Homework 4

Manyara Bonface Baraka - mbaraka

April 21, 2025

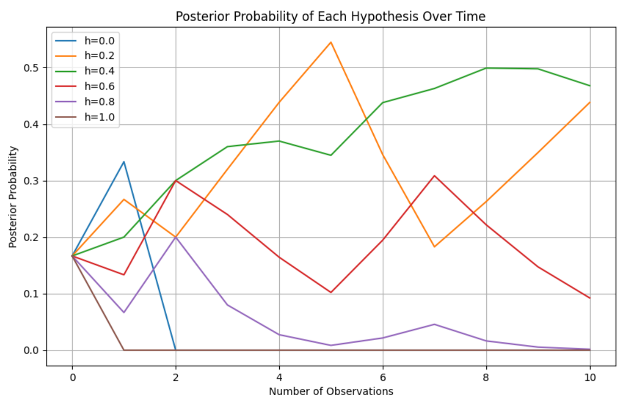
Problem 1: Bayesian Learning

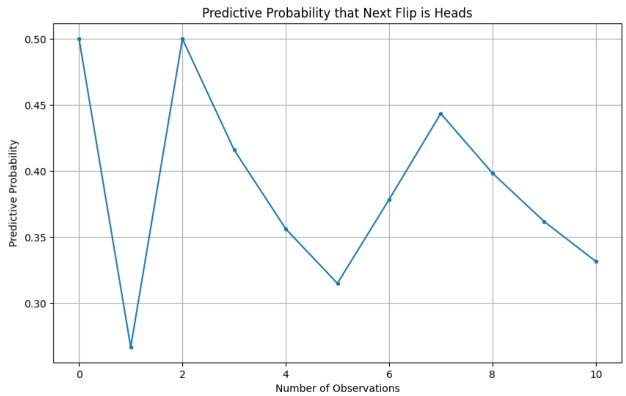
Bayesian Learning calculates the probability of each hypothesis given the data, and makes predictions by weighing all hypothesis by their probabilities

- Will asssumed eaqual prior start for each hypothesis $P(h_i) = \frac{1}{6}$ for all i

Observations

Plots



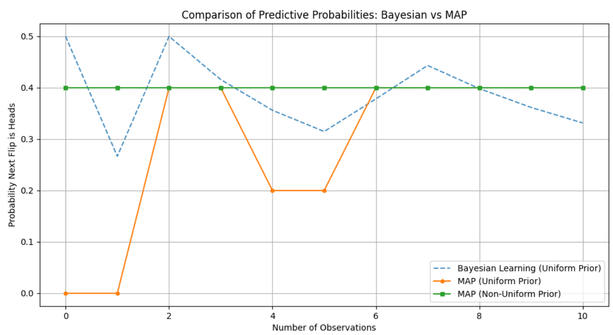


Most likely hypothesis after all observations: h=0.4

Insights

The first plot: Posterior Probability plot: the plot shows the evolution of the posterior probabilities over time for each hypothesis. At the start all hypthesis have the same prior however as the observation came in the posterior changes am having the likely of h=0.4.
The second plot: Predictive Probabilit plot: the second plot depicts teh predictive probability that the next flip will be heads, calculate after each observation. The model flactuates at first but it tends to stabilize around 0.4. The highest posterior probabily influences the dominance of the hypthesis on predictions. The predictive probability tends to be stable aroung 0.4 since the posterior of h3 becaomes quite large.

Problem 2: Maximum a Posteriori (MAP) Estimation



Observation and Insights

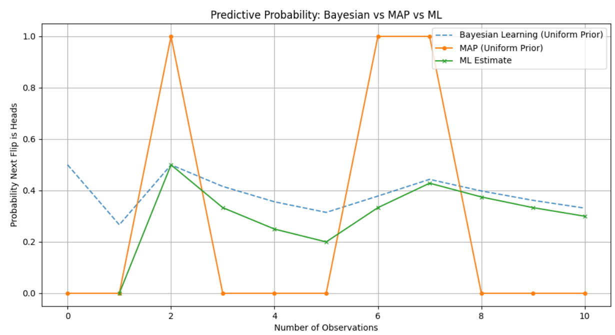
MAP vs Bayesian

- Smoothness: Bayesian with the blue dotted it smoothly converge to the true bias by averagin all hypthesis however, the MAP prediction jumps discretely as the most probable hypothesis changes.

- Prior sensistivity: With the uniform prior MAP predictions astarts at 0 due to limitation in tie-breaking but it adjusts faster as the data comes in. While with non-uniform prior is biased MAP initially predicts 0.4 and its constant for that.

Generally the MAP is computational efficinte however its prone to abrupt changes and prior bias, especially with limited data. It trades off accuracy for simplicity while Bayesian provides a robust predictions by combining uncertinity but requires more computation.

Problem 3: Maximum Likelihood (ML) Estimation



- Is the estimate equal to the estimates we found as a result of Bayesian Learning and MAP estimation? What is the major difference?
- Which is likely to be more accurate: the estimate found with ML estimation or the estimate found using the other methods?
- In what case (specifically, for what prior distribution) does MAP estimation reduce to the ML estimate? Feel free to consult Section 20.2 Learning with Complete Data in Russel & Norvig.
- Provide a brief discussion comparing Bayesian Learning, MAP, and ML estimation. What are the pros and cons of each?