COMPUTER SCIENCE TRIPOS Part IA – EXAMPLE – Paper 3

8   Machine Learning and Real-world Data (AAC)

This question concerns a variant of the fraudulent croupier scenario discussed in the course. In this version, the croupier has three six-sided dice F, L1 and L2 where F is fair and L1 and L2 are both loaded. The croupier secretly switches between dice. You have some training data sequences, of different lengths, with the observed dice rolls paired with the identifier of the dice.

(a) Assuming a first-order HMM, show what the two HMM probability matrices would look like for this scenario, explaining the notation you use.   [6 marks]

(b) Draw a state diagram, labelling the transitions with sample terms from the matrices to illustrate the different types of probabilities involved. (i.e., you do not have to label all the arcs, and should avoid doing so because it will make the diagram difficult to read.)   [3 marks]

(c) Suppose the the croupier is acting according to certain rules which determine when the dice is switched. For each of the following rules, describe the effect in terms of probabilities and training data counts, and discuss whether the behaviour would be modelled by an HMM.

   (i)   The croupier never switches directly from F to L2.   [2 marks]

   (ii) The croupier knows in advance how many dice throws the sequence will contain and makes sure that the dice is always F on the last roll.

   [3 marks]

   (iii) The croupier never rolls L2 more than twice in a row.   [3 marks]

   (iv) The croupier always switches dice after rolling a 6.   [3 marks]