

CS 486/686 Assignment 3
Fall 2022
(125 marks)

Blake VanBerlo

Due Date: 11:59 pm ET on Tuesday, November 15, 2022

Instructions

- Submit the signed academic integrity statement any written solutions in a file to the Q0 box in the A3 project on Crowdmark. **(5 marks)**.
- Submit your written answers to questions 1 and 2 as PDF files to the Q1 and Q2 boxes respectively in the A3 project on Crowdmark. I strongly encourage you to complete your write-up in LaTeX, using this source file. If you do, in your submission, please replace the author with your name and student number. Please also remove the due date, the Instructions section, and the Learning Goals section. Thank you!
- Submit any code to Marmoset at <https://marmoset.student.cs.uwaterloo.ca/>. Be sure to submit your code to the project named **Assignment 3 - Final**.
- No late assignment will be accepted. This assignment is to be done individually.
- Lead TAs:
 - Connor Raymond Stewart (crstewart@uwaterloo.ca)
 - Dake Zhang (dake.zhang@uwaterloo.ca)

The TAs' office hours will be scheduled and posted on LEARN and Piazza.

Learning goals

Hidden Markov Models (HMM)

- Derive the formula for the prediction inference task in HMMs
- Trace the forward-backward algorithm

Decision Trees

- Implement the decision tree learner algorithm to learn a decision tree using a dataset with real-valued features.
- Determine the accuracy of a decision tree on a dataset.
- Perform pre-pruning and post-pruning. Determine the best parameter value for pruning using k -fold cross validation.

1 Hidden Markov Models (30 marks)

- (a) In Lectures 10 and 11, we used a different form of the Bayes rule in our derivations for the filtering and smoothing formulas. Show that $P(A | B \wedge C) = \alpha P(B | A \wedge C)P(A | C)$, where α is a normalizing constant. In your derivation, clearly identify any probability rules that you use (see the rules in Lecture 6).

Marking Scheme:

(2 marks) Correct derivation

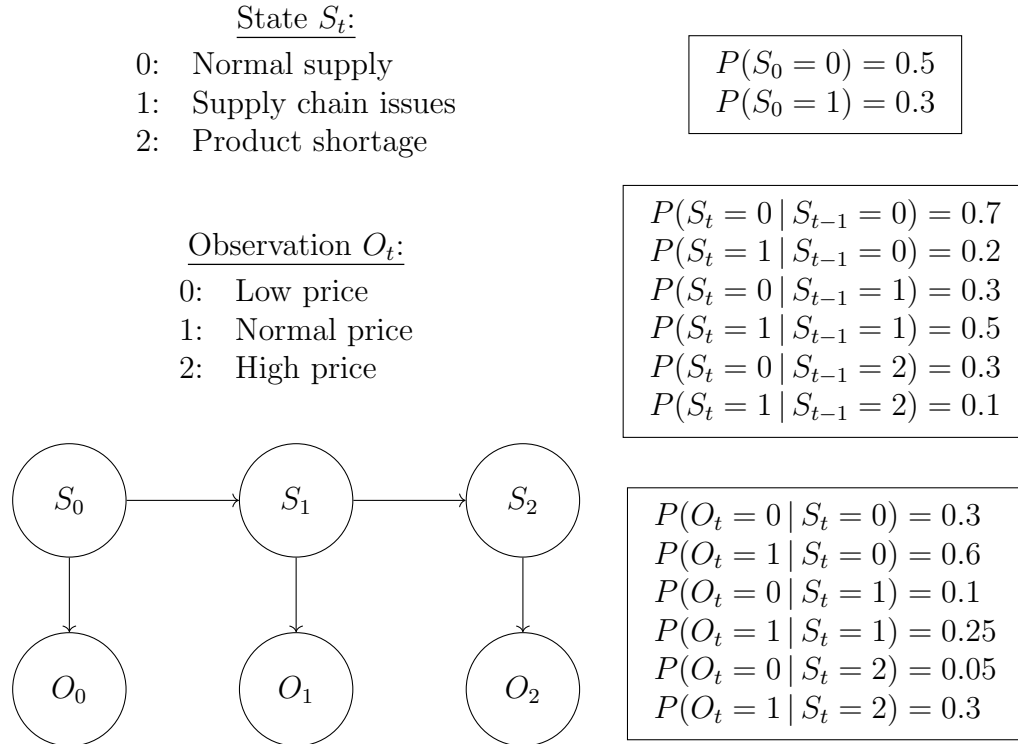
(2 marks) Correct rules cited

Solutions:

$$\begin{aligned} P(A|B \wedge C) &= P(A \wedge B \wedge C)/P(B \wedge C) - \text{Bayes rule} \\ &= P(B|A \wedge C)P(A \wedge C)/P(B \wedge C) - \text{Bayes rule} \\ &= P(B|A \wedge C)P(A|C)P(C)/P(B \wedge C) - \text{Bayes rule} \\ &= \alpha P(B | A \wedge C)P(A | C) , \text{ where } \alpha = P(C)/P(B \wedge C) \end{aligned}$$

- (b) Several commodities have been experiencing volatile prices recently, often due to supply chain issues and product shortages. Olive oil is an example of a product that might experience product shortages this year due to [droughts in Europe](#).

Suppose that you would like to determine the state of olive oil availability, given the price that you observe during your weekly grocery shopping trip. You model this problem as a Hidden Markov Model (HMM), where the hidden state S_t is the state of olive oil supply (normal, supply chain backup, or shortage) and the observation is the price you see in the grocery store for a 1 L bottle (low, normal, high). You make some assumptions regarding the transition and sensor distributions. For 3 weeks, you take note of the olive oil price, obtaining the HMM shown below:



You observed that the price of olive oil was normal in week 0, high in week 1, and high in week 2. Execute the Forward Backward Algorithm to determine the probability distribution for the hidden state at each of weeks 0, 1, and 2. That is, calculate $P(S_0 | O_0 = 1 \wedge O_1 = 2 \wedge O_2 = 2)$, $P(S_1 | O_0 = 1 \wedge O_1 = 2 \wedge O_2 = 2)$, and $P(S_2 | O_0 = 1 \wedge O_1 = 2 \wedge O_2 = 2)$. Normalize each forward message. Show all calculations and round each calculation to 4 decimal places.

Marking Scheme:

(10 marks) Correct execution of the Forward Backward Algorithm

(2 marks) Calculations are displayed clearly

Solutions:

[Refer to attached image](#)

- (c) Recall that *prediction* in Hidden Markov Models is the task of computing the probability distribution for the hidden state at some future time step. More specifically, prediction computes the distribution $P(S_k | o_{0:t})$, where t is the current time step and k is some future time step ($k > t$).

Given the state distribution for time step $k \geq t$, show that the distribution for the state at time step $k + 1$ can be recursively computed as follows:

$$P(S_{k+1} | o_{0:t}) = \sum_{s_k} P(S_{k+1} | s_k) P(s_k | o_{0:t})$$

In your derivation, clearly identify any conditional independence assumptions and/or probability rules that you use.

Marking Scheme:

(6 marks) Correct derivation

(2 marks) Correct rules/assumptions cited

Solutions:

empty

- (d) After many time steps (i.e., the *mixing time*), the distribution returned by prediction will converge to a fixed point. In other words, once convergence is reached, $P(S_k | o_{0:t})$ does not change as k increases. This distribution is referred to as the *stationary distribution* of the HMM.

What is the stationary distribution of the HMM in Question 1(b)? Show your calculations.

Marking Scheme:

(4 marks) Correct answer and justification

(2 marks) Calculations are displayed clearly

Solutions:

empty