for Discounted Reward MDP (discount factor is 0.9), perform value iteration with Python (100 iterations to converge) and report the optimal policy and value. Note that the solution to Problem 2 is NOT necessarily the same as Problem 1.