TDT4171 Artificial Intelligence Methods

Assignment 2 - Probabilistic Reasoning over Time

February 04, 2021

- **Delivery deadline: February 17, 2021 by 23:59 sharp**.
- Required reading for this assignment: Chapter 15 (the parts in the curriculum).
- Deliver your solution on *Blackboard*. Please upload your report as a PDF file. For the programming part, deliver the source code alongside the PDF file. Please **do not** put it into an archive.
- Students can **NOT** work in groups. Each student can only submit solution individually.
- This assignment totals 10 points. The number of points for each problem is labeled. All sub-problems count equally towards that problem's total.
- Copying ("koking") from other students is not accepted, and if detected, will lead to immediate failure of the course. The consequence will apply to both the source and the one cribbing.
- For help and questions related to the assignment, ask the student assistants during the guidance hours or use Piazza. The guidance hours and link to Piazza can be found under "Assignments" on Blackboard. For other inquires, an email can be sent to tdt4171@idi.ntnu.no.

1 Hidden Markov Model

5 points

Some tourists are curious if there are fish in a nearby lake. They are unable to observe whether this is true or not by staring into the lake. However, they can observe whether or not there are birds nearby that affect the presence of fish. Based on their instincts, the tourists propose the following domain theory:

1. The prior probability of fish nearby (that is, without any observation) is 0.5.
2. The probability of fish nearby on day t is 0.8 given there are fish nearby on day $t-1$, and 0.3 if not.
3. The probability of birds nearby on day t if there are fish nearby on the same day is 0.75, and 0.2 if not.

The following evidence is given

- $\mathbf{e}_1 = \{\text{birds nearby}\}$
- $\mathbf{e}_2 = \{\text{birds nearby}\}$
- $\mathbf{e}_3 = \{\text{no birds nearby}\}$
- $\mathbf{e}_4 = \{\text{birds nearby}\}$
- $\mathbf{e}_5 = \{\text{no birds nearby}\}$
- $\mathbf{e}_6 = \{\text{birds nearby}\}$

We will denote the state variable for fish nearby on day t by X_t.

Instructions

Use programming to solve all exercises in this section involving computation. The results need to be extracted from the program and well documented in a human-readable format that is easy to understand in a PDF file. Additionally, write a few sentences to give the results some context. The results can, for example, be plotted using Matplotlib [1] to give a more straightforward overview.

The code must be runnable without any modifications after delivery. Moreover, the code must be readable and contain comments explaining it. We recommend that Python with the package NumPy [2] be used for the programming exercises. It is not allowed to use libraries, such as Scikit-learn [3] to solve the tasks.

Problems

(a) Formulate the information given above as a hidden Markov model, and provide the complete probability tables for the model.

(b) Compute

$$\mathbf{P}(X_t|\mathbf{e}_{1:t}), \text{ for } t = 1, ..., 6. \tag{1}$$

What kind of operation is this (filtering, prediction, smoothing, likelihood of the evidence, or most likely sequence)? Describe in words what kind of information this operation provides us.

(c) Compute

$$\mathbf{P}(X_t|\mathbf{e}_{1:6}), \text{ for } t = 7, ..., 30. \tag{2}$$

What kind of operation is this (filtering, prediction, smoothing, likelihood of the evidence, or most likely sequence)? Describe in words what kind of information this operation provides us. What happens to the distribution in Equation (2) as t increases?

(d) Compute

$$\mathbf{P}(X_t|\mathbf{e}_{1:6}), \text{ for } t = 0, ..., 5. \tag{3}$$

What kind of operation is this (filtering, prediction, smoothing, likelihood of the evidence, or most likely sequence)? Describe in words what kind of information this operation provides us.

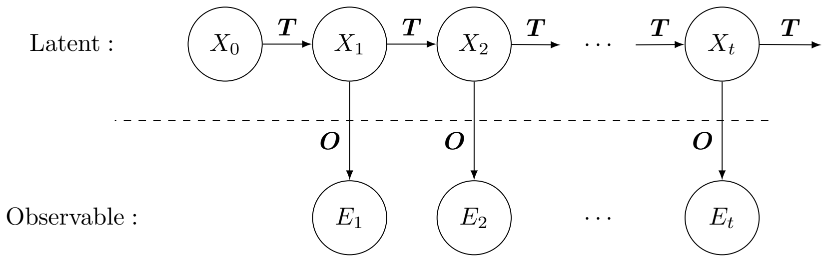
(e) Compute

$$\underset{x_1,...,x_{t-1}}{\arg\max}\, \mathbf{P}(x_1, ..., x_{t-1}, X_t|\mathbf{e}_{1:t}), \text{ for } t = 1, ..., 6. \tag{4}$$

What kind of operation is this (filtering, prediction, smoothing, likelihood of the evidence, or most likely sequence)? Describe in words what kind of information this operation provides us.

Solution

(a)



- $$P(X_0 = true) = \alpha = 0.5$$
  $$P(X_t = true|X_{t-1} = true) = \beta_1 = 0.8$$
  $$P(X_t = true|X_{t-1} = false) = \beta_0 = 0.3$$
  $$P(E_t = true|X_t = true) = \gamma_1 = 0.75$$
  $$P(E_t = true|X_t = false) = \gamma_0 = 0.2$$

- $\boldsymbol{T} = \begin{bmatrix} \beta_1 & 1-\beta_1 \\ \beta_0 & 1-\beta_0 \end{bmatrix}$

- $\boldsymbol{O}(E_t = false) = \begin{bmatrix} 1-\gamma_1 & 0 \\ 0 & 1-\gamma_0 \end{bmatrix}$

- $\boldsymbol{O}(E_t = true) = \begin{bmatrix} \gamma_1 & 0 \\ 0 & \gamma_0 \end{bmatrix}$

(b) **Filtering**

- The description of the operation needs to explain what the operation provides rather than just repeating the task. Preferably with the students' own words! See Page 570 and 571 in AIMA for more detail.

- $\mathbf{P}(X_t|\mathbf{e}_{1:t}) = c \cdot \mathbf{P}(\mathbf{e}_t|X_t) \sum_{x_{t-1}} \mathbf{P}(X_t|x_{t-1}) P(x_{t-1}|\mathbf{e}_{1:t-1})$.
- $\mathbf{f}_{1:t} = \mathbf{P}(X_t|\mathbf{e}_{1:t}) = c \cdot \text{forward}(\mathbf{f}_{1:t-1}, \mathbf{e}_t)$.
- $\mathbf{f}_{1:0} = \mathbf{P}(X_0)$.

Probabilities
The probabilities are shown in Figures 1 and 2.

(c) **Prediction**

The description of the operation needs to explain what the operation provides rather than just repeating the task. Preferably with the students' own words! See Page 570 and 571 in AIMA for more detail.

$\mathbf{P}(X_t|\mathbf{e}_{1:t-1}) = \sum_{x_{t-1}} \mathbf{P}(X_t|x_{t-1}) P(x_{t-1}|\mathbf{e}_{1:t-1})$.

The distribution converges to the stationary distribution of the Markov process defined by the transition model as the time t increases without observing new evidence.

Probabilities
The probabilities are shown in Figures 1 and 2.

(d) **Smoothing**

- The description of the operation needs to explain what the operation provides rather than just repeating the task. Preferably with the students' own words! See Page 570 and 571 in AIMA for more detail.
- $\mathbf{P}(\mathbf{e}_{k:t}|X_{k-1}) = \sum_{x_k} P(\mathbf{e}_k|x_k) P(\mathbf{e}_{k+1:t}|x_k) \mathbf{P}(x_k|X_{k-1})$.
- $\mathbf{b}_{k:t} = \mathbf{P}(\mathbf{e}_{k:t}|X_{k-1}) = \text{backward}(\mathbf{b}_{k+1:t}, \mathbf{e}_k)$.
- $\mathbf{P}(X_k|\mathbf{e}_{1:t}) = c \cdot \mathbf{f}_{1:k} \times \mathbf{b}_{k+1:t} = c \cdot \mathbf{P}(X_k|\mathbf{e}_{1:k}) \times \mathbf{P}(\mathbf{e}_{k+1:t}|X_k),\ 0 \le k < t$

Probabilities
The probabilities are shown in Figures 1 and 2.

(e) **Most likely sequence**

- The description of the operation needs to explain what the operation provides rather than just repeating the task. Preferably with the students' own words! See Page 570 and 571 in AIMA for more detail.
- $$\begin{aligned}&\arg\max_{x_1,\ldots,x_{t-1}} \mathbf{P}(x_1, ..., x_{t-1}, X_t|\mathbf{e}_{1:t}) = \\ &\mathbf{P}(\mathbf{e}_t|X_t) \arg\max_{x_{t-1}}[\mathbf{P}(X_t|x_{t-1}) \arg\max_{x_1,\ldots,x_{t-2}} P(x_1, ..., x_{t-2}, x_{t-1}|\mathbf{e}_{1:t-1})]\end{aligned}$$
- $\mathbf{m}_{1:t} = \arg\max_{x_1,\ldots,x_{t-1}} \mathbf{P}(x_1, ..., x_{t-1}, X_t|\mathbf{e}_{1:t}) = \text{forward}(\mathbf{m}_{1:t-1}, \mathbf{e}_t)$

Most likely sequences for $t = 1, ..., 6$.

$$\mathbf{m}_{1:6} = \{X_1 = true, X_2 = true, X_3 = true, X_4 = true, X_5 = true, X_6 = true\}$$
$$\mathbf{m}_{1:6} = \{X_1 = true, X_2 = true, X_3 = true, X_4 = true, X_5 = false, X_6 = false\}$$
$$\mathbf{m}_{1:5} = \{X_1 = true, X_2 = true, X_3 = true, X_4 = true, X_5 = true\}$$
$$\mathbf{m}_{1:5} = \{X_1 = true, X_2 = true, X_3 = true, X_4 = true, X_5 = false\}$$
$$\mathbf{m}_{1:4} = \{X_1 = true, X_2 = true, X_3 = true, X_4 = true\}$$
$$\mathbf{m}_{1:4} = \{X_1 = true, X_2 = true, X_3 = false, X_4 = false\}$$
$$\mathbf{m}_{1:3} = \{X_1 = true, X_2 = true, X_3 = true\}$$
$$\mathbf{m}_{1:3} = \{X_1 = true, X_2 = true, X_3 = false\}$$
$$\mathbf{m}_{1:2} = \{X_1 = true, X_2 = true\}$$
$$\mathbf{m}_{1:2} = \{X_1 = true, X_2 = false\}$$
$$\mathbf{m}_{1:1} = \{X_1 = true\}$$
$$\mathbf{m}_{1:1} = \{X_1 = false\}$$

Probabilities
The probabilities are shown in Figures 1 and 2. However, they are not an answer to the problem since the task asks for $\arg\max$.

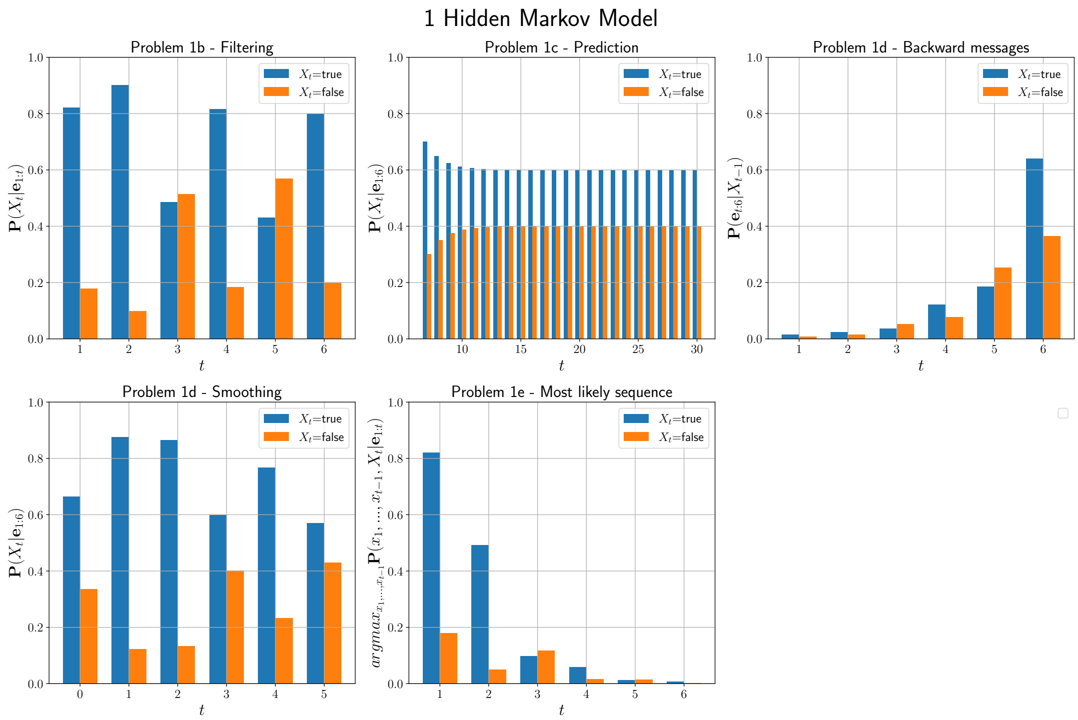


Figure 1: Problem 1e: For each time t, we have shown the values of the message $\mathbf{m}_{1:t}$, which gives the probability of the best sequence reaching each state at time t.

```
Problem 1b - Filtering
              1         2         3         4         5         6
True   0.820896  0.901971  0.485185  0.816459  0.431349  0.799709
False  0.179104  0.098029  0.514815  0.183541  0.568651  0.200291

Problem 1c - Prediction
             7         8         9        10  ...   27   28   29   30
True   0.699854  0.649927  0.624964  0.612482  ...  0.6  0.6  0.6  0.6
False  0.300146  0.350073  0.375036  0.387518  ...  0.4  0.4  0.4  0.4

[2 rows x 24 columns]

Problem 1d - Smoothing
              0         1         2         3         4         5
True   0.664852  0.876407  0.865787  0.597927  0.766637  0.570826
False  0.335148  0.123593  0.134213  0.402073  0.233363  0.429174

Problem 1e - Most likely sequence probabilities
              1         2         3         4         5         6
True   0.820896  0.492537  0.098507  0.059104  0.011821  0.007093
False  0.179104  0.049254  0.118209  0.016549  0.014185  0.001986
```

Figure 2: Showing same information as Figure 1.

2 Dynamic Bayesian Network

5 points

Some tourists visiting a cabin are interested in finding out if there are animals nearby. They can observe outside of their window every day whether there are animal tracks and whether the food they placed outside is gone. Furthermore, they believe that the animal tracks and the food placed outside are conditionally independent given animals nearby ($AnimalTracks_t \perp\!\!\!\perp FoodGone_t \mid AnimalsNearby_t$). Based on gut feelings, the tourists provide the following domain theory:

1. The prior probability of animals nearby (that is, without any observation) is 0.7.
2. The probability of animals nearby on day t is 0.8 given that there were animals nearby on day $t-1$, and 0.3 if not.
3. The probability of animal tracks on day t if there are animals nearby on the same day is 0.7, and 0.2 if not.
4. The probability of the food gone on day t if there are animals nearby on the same day is 0.3, and 0.1 if not.

The following evidence is given

- $\mathbf{e}_1 = \{\text{animal tracks, food gone}\}$
- $\mathbf{e}_2 = \{\text{no animal tracks, food gone}\}$
- $\mathbf{e}_3 = \{\text{no animal tracks, food not gone}\}$
- $\mathbf{e}_4 = \{\text{animal tracks, food not gone}\}$

We will denote the state variable for animals nearby on day t by X_t.

Instructions

Solve by hand all exercises in this section involving computation, not by programming. The results must be accompanied by steps to solve them and justifications, not only the final results.

Problems

(a) Formulate the information given above as a dynamic Bayesian network and provide the complete probability tables for the model.

(b) Compute

$$\mathbf{P}(X_t|\mathbf{e}_{1:t}), \text{ for } t = 1, 2, 3, 4. \tag{5}$$

(c) Compute

$$\mathbf{P}(X_t|\mathbf{e}_{1:4}), \text{ for } t = 5, 6, 7, 8. \tag{6}$$

(d) By forecasting further and further into the future, you should see that the probability converges towards a fixed point. Verify that

$$\lim_{t\to\infty} \mathbf{P}(X_t|\mathbf{e}_{1:4}) = \langle 0.6, 0.4 \rangle. \tag{7}$$

(e) Compute

$$\mathbf{P}(X_t|\mathbf{e}_{1:4}), \text{ for } t = 0, 1, 2, 3. \tag{8}$$

Solution

(a) The dynamic Bayesian network has three variables:

- A_t, whether there are animals.
- T_t, whether there are animal tracks.
- F_t, whether the food placed is gone.
- $\mathbf{E}_t = \{T_t, F_t\}$

A_t is the latent variable, and T_t and F_t are the observed variables at the time step t. A_t is a parent of A_{t+1}, T_t and F_t. Preferably draw this when handing in the exercise.

We have the following probabilities:

$$\begin{aligned} P(A_0 = true) &= 0.7 \\ P(A_t = true|A_{t-1} = true) &= 0.8 \\ P(A_t = true|A_{t-1} = false) &= 0.3 \\ P(T_t = true|A_t = true) &= 0.7 \\ P(T_t = true|A_t = false) &= 0.2 \\ P(F_t = true|A_t = true) &= 0.3 \\ P(F_t = true|A_t = false) &= 0.1 \end{aligned}$$

(b) Filtering

$$
\begin{aligned}
\mathbf{P}(A_0) &= \langle 0.7, 0.3 \rangle \\
\mathbf{P}(A_1) &= \sum_{a_0} \mathbf{P}(A_1|a_0)P(a_0) \\
&= \langle 0.8, 0.2 \rangle 0.7 + \langle 0.3, 0.7 \rangle 0.3 \\
&= \langle 0.65, 0.35 \rangle \\
\mathbf{P}(A_1|\mathbf{e}_1) &= c\mathbf{P}(\mathbf{e}_1|A_1)\mathbf{P}(A_1) \\
&= c\langle 0.7 \cdot 0.3, 0.2 \cdot 0.1 \rangle \langle 0.65, 0.35 \rangle \\
&= c\langle 0.1365, 0.007 \rangle \\
&= \langle 0.9512, 0.0488 \rangle \\
\mathbf{P}(A_2|\mathbf{e}_1) &= \sum_{a_1} \mathbf{P}(A_2|a_1)P(a_1|\mathbf{e}_1) \\
&= \langle 0.8, 0.2 \rangle 0.9512 + \langle 0.3, 0.7 \rangle 0.0488 \\
&= \langle 0.7756, 0.2244 \rangle \\
\mathbf{P}(A_2|\mathbf{e}_{1:2}) &= c\mathbf{P}(\mathbf{e}_2|A_2)\mathbf{P}(A_2|\mathbf{e}_1) \\
&= c\langle 0.3^2, 0.1 \cdot 0.8 \rangle \langle 0.7756, 0.2244 \rangle \\
&= \langle 0.7954, 0.2045 \rangle \\
\mathbf{P}(A_3|\mathbf{e}_{1:2}) &= \sum_{a_2} \mathbf{P}(A_3|a_2)P(a_2|\mathbf{e}_{1:2}) \\
&= \langle 0.8, 0.2 \rangle 0.7954 + \langle 0.3, 0.7 \rangle 0.2045 \\
&= \langle 0.6978, 0.3022 \rangle \\
\mathbf{P}(A_3|\mathbf{e}_{1:3}) &= c\mathbf{P}(\mathbf{e}_3|A_3)\mathbf{P}(A_3|\mathbf{e}_{1:2}) \\
&= c\langle 0.3 \cdot 0.7, 0.8 \cdot 0.9 \rangle \langle 0.6978, 0.3022 \rangle \\
&= \langle 0.4024, 0.5976 \rangle \\
\mathbf{P}(A_4|\mathbf{e}_{1:3}) &= \sum_{a_3} \mathbf{P}(A_4|a_3)P(a_3|\mathbf{e}_{1:3}) \\
&= \langle 0.8, 0.2 \rangle 0.4024 + \langle 0.3, 0.7 \rangle 0.5976 \\
&= \langle 0.5012, 0.4988 \rangle \\
\mathbf{P}(A_4|\mathbf{e}_{1:4}) &= c\mathbf{P}(\mathbf{e}_4|A_4)\mathbf{P}(A_4|\mathbf{e}_{1:3}) \\
&= \langle 0.7^2, 0.2 \cdot 0.9 \rangle \langle 0.5012, 0.4988 \rangle \\
&= \langle 0.7323, 0.2677 \rangle
\end{aligned}
$$

(c) Prediction

$$
\begin{aligned}
\mathbf{P}(A_5|\mathbf{e}_{1:4}) &= \sum_{a_4} \mathbf{P}(A_5|a_4)P(a_4|\mathbf{e}_{1:4}) \\
&= \langle 0.8, 0.2\rangle 0.7323 + \langle 0.3, 0.7\rangle 0.2677 \\
&= \langle 0.6661, 0.3339\rangle \\
\mathbf{P}(A_6|\mathbf{e}_{1:4}) &= \sum_{a_5} \mathbf{P}(A_6|a_5)P(a_5|\mathbf{e}_{1:4}) \\
&= \langle 0.8, 0.2\rangle 0.6661 + \langle 0.3, 0.7\rangle 0.3339 \\
&= \langle 0.6331, 0.3669\rangle \\
\mathbf{P}(A_7|\mathbf{e}_{1:4}) &= \sum_{a_6} \mathbf{P}(A_7|a_6)P(a_6|\mathbf{e}_{1:4}) \\
&= \langle 0.8, 0.2\rangle 0.6331 + \langle 0.3, 0.7\rangle 0.3669 \\
&= \langle 0.6165, 0.3835\rangle \\
\mathbf{P}(A_8|\mathbf{e}_{1:4}) &= \sum_{a_7} \mathbf{P}(A_8|a_7)P(a_7|\mathbf{e}_{1:4}) \\
&= \langle 0.8, 0.2\rangle 0.6165 + \langle 0.3, 0.7\rangle 0.3835 \\
&= \langle 0.6083, 0.3917\rangle
\end{aligned}
$$

(d)

$$
\begin{aligned}
P(A_t = true|\mathbf{e}_{1:4}) =& \sum_{a_{t-1}} P(A_t = true|a_{t-1})P(a_{t-1}|\mathbf{e}_{1:4}) \\
=& P(A_t = true|A_{t-1} = true)P(A_{t-1} = true|\mathbf{e}_{1:4}) \\
& + P(A_t = true|A_{t-1} = false)P(A_{t-1} = false|\mathbf{e}_{1:4}) \\
=& 0.8P(A_{t-1} = true|\mathbf{e}_{1:4}) + 0.3P(A_{t-1} = false|\mathbf{e}_{1:4}) \\
=& 0.8P(A_{t-1} = true|\mathbf{e}_{1:4}) + 0.3(1 - P(A_{t-1} = true|\mathbf{e}_{1:4})) \\
& (\text{Assume } P(A_t = true|\mathbf{e}_{1:4}) = P(A_{t-1} = true|\mathbf{e}_{1:4})) \\
=& 0.6 \\
\mathbf{P}(A_t|\mathbf{e}_{1:4}) =& \langle 0.6, 0.4\rangle
\end{aligned}
$$

(e) Backward

$$
\begin{aligned}
\mathbf{P}(\mathbf{e}_4|A_4) &= \langle 0.7^2, 0.2\cdot 0.9\rangle = \langle 0.49, 0.18\rangle \\
\mathbf{P}(\mathbf{e}_4|A_3) &= \sum_{a_4} P(\mathbf{e}_4|a_4)P(\mathbf{e}_5|a_4)\mathbf{P}(a_4|A_3) \\
&= \langle 0.8, 0.3\rangle 0.49 + \langle 0.2, 0.7\rangle 0.18 \\
&= \langle 0.428, 0.273\rangle \\
\mathbf{P}(\mathbf{e}_{3:4}|A_2) &= \sum_{a_3} P(\mathbf{e}_3|a_3)P(\mathbf{e}_4|a_3)\mathbf{P}(a_3|A_2) \\
&= \langle 0.1112, 0.1646\rangle \\
\mathbf{P}(\mathbf{e}_{2:4}|A_1) &= \sum_{a_2} P(\mathbf{e}_2|a_2)P(\mathbf{e}_{3:4}|a_2)\mathbf{P}(a_2|A_1) \\
&= \langle 0.0106, 0.0122\rangle \\
\mathbf{P}(\mathbf{e}_{1:4}|A_0) &= \sum_{a_1} P(\mathbf{e}_1|a_1)P(\mathbf{e}_{2:4}|a_1)\mathbf{P}(a_1|A_0) \\
&= \langle 0.0018, 0.0008\rangle
\end{aligned}
$$

Smoothing

$$
\begin{aligned}
\mathbf{P}(A_0|\mathbf{e}_{1:4}) &= c\mathbf{P}(A_0)\mathbf{P}(\mathbf{e}_{1:4}|A_0) \\
&= c\langle 0.7, 0.3\rangle\langle 0.0018, 0.0008\rangle \\
&= \langle 0.8359, 0.1641\rangle \\
\mathbf{P}(A_1|\mathbf{e}_{1:4}) &= c\mathbf{P}(A_1|\mathbf{e}_1)\mathbf{P}(\mathbf{e}_{2:4}|A_1) \\
&= c\langle 0.9512, 0.0488\rangle\langle 0.0106, 0.0122\rangle \\
&= \langle 0.9444, 0.0556\rangle \\
\mathbf{P}(A_2|\mathbf{e}_{1:4}) &= c\mathbf{P}(A_2|\mathbf{e}_{1:2})\mathbf{P}(\mathbf{e}_{3:4}|A_2) \\
&= \langle 0.7244, 0.2756\rangle \\
\mathbf{P}(A_3|\mathbf{e}_{1:4}) &= c\mathbf{P}(A_3|\mathbf{e}_{1:3})\mathbf{P}(\mathbf{e}_4|A_3) \\
&= \langle 0.5135, 0.4865\rangle
\end{aligned}
$$

References

[1] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.

[2] Charles R. Harris, K. Jarrod Millman, Stéfan J van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585:357–362, 2020.

[3] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.