8 Machine Learning and Real-world Data (av308)

A farmer has been collecting data about the performance of a crop. Every year, one of three methods is used for cultivation: fertilizer (F), pesticide (P) or none (N). The farmer characterizes the performance of the crop as good (G), average (A) or bad (B). The following data has been collected:

\[
\begin{array}{cccccccc}
\end{array}
\]

The farmer wants to apply a Hidden Markov Model (HMM) to understand the relation between cultivation methods and crop performance, treating the cultivation method as the hidden states and the crop performance as the observations.

(a) Describe the components of the first order HMM using equations where appropriate. For the parameters you define, also describe how you would estimate them from the data given above. [7 marks]

(b) What are the assumptions made in the first order HMM? What do they mean in the context of the farming application described above? [4 marks]

(c) Given the HMM and the farming data above, answer the following questions, explaining your answers. [6 marks]

(i) Which cultivation method is more likely to result in a good crop?

(ii) Which succession of cultivation methods is more likely to result in a succession of two good crop years? Assume that for the first of the two years the choice of cultivation method is uniform.

(iii) Is a bad crop year more likely to be followed by an average or a good crop year?

(d) By this point you have solved the labelled learning problem for the HMM. Name the other three problems described below: [3 marks]

(i) Given the observations about crop performances and knowledge of the possible cultivation options, find the parameters of the HMM.

(ii) Given the parameters of the HMM and a sequence of crop performances, determine how likely the sequence of performances is.

(iii) Given the parameters of the HMM and a sequence of crop performances, infer the most likely sequence of cultivation methods used.