Hidden Markov Model Integration for F1 Racing Line Optimization

Project Context

My capstone project involves training an RL agent to discover optimal racing lines on F1 tracks and comparing these AI-generated paths against actual driver data. The integration of Hidden Markov Models adds a probabilistic layer to understand driving behavior patterns and racing strategies.

1. Observations - Measurable Data

The HMM would observe the following sequential data points collected at regular time intervals during racing:

- **Speed measurements** (km/h) at track positions
- **Steering angle inputs** (-1 to 1 normalized)
- **Throttle/brake positions** (0 to 1 normalized)
- **Track sector information** (straight, corner entry, apex, exit)
- **Lateral G-forces** indicating cornering intensity
- **Distance to track boundaries** (left/right edges)

These observations form time-series sequences representing driving behavior throughout lap segments.

2. Type of HMM Problem

This represents an **unsupervised learning problem** since we don't know the hidden driving states in advance. We want the model to discover latent driving modes (aggressive cornering, fuel-saving, defensive positioning, etc.) from observed telemetry data without predefined labels.

3. Training Algorithm

a. Known Values at Start

- **Observation sequences**: Telemetry data from F1 drivers and RL agent
- **Number of hidden states**: Initially estimated (e.g., 5-8 driving modes)
- **Observation dimensionality**: 6 features per time step

b. Unknown Values to Learn

- **Hidden state meanings**: What each latent state represents behaviorally
- **Transition probabilities**: How drivers switch between driving modes
- **Emission probabilities**: How each hidden state generates observable telemetry
- **Initial state distribution**: Starting probability for each driving mode

4. Parameter Updates

The **Expectation-Maximization (EM) algorithm** will iteratively update:

- **Transition matrix (A)**: Probabilities of switching between hidden driving states
- **Emission parameters (B)**: Mean and covariance matrices for Gaussian emissions from each state
- **Initial state probabilities (π)**: Likelihood of starting in each driving mode

Application to Racing Line Optimization

The trained HMM will:

1. **Identify distinct driving strategies** from F1 driver telemetry
2. **Classify RL agent behavior** into human-like or novel driving modes
3. **Predict optimal state transitions** for different track sections
4. **Enable strategy comparison** between AI and human approaches

This probabilistic framework adds interpretability to the RL agent's decisions while providing insights into when and why certain racing lines emerge, ultimately contributing to more energy-efficient and safer transportation systems through motorsports innovation.