

8 Machine Learning and Real-world Data (av308)

You are interested in predicting snowfall in order to plan for a winter trip with friends. You know that for snowfall to occur, we need to have freezing air temperatures, even if this is not always sufficient.

You decide to model this using a first order hidden Markov model (HMM), with air temperature as the hidden state (Freezing or NotFreezing) and Snowfall, Rain or Dry as the observations. Then you take a look at data from recent months and see:

| Month | Oct | Nov | Dec | Jan | Feb | Mar | Apr | May | Jun |
|---------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| AirTemp | NF | NF | F | F | F | NF | F | NF | NF |
| Weather | R | R | S | S | D | R | D | D | D |

- (a) Define and estimate the components of an appropriate HMM for this application, without smoothing. Assume that all hidden states are equally likely to start the sequence. Ignore the start and end states.

[4 marks]

- (b) Answer the following questions using the model you estimated. Provide the calculations needed to arrive in each answer:

(i) What are is most likely combination of air temperature and weather for the next month (July) ?

(ii) What is the most likely combination of air temperature and weather in three months from now (September)?

[4 marks]

- (c) We are in June, and your friends predict that July and August will be dry, followed by rain in September and snow in October. If they are right, what is the most likely sequence of air temperatures for July to October?

[6 marks]

- (d) The current model you have estimated ignores the time of the year we are in, e.g., the probability of transitioning from NotFreezing to Freezing is independent of whether we are in summer or in winter. Build a better model by taking into account the time of the year, by changing the definition of hidden states. Describe any transformations necessary to the data and estimate the parameters of the new model.

[6 marks]