CIS 520, Machine Learning, Fall 2019
Homework 7
Due: Monday, November 18th, 11:59pm
Submit to Gradescope

Matthew Scharf

November 19, 2019

1 Hidden Markov Models

1. (a) Find the joint probability of the observation sequence $X_{1:2} = (\mathsf{sing}, \mathsf{TV})$ together with each possible hidden state sequence that could be associated with it, i.e. find the four probabilities below. Show your calculations, in particular the formulas to calculate the 4 joint probabilities.

$$P\big(X_{1:2} = (\mathsf{sing}, \mathsf{TV}), Z_{1:2} = (\mathsf{happy}, \mathsf{happy})\big) = \frac{1}{2}\frac{6}{10}\frac{4}{5}\frac{1}{10} = .024$$
$$P\big(X_{1:2} = (\mathsf{sing}, \mathsf{TV}), Z_{1:2} = (\mathsf{happy}, \mathsf{sad})\big) = \frac{1}{2}\frac{6}{10}\frac{1}{5}\frac{7}{10} = .042$$
$$P\big(X_{1:2} = (\mathsf{sing}, \mathsf{TV}), Z_{1:2} = (\mathsf{sad}, \mathsf{happy})\big) = \frac{1}{2}\frac{1}{10}\frac{1}{3}\frac{1}{10} = .0017$$
$$P\big(X_{1:2} = (\mathsf{sing}, \mathsf{TV}), Z_{1:2} = (\mathsf{sad}, \mathsf{sad})\big) = \frac{1}{2}\frac{1}{10}\frac{2}{3}\frac{7}{10} = .023$$

(b) Based on these probabilities, what is the most likely hidden state sequence $z_{1:2}$ that corresponds with an observation sequence (sing, TV)?

(happy,sad)

(c) Now, find the joint probability of this observation sequence together with each possible hidden state for day 2, i.e. the two probabilities below. Show your calculations for finding the 2 joint probabilities.

$$P\big(X_{1:2} = (\mathsf{sing}, \mathsf{TV}), Z_2 = \mathsf{happy}\big) =$$
$$P\big(X_{1:2} = (\mathsf{sing}, \mathsf{TV}), Z_{1:2} = (\mathsf{happy}, \mathsf{happy})\big) + P\big(X_{1:2} = (\mathsf{sing}, \mathsf{TV}), Z_{1:2} = (\mathsf{sad}, \mathsf{happy})\big) = .024 + .0017 = .0257$$

$$P\big(X_{1:2} = (\mathsf{sing}, \mathsf{TV}), Z_2 = \mathsf{sad}\big) =$$
$$P\big(X_{1:2} = (\mathsf{sing}, \mathsf{TV}), Z_{1:2} = (\mathsf{happy}, \mathsf{sad})\big) + P\big(X_{1:2} = (\mathsf{sing}, \mathsf{TV}), Z_{1:2} = (\mathsf{sad}, \mathsf{sad})\big) = .042 + .023 = .0653$$

(d) Based on these probabilities, what is the individually most likely hidden state on day 2 that corresponds with an observation sequence (sing, TV)?

sad

2. Estimate the most likely activity on day 3 as follows:

(a) Compute the marginal probability of Alice being happy or sad on day 2, i.e. calculate $P\big(Z_2 = \mathsf{happy}\big)$ and $P\big(Z_2 = \mathsf{sad}\big)$. They should sum up to 1.

$$P\big(Z_2 = \mathsf{happy}\big) = \frac{1}{2}\frac{4}{5} + \frac{1}{2}\frac{1}{3} = .57$$
$$P\big(Z_2 = \mathsf{sad}\big) = .43$$

(b) Write the formulas and give the values for the probabilities of Alice's mood Z_3 on day 3:

$$
\begin{aligned}
&P\big(Z_3 = \textsf{happy}\big) = .57\tfrac{4}{5} + .43\tfrac{1}{3} = .6 \\
&P\big(Z_3 = \textsf{sad}\big) = .4
\end{aligned}
$$

(c) Write the formulas and give the values for the probabilities of Alice's activity X_3 on day 3:

$$
\begin{aligned}
&P\big(X_3 = \textsf{sing}\big) = .6\tfrac{6}{10} + .4\tfrac{1}{10} = .4 \\
&P\big(X_3 = \textsf{walk}\big) = \dot{6}\tfrac{3}{10} + .4\tfrac{2}{10} = .26 \\
&P\big(X_3 = \textsf{TV}\big) = .34
\end{aligned}
$$

(d) What is the most likely activity on day 3?
sing

2 Recurrent Neural Networks

1. Estimate the most likely activity on day 3 as follows:

(a) Give the formula to calculate Z_2 and O_2.

$$Z_2 = Relu(UX_2 + WZ_1)$$

$$O_2 = softmax(VZ_2)$$

(b) Find the values of Z_2 and O_2.

$$Z_2 = [2.35, 1.95]$$

$$O_2 = [.243, .178, .579]$$

(c) What is the most likely predicted activity on day 3?

TV

2. Now estimate the most likely activity on day 4. To do this, use $X_3 = O_2$ and proceed as follows:

(a) Give the formula to calculate Z_3 and O_3.

$$Z_3 = ReLU(UO_2 + WZ_2)$$

$$O_3 = softmax(VZ_3)$$

(b) Find the values of Z_3 and O_3.

$$Z_3 = [4.468, 8.785]$$

$$O_3 = [.905, .066, .029]$$

(c) What is the most likely predicted activity on day 4?
sing

3 Bayesian Networks

1. Write an expression for the joint probability mass function $p(X_1, X_2, X_3, X_4, X_5, X_6)$ that makes the same (conditional) independence assumptions as the Bayesian network:

$$p(X_1, X_2, X_3, X_4, X_5, X_6) = p(X_1)p(X_2)p(X_3 \mid X_1, X_2)p(X_4 \mid X_2)p(X_5 \mid X_3)p(X_6 \mid X_3, X_4)$$

2. Consider a joint probability distribution satisfying the following factorization:

$$p(X_1, X_2, X_3, X_4, X_5, X_6) = p(X_1)p(X_2)p(X_3)p(X_4)p(X_5 \,|\, X_3)p(X_6 \,|\, X_3)\,.$$

   Is this distribution included in the class of joint probability distributions that can be represented by the Bayesian network? Briefly explain your answer.

   Yes it is included. It just has the added the added restrictions that $p(X_3) = p(X_3 \mid X_1, X_2)$, $p(X_4) = p(X_4 \mid X_2)$, and $p(X_6 \mid X_3) = p(X_6 \mid X_3, X_4)$.

3. If the edge from X_3 to X_6 is removed from the above network, will the class of joint probability distributions that can be represented by the resulting Bayesian network be smaller or larger than that associated with the original network? Briefly explain your answer.

4. Given the above figure, determine whether each of the following is true or false. Briefly justify your answer.

   (a) $p(X_1, X_2) = p(X_1)p(X_2)$

   True. There are no active trails.

   (b) $p(X_4, X_5|X_3) = p(X_4|X_3)p(X_5|X_3)$

   True. There are no active trails.

   (c) $p(X_1, X_2 \,|\, X_6) = p(X_1 \,|\, X_6)p(X_2 \,|\, X_6)$

   False. There is an active trail through X_6.

   (d) $p(X_1, X_6 \,|\, X_3) = p(X_1 \,|\, X_3)p(X_6 \,|\, X_3)$

   False. There is an active trail from X_1 down to X_3, and then back up and around to X_6.

4 Belief Net Construction

1. Add A
   $p(A) = .68$

2. Add B and decide whether to add a link from A to B
   $p(B) = .55$
   $p(B \mid A) = .53$
   $p(B \mid A^C) = .57$
   So, A and B are independent.

3. Add C and decide whether to add a link from A to C
   $p(C) = .68$
   $p(C \mid A) = .67$
   $p(C \mid A^C) = .71$
   So, A and C are independent.

4. Decide whether to add a link from B to C
   $p(C \mid B) = .58$
   So, C is dependent on B and we will add the link.

   Furthermore, $p(C \mid A, B) = .5$ which is more than .05 from $p(C \mid B)$ and so C is conditional dependent on A conditioned on B. So, I am adding a link from A to C.

5. Add D and decide whether to add a link from A to D
   $p(D) = .32$
   $p(D \mid A) = .33$
   $p(D \mid A^C) = .29$
   So, no link needs to be added from A to D.

6. Decide whether to add a link from B to D
   $p(D \mid B) = .33$
   $p(D \mid B^C) = .3$
   So, no link needs to be added from B to D.

   $p(D \mid A, B) = .38 \approx p(D \mid B) = p(D \mid A)$ so no more links need to be added from A or B to D.

7. Decide whether to add a link from C to D
   $p(D \mid C) = .47$

   So, D is dependent on C. So, I am adding a link from C to D.

8. Final network:

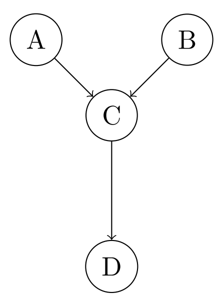

Figure 1: Belief Net

Be sure to give the **conditional probability tables** for downstream variables when appropriate.