Summer of Science: Plan of Action
(Revised after midterm)

Tanay Jha
24B1040

May 15, 2025

Week 1: Rigorous Foundations - Logic, Automata, and Computability Links

Status: Completed **Topics Covered:**

- **Propositional Logic:**
  - Syntax, semantics, truth assignments.
  - Logical connectives, truth tables, normal forms (CNF, DNF).
  - Tautologies, contradictions, satisfiability (SAT problem intro).
  - (Optional: Logical entailment - conceptual understanding).
- **(Core Focus: Finite Automata)**
  - Deterministic and Nondeterministic Finite Automata (DFA/NFA): formal defini tions, transition functions, language acceptance.
  - Equivalence of DFA and NFA (understand constructive proof idea).
  - Regular expressions: syntax, semantics, Kleene's Theorem (equivalence with FA - understand the proof sketch conceptually).
  - (Optional: Pumping Lemma for non-regular languages - basic idea).
- **Connections:** Discuss how these formalisms (especially FA) model aspects of comput a tion, state, and transitions, laying groundwork for sequential decision-making.

Week 2: Stochastic Processes - Markov Chains In-Depth

Status: Completed **Topics Covered:**

- **Markov Chains (MCs):**
  - Formal definition, state space (focus on discrete time), transition matrix/kernel.
  - Chapman-Kolmogorov equations.

  - N-step transition probabilities.

- **Classification of States:**
  - Accessibility, communicating classes.
  - Recurrence (positive/null), transience, periodicity.
  - Irreducible MCs, aperiodic MCs, ergodic MCs.

- **Long-Term Behavior:**
  - Limiting distributions, stationary distributions: existence and uniqueness (conditions like ergodicity).
  - Convergence to stationary distribution.
  - (Optional: First passage times, mean recurrence times - basic concepts).

- **Absorbing Markov Chains:**
  - Canonical form, fundamental matrix, absorption probabilities.

- **Implementation:** Simulate a few small MCs, compute stationary distributions for simple examples.

Week 3: Markov Decision Processes - Core Concepts (and K-Armed Bandits)

Status: Completed **Topics Covered:**

- **MDP Components:** Thorough understanding of states, actions, transition probabilities (model dynamics $P(s'|s,a)$), rewards $R(s,a,s')$.

- **Policies and Value Functions:**
  - Deterministic and stochastic policies ($\pi(a|s)$).
  - State-value function $V^\pi(s)$, Action-value function $Q^\pi(s,a)$.

- **Bellman Equations:**
  - Bellman expectation equation for V^π and Q^π (understand the equations).
  - Bellman optimality equation for V^* and Q^* (understand the equations).
  - (Optional: Bellman operators (T^π, T^*) and properties like contraction - conceptual overview).

- **K-Armed Bandits (Implementation Completed):**
  - Problem definition, exploration-exploitation dilemma.
  - Implemented and analyzed Greedy, ϵ-Greedy, UCB, Gradient Bandit, and Optimistic Initial Values algorithms.

- **Focus:** Deeply understand the components and the meaning of Bellman equations. Defer detailed algorithm study and implementation for MDP solution methods to later weeks initially, though bandit implementations provided foundational RL experience.

Week 4: Buffer Week & Midterm Report (Preparation for Mid June Submission)

Status: Completed **Activities:**

- Consolidate understanding of Weeks 1-3, focusing on theoretical concepts.
- Review notes and clarify any doubts on Logic, Automata, MCs, K-Armed Bandits, and MDP fundamentals.
- Midterm Report preparation and review.

Week 5: Formulating RL Problems & Intro to MDP Solutions (Midterm Submission: Mid June)

Status: In Progress (DP Solutions Implemented) **Topics Activities:**

- Midterm Report Submission (Target: Mid June)
- **Introduction to Exact MDP Solution Algorithms (Implementation Completed):**
  - Value Iteration (VI): Understand the algorithm conceptually and implement for solving the Gambler's Problem.
  - Policy Iteration (PI): Understand the algorithm conceptually (policy evaluation, pol icy improvement) and implement for solving the Grid World problem.
- **Introduction to Hidden Markov Models (HMMs) (Upcoming):**
  - Definition, key problems (filtering, smoothing, decoding) - brief.
  - Contrast with observable MCs and fully observable MDPs.
- **Reward Engineering (Upcoming):**
  - Principles of good reward design.
  - Sparse vs. dense rewards.
  - (Optional: Common pitfalls and unintended consequences - brief discussion).
- **Problem Formulation for RL (Upcoming):**
  - Episodic vs. Continuing tasks.
  - Horizon: Finite, infinite.
  - Discounting factor (γ): role, interpretation.

Week 6: Model-Free Reinforcement Learning - Core Algorithms (Upcoming)

Topics:

- **Monte Carlo (MC) Methods:**
  - First-visit vs. Every-visit MC prediction.
  - MC control (exploring starts, on-policy ϵ-soft).
  - (Optional: Off-policy MC control using importance sampling - conceptual overview).
- **Temporal Difference (TD) Learning:**
  - TD(0) prediction.
  - Advantages of TD over MC.
  - SARSA (On-policy TD control): Algorithm, convergence properties (conceptual).
  - Q-Learning (Off-policy TD control): Algorithm, convergence proof sketch (concep tual). Difference from SARSA.
- **Exploration vs. Exploitation:**
  - ϵ-greedy, ϵ-decreasing strategies.
  - (Optional: Optimistic initialization, UCB - brief mention).
- **Implementation:** Implement Q-learning and SARSA for simple grid-world environments or a Gym environment like FrozenLake. Experiment with exploration.

Week 7: Advanced RL - Function Approximation & Policy Gradients (Upcoming)

Topics:

- **Function Approximation in RL:**
  - Value function approximation: $V(s, \mathbf{w}) \approx V^{\pi}(s)$, $Q(s, a, \mathbf{w}) \approx Q^{\pi}(s, a)$.
  - Linear function approximation: features, gradient descent methods (SGD).
  - (Optional: Semi-gradient TD(0), Semi-gradient SARSA - conceptual.)
- **Deep Q-Networks (DQN) - Introduction:**
  - Basic Architecture using Neural Networks.
  - Experience Replay.
  - Target Networks.
- **Policy Gradient Methods - Introduction:**

  - Policy approximation $\pi(a|s, \theta)$.
  - Policy Gradient Theorem (understand derivation idea).
  - REINFORCE algorithm (Monte Carlo Policy Gradient).
  - (Optional: REINFORCE with Baseline - conceptual understanding).
- **Applications (Conceptual Overview):**
  - Brief overview of successes like AlphaGo/AlphaZero.
  - Challenges in robotics (continuous spaces, sim-to-real).

Week 8: Buffer Week, Project Completion, and Final Report Submission (Endterm: Mid July) (Upcoming)

Activities:

- Intensive work on the coding project (e.g., Flappy Bird with tabular Q-learning, poten tially simplified DQN).
- Debugging, experimentation, hyperparameter tuning for the project.
- **Final Report (Target: Mid July):** Comprehensive document detailing:
  - Theoretical understanding (based on covered topics).
  - Project design and implementation details.
  - Experimental results (learning curves, performance metrics).
  - Challenges encountered and future work.
- Prepare a short presentation of your project if required.

Main Coding Project: Mini Game with RL Agent (Upcoming)

Description:

- **Game:** Create or adapt a simplified version of Flappy Bird (or a similar simple game like CartPole from Gym if preferred for easier setup).
- **Agent Baseline:** Implement with tabular Q-learning (discretized state space if needed for custom game).
- **Agent Advanced (Stretch Goal):** Implement with Deep Q-Learning (DQN) using a simple neural network (e.g., PyTorch or TensorFlow/Keras). Focus on getting a basic DQN working rather than extensive optimization.
- **Experimentation:**
  - Compare performance and learning speed of tabular Q-learning (vs. DQN if imple mented).

  - If DQN is implemented: Analyze the effect of learning rate, exploration strategy. (Keep network architecture simple).

- **Visualization:** Plot learning curves (e.g., rewards per episode). Demonstrate the agent improving over episodes.

Additional Optional Mini-Projects (to reinforce weekly concepts, if time per mits):

- **Grid World Solver (Value Iteration & Policy Iteration): Status: Completed.** (Implement VI and PI for a simple grid world).
- **Markov Chain Analyzer (Upcoming/Review):** Implement functions to calculate N-step transition proba bilities and identify communicating classes for a given MC. (Week 2 material).
- **Tabular Q-Learning/SARSA on Gym's FrozenLake/CliffWalking (Upcoming):** Revisit and solidify understanding. (Week 6).

References

This list includes primary textbooks and lecture series that will be consulted. Specific research papers or supplementary materials may be added as the SoS progresses.

For Week 1 Topics (Logic, Automata):

- Huth, M., & Ryan, M. (2004). *Logic in Computer Science: Modelling and Reasoning about Systems* (2nd ed.). Cambridge University Press. (Filename: 'huth and ryan.pdf')
- Baier, C., & Katoen, J.-P. (2008). *Principles of Model Checking*. MIT Press. (Filename: 'baier-katoen.pdf') (Relevant chapters for automata and foundational concepts)

For MDPs (before midsem):

- Baier, C., & Katoen, J.-P. (2008). *Principles of Model Checking*. MIT Press. (Primary reference for MDPs)
- Parker, D. Slides on Model Checking (PRISM). Available: https://www.prismmodelchecker.org/lectures/pmc/
- Relevant lectures from David Silver's UCL Course on Reinforcement Learning (early lec tures covering MDPs). YouTube Playlist: https://youtube.com/playlist?list=PLqYmG7hTraZDVH599E...

For RL part from Week 4 onwards:

- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press. (Standard reference)

- David Silver's UCL Course on Reinforcement Learning. YouTube Playlist: https://youtube.com/playlist?list=PLqYmG7hTraZDVH599EltlEWSuOSJbAodm&si=S1Bu6CvQAlYAY38
- NPTEL Course on Reinforcement Learning by Prof. Balaraman Ravindran (IIT Madras). YouTube Playlist: https://youtube.com/playlist?list=PLWbRJQ4m4UjJnymuBM9Rdm si=STuLazJq29Cbi9_H