

8 Machine Learning and Real-world Data (AAC)

- (a) A team plays several seasons with 20 matches in each season. If the team is at home, the probability of winning is 0.7. The probability of winning away is 0.4. No games are draws. Home and away games usually alternate: ignoring the end of the season, there is a probability of 0.75 that a home game is followed by an away game, and also a probability of 0.75 that an away game is followed by a home game. At the start of the season, there is an equal chance of a home or an away game. Show how this information can be partially modelled as a Hidden Markov Model (HMM), treating home and away as hidden states. Give the parameters of the HMM. [6 marks]
- (b) What aspect of the scenario described is not correctly covered by the HMM? [1 mark]
- (c) If the team's results are 'win, lose, win' at the start of the season, what probability estimate for 'home, away, home' is given by this HMM? [2 marks]
- (d) You are given a complete record of individual games, including a record of the opponents. The win/loss ratio varies depending on the opponent. Explain how you could use such information to derive the parameters of a more complex HMM (treating home and away as hidden states, as before). [4 marks]
- (e) It is suggested that you could use an HMM to predict the results of next season's games, since it is known who the opponent will be and whether the match will be at home or away. How might you do this? How successful do you think this would be, compared to predicting whether a sequence of matches were home or away based on a sequence of match results? [7 marks]