T-61.5020 Statistical Natural Language Processing

Exercises 5 — Markov chains and Hidden Markov Models
Version 1.0

1. You have studied the changes of the weather. Every day 12am you have looked at the sky and wrote down whether it is sunny (S_1), cloudy (S_2) or rainy (S_3). From these observations you have calculated the following transition probabilities:

$$A = \begin{bmatrix} 0.8 & 0.15 & 0.05 \\ 0.4 & 0.5 & 0.1 \\ 0.3 & 0.3 & 0.4 \end{bmatrix}$$

   Element a_{ij} stands for transition from state S_i to state S_j. E.g. the probability that it is sunny day after a cloudy one is 0.4.

   a) Draw Markov chain for the weather.

   b) Today it is cloudy. What is the probability of the following sequence of five days: Tomorrow it rains, day after tomorrow is cloudy, and after it we have three sunny days.

   c) On average, how many successive sunny days there are?

2. You are on holidays in Hawaii when you begin to wonder what kind of weather it would be in home. The local newspapers tell only the temperature of the region. However, you can estimate the probability $P(x < -5°C \mid q = S_1) = b_1(x \leq -5°C)$, i.e. that it is over 5 degrees of frost in a sunny day. And more:

$$\begin{aligned}
b_1(x \leq -5°C) &= 0.8 \\
b_1(-5°C \leq x \leq 5°C) &= 0.05 \\
b_1(x \geq 5°C) &= 0.15 \\
b_2(x \leq -5°C) &= 0.1 \\
b_2(-5°C \leq x \leq 5°C) &= 0.7 \\
b_2(x \geq 5°C) &= 0.2 \\
b_3(x \leq -5°C) &= 0.3 \\
b_3(-5°C \leq x \leq 5°C) &= 0.4 \\
b_3(x \geq 5°C) &= 0.3
\end{aligned}$$

   When you left home, it was a sunny day there.

   a) After three days the holiday ends and it is time to fly back. You want to estimate what kind of weather it is when you come back, in order to dress appropriately. The temperatures in the home town were $7°C$, $3°C$ and $-8°C$. Calculate the probabilities of the three weather types on the day of returning. It is recommended to use the forward algorithm, brute force will be very time consuming.

b) In the airplane you speculate what kind of weather it has been in home during the holidays. Find the most probable state sequence. Viterbi algorithm uses the least number of calculations.