ANALYSIS REPORT - MACHINE LEARNING

COURSE: Machine learning (UE23CS352A)

TEAM 2 ‘A’ SEC:

Aaron TM (PES1UG23AM005)

Preetham (PES1UG23AM913)

Aarya Upadhya (PES1UG23AM006)

Aarav Adarsh (PES1UG23AM003)

DATE: 03-11-2025

Key challenges:

- The initial HMM on the test set was 33.5%-win rate suggest that relying on statistical model patterns from the corpus has its limits
- The training curves for the initial RL agent show fluctuations in average reward and win rate indicating potential instability during training
- The parameter tuning for the HMM demonstrated that the win rate is sensitive to smoothing alpha, signal weights, and bin boundaries
- Similliar The RL agent is sensitive to the learning rate, epsilon decay, and batch size etc

Insights gained:

- HMM are effective at capturing basic letter patterns but lack the ability to adapt when the environment dynamic state changes or when complex sequential targets are present
- RL can learn these strategies based on the reward system and gameplay experience
- Combining these both is a promising approach as it gives more of a probabilistic approach than a random start that is usually seen in Reinforcement learning

Strategies:

- HMM Design choices:
    - Dividing words into length bins allowed the HMM to learn pattern specific to the length in those ranges
    - Learning start and ending transition and, position specific and overall frequency captures the word structure
    - Laplacian smoothing for unseen letters
- RL Design Choices:
    - The includes masked_word, guessed_letters, lives_left, word_length, num_blanks, and num_guessed. This captures the essential information needed to understand the current game situation. Including num_blanks and num_guessed provides a simple measure of game progress

Exploration vs Exploitations

- Epsilon: The agent uses the epsilon greedy approach to balance exploration and exploitation

  - Exploitation:
    - The agent chooses an action randomly; this allows the agent to discover new better strategies or letter combos that it has not tried frequently before.
  - Exploitation
    - The agent chooses that action this predicted to have the highest Q value based on its current Neural Network. This leverages the knowledge what it already has learn (learns the best moves that gives the best output)
- The epsilon decays over time so that exploration is high at initial epochs and then the exploration decrease and model uses moves that learn from the neural network (Exploitation)

Future Improvements:

- More Advanced State Representation
- Vectorization and Word embeddings that can be passed into a LSTM etc
- Better RL algorithms such as: Double DDQN. PP0 etc
- Ensambling and Better Hyperparameter Tuning could've been tried