Engineering and Applied Science Programs for Professionals
Whiting School of Engineering
Johns Hopkins University
705.625 Agentic AI
Homework 2
Assigned with Module 4
Due at the end of Module 6

Total Points 100/100

Submission Instructions and Problem Selection:

While student discussions, sharing of recommendations, and some collaboration are encouraged for team building and peer learning, **each student is required to submit their own complete programming solutions.** Every submission must demonstrate the individual student's ability to design, document, implement in code, and test their work independently.

Grading Criteria (100% for each problem)

Completeness & Problem Coverage (20%): Implement all required parts (pseudocode, code, performance analysis, outputs) with nothing missing or vague.

Writing Quality, Technical Accuracy & Justification (20%): Explanations are clear, concise, and technically correct; design choices and conclusions are justified with sound reasoning.

Quantitative Work: Assumptions, Derivations & Calculations (20%): List assumptions up front; show derivations or intermediate formulas/metrics; present final results correctly and clearly.

Code Quality, Documentation & Execution (20%): The notebook runs end-to-end without errors; names are clear; formatting is consistent; comments explain key logic; the work is organized and reproducible.

Examples, Test Cases & Visuals (10%): Include labeled test cases and informative figures/tables with titles, captions, and axes; use appropriate metrics (e.g., precision, recall, F1, ROC/PR curves) rather than accuracy alone [1].

Notebook README & Reproducibility (10%): Provide Python version, package list with install steps, dataset details and download instructions, and step-by-step run instructions; use relative paths and fixed random seeds.

1. **Environments and Rationality — Modules 4, and 5**

   20 points total (Each subpart is worth 5 points each).

   Problem Statement:

   A large restaurant chain (agent 1) is considering entering a new city where a local restaurant (agent 2) currently operates. The sequence of play is:

   I. The chain decides whether to Enter or Stay Out.

   II. If the chain enters, the local restaurant must decide whether to Compete Aggressively or Accommodate.

   Payoffs (Chain, Local):

   - If Stay Out: (0, 10)
   - If Enter → Accommodate: (8, 5)
   - If Enter → Compete Aggressively: (-2, 2)

   Questions:

   (a) Draw the extensive-form game tree with proper notation.

   (b) Use backward induction to find the subgame perfect equilibrium.

   (c) Explain why the local restaurant's threat to "compete aggressively if you enter" is not credible.

   (d) What would be the outcome if this were played as a simultaneous game instead?

2. **Decision-Making in Agents — Modules 5**

   25 points total (Each subpart is worth 5 points each).

   Problem Statement:

   Gru and Vector must decide when to launch competing products. Gru moves first and chooses Early Launch or Wait. Vector observes Gru's decision and then chooses its timing. Market conditions create the following payoff structure:

   - If both launch early: intense competition reduces profits.
   - If one launches early and the other waits: first-mover advantage.
   - If both wait: market opportunity diminishes.

   Questions:

   (a) Design a payoff matrix that captures the strategic structure described above.

   (b) Solve for the subgame perfect equilibrium using backward induction.

   (c) Compare this SPE outcome to what would happen in a simultaneous-move version of the game.

   (d) Explain how the concept of sequential rationality prevents Gru from making "empty threats" about its launch timing.

   (e) Discuss why this game structure might lead to inefficient market outcomes despite rational play.

3. **Learning and Adaptation — Modules 4, 5, and 6**

   25 points total (Each subpart is worth 5 points each).

   Problem Statement: In traditional economic theory, more information is generally considered beneficial for decision-makers. However, game theory reveals situations where having more information can actually make players worse off. Consider the following statement:

   "*In sequential games, the player who moves second (and thus has more information) always has an advantage over the player who moves first.*"

   Questions:

   (a) Theoretical Analysis: Is this statement true or false? Provide a clear theoretical argument explaining why, drawing on concepts of sequential rationality and backward induction.

   (b) Counterexample Construction: Create a simple 2-player extensive-form game where the first-mover actually benefits from moving first, despite the second-mover having more information. Explain the payoff structure that makes this possible.

   (c) The Commitment Value: Explain how the concept of "commitment" relates to first-mover advantage. Why might players sometimes benefit from limiting their own future options?

   (d) Strategic Implications: Discuss how this paradox applies to real-world strategic situations. When might someone deliberately choose to act first or "burn bridges" to gain strategic advantage?

   (e) Information vs. Flexibility Trade-off: Analyze the fundamental trade-off between having more information (observing opponent's moves) versus having more strategic flexibility (being able to commit to actions). Under what conditions does each advantage dominate?

4. **Decision Making in Agents — Modules 4, 5, and 6**

30 points total (Each subpart is worth 6 points each).

Problem Statement: Consider an agent operating in a **partially observable, stochastic environment** with a finite horizon of $T = 3$ time steps and a discount factor $\gamma = 0.9$. At each time step $t \in \{0, 1, 2\}$, the agent may choose one of two actions:

- **Action A**: deterministic reward $R_t(A) = 3$.
- **Action B**: stochastic reward

$$R_t(B) = \begin{cases} 10 & \text{with probability } 0.2, \\ 0 & \text{with probability } 0.8, \end{cases}$$

independently at each step.

Define the total discounted return for a trajectory $\tau = (a_0, r_0, a_1, r_1, a_2, r_2)$ as

$$U(\tau) = \sum_{t=0}^{2} \gamma^t r_t.$$

Questions:

(a) Construct the decision tree for both policies π_A (always choose Action A) and π_B (always choose Action B), labeling rewards and leaf utilities.

(b) Derive the expected utility $\mathrm{E}[U \mid \pi_A]$ and $\mathrm{E}[U \mid \pi_B]$ using the value-function recursion

$$V_t^\pi = \bar{r}^\pi + \gamma V_{t+1}^\pi, \qquad V_3^\pi = 0,$$

where $\bar{r}^\pi$ is the expected one-step reward under policy π.

(c) Determine which policy is **rational** under the expected utility criterion.

(d) Suppose instead the agent is **boundedly rational** and maximizes only the *immediate expected reward* at each step (ignoring horizon and discounting). Which action would it choose?

(e) (Optional extension) Compute the variance of the total return under π_A and π_B. Discuss how a **risk-averse agent** might evaluate the two policies.