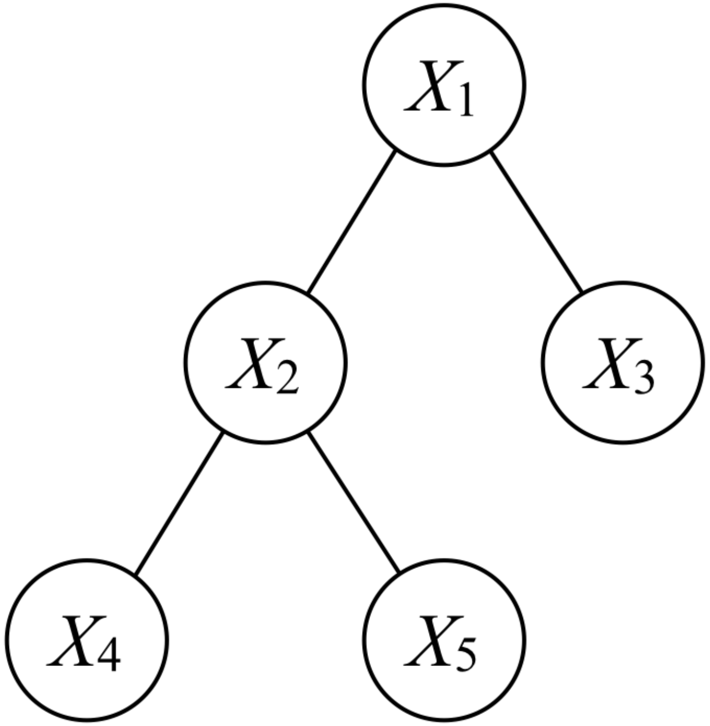
Exercise: Twitter Follower Network

Bookmark this page

Exercise: Twitter Follower Network

10/10 points (graded)

The figure below shows the Twitter follower network among 5 users $u = 1, 2, 3, 4, 5$, where each node u represents user u and each edge indicates that the user corresponding to the child node is following the user corresponding to the parent node. That is, user 1 does not follow anyone, users 2 and 3 follow user 1, and users 4 and 5 follow user 2. Suppose that user 1 is the source of all tweets, and each of the other users does not tweet anything but only retweets the tweets of the user that she is following. On any given day, user 1 tweets about something with some probability θ_1, and each user u retweets any (re)tweet of the user she follows (user v) with probability $\theta_{u|v}$, regardless of the content. We would like to estimate how likely each user retweets a tweet based on observations during one week.



Let X_1 be a binary random variable associated with user 1, which is set to 1 if user 1 tweets about something on a given day. Let X_u be a binary random variable associated with user u, which is set to 1 if user u retweets a tweet for $u = 2, 3, 4, 5$. Suppose that we observe whether each user (re)tweets for 7 days as the table below.

i	$(x_1^{(i)}, x_2^{(i)}, x_3^{(i)}, x_4^{(i)}, x_5^{(i)})$
1	(1, 1, 1, 0, 0)
2	(1, 1, 0, 1, 0)
3	(1, 1, 1, 1, 1)
4	(1, 0, 1, 0, 0)
5	(1, 1, 1, 0, 0)
6	(1, 0, 1, 0, 0)
7	(0, 0, 0, 0, 0)

- Find the maximum likelihood estimate for the parameters θ_1 and $\theta_{u|v}$ for each pair of nodes $(u, \pi(u))$ where $\pi(u)$ is the parent of u.

  Hint: Feel free to apply the final result that we saw in the video which is to look at empirical distributions.

  (Please be precise with at least 3 decimal places, unless of course the answer doesn't need that many decimal places. You could also put a fraction.)

$\hat{\theta}_1 =$ 0.857142857143 ✔ **Answer:** 6/7

$\hat{\theta}_{2|1} =$ 0.666666666667 Answer: 4/6

$\hat{\theta}_{3|1} =$ 0.833333333333 Answer: 5/6

$\hat{\theta}_{4|2} =$ 0.5 Answer: 2/4

$\hat{\theta}_{5|2} =$ 0.25 Answer: 1/4

Solution:

The ML estimates are given by the empirical distributions:

$$\hat{\theta}_1 = p_{X_1}(1) = \hat{p}_{X_1}(1) = \boxed{\frac{6}{7}}$$

$$\hat{\theta}_{2|1} = p_{X_2|X_1}(1|1) = \hat{p}_{X_2|X_1}(1|1) = \boxed{\frac{4}{6}}$$

$$\hat{\theta}_{3|1} = p_{X_3|X_1}(1|1) = \hat{p}_{X_3|X_1}(1|1) = \boxed{\frac{5}{6}}$$

$$\hat{\theta}_{4|2} = p_{X_4|X_2}(1|1) = \hat{p}_{X_4|X_2}(1|1) = \boxed{\frac{2}{4}}$$

$$\hat{\theta}_{5|2} = p_{X_5|X_2}(1|1) = \hat{p}_{X_5|X_2}(1|1) = \boxed{\frac{1}{4}}.$$

- Suppose now that user 5 decided to unfollow user 2 and follow another user among users 1, 2, 3 and 4 (note that user 5 could change her mind and decide to re-follow user 2), but we do not know whom user 5 decided to follow. We observe $(X_1, \dots, X_5)$ for the next week, and coincidentally, we observe the same values as the previous week (so please again use the same table above). In this part, we determine who user 5 is following now, using maximum likelihood. We assume that the probability that each user (re)tweets remains the same as in the previous part.

  Hint: We are basically removing edge (2, 5) and then deciding which edge to add that includes user 5. Remember each edge $(u, \pi(u))$ contributes a piece to the overall log likelihood:

$$\ell_{u|\pi(u)} \triangleq \sum_{i=1}^{7} \log p_{X_u|X_{\pi(u)}}(x_u^{(i)} \mid x_{\pi(u)}^{(i)}; \theta_{u|\pi(u)}).$$

  (Please be precise with at least 3 decimal places, unless of course the answer doesn't need that many decimal places. You could also put a fraction.)

  Suppose that the edge we're adding is (1,5) so that $\pi(5) = 1$. In this case, what is $\ell_{u|\pi(u)}$? Please use natural log. -2.82470472338 ✔ **Answer:** -2.8247

  Solution:

  edXmath

$$\begin{aligned}\ell_{5|\pi(5)} &= \sum_{i=1}^{7} \log p_{X_5|X_1}(x_5^{(i)}|x_1^{(i)}) \\ &= 5 \log p_{X_5|X_1}(1|0) + 1 \log p_{X_5|X_1}(1|1) + 1 \log p_{X_5|X_1}(0|0)\end{aligned}$$

$$= \boxed{5\log\frac{3}{4} + 1\log\frac{1}{4} + 1\log 1}$$
$$\approx -\mathbf{2.8247.}$$

\end{edXmath}

Suppose that the edge we're adding is (2,5) so that $\pi(5) = \mathbf{2}$. In this case, what is $\ell_{u|\pi(u)}$? Please use natural log. -2.24934057848 ✔ **Answer:** -2.2493

Solution:

Using similar reasoning as in the previous case, we end up with

$$\ell_{5|\pi(5)} = \boxed{3\log\frac{3}{4} + 1\log\frac{1}{4} + 4\log 1} \approx -\mathbf{2.2493.}$$

Suppose that the edge we're adding is (3,5) so that $\pi(5) = \mathbf{3}$. In this case, what is $\ell_{u|\pi(u)}$? Please use natural log. -2.53702265093 ✔ **Answer:** -2.5370

Solution:

Using similar reasoning as in the earlier cases, we end up with

$$\ell_{5|\pi(5)} = \boxed{4\log\frac{3}{4} + 1\log\frac{1}{4} + 1\log 1} \approx -2.5370.$$

Suppose that the edge we're adding is (4,5) so that $\pi(5) = 4$. In this case, what is $\ell_{u|\pi(u)}$? Please use natural log. -1.67397643357 **Answer:** -1.6740

Solution:

Using similar reasoning as in the earlier cases, we end up with

$$\ell_{5|\pi(5)} = \boxed{1\log\frac{3}{4} + 1\log\frac{1}{4} + 4\log 1} \approx -1.6740.$$

Given the four numbers you computed above, who is user 5 following according to maximum likelihood?

- 1
- 2
- 3

4 ✔

Solution:

The overall log likelihood will be the sum of the same terms except for the $\ell_{5|\pi(5)}$, which is different across the four cases considered above, in which the one with highest value corresponds to adding edge (4, 5), so the ML estimate for who user 5 is following is user $\boxed{4}$.

Submit You have used 2 of 5 attempts

✔ Correct (10/10 points)

Discussion

Topic: Parameter Learning - Finite Random Variables and Trees / Exercise: Twitter Follower Network

Hide Discussion

Add a Post

posted 2 days ago by **seanedXacc**

I've found this exercise formulation very hard to grok. With @RADUGROSU help I've managed to get right answer, but it's unclear to me, howcome *conditional* probability of retweeting could magically be converted into inherent unconditional one.

posted 2 days ago by **Mark_B2** **Community TA**

> Or does it mean, user 5's probability of retweeting remains (fraction from part 1) and if he follows somebody who tweets more, he will also retweet more?

That's what it means. So, given the new observations, some users make more likely parents than others.

posted 2 days ago by **RADUGROSU** **Community TA**

using Bayes (log(x) + log(y|x)) to find the log p of y|x and taking max to find most likely parent?

posted 2 days ago by **seanedXacc**

@RADUGROSU Are we saying that $P_{5|n}$ is the same as $P_{5|2}$ from the previous part no matter where we hang X_5?

— **Collapse discussion**

[Parameter Learning Exercise] typo in answer

discussion posted 2 days ago by **RADUGROSU** **Community TA**

The coefficients of the cases where the parent doesn't tweet are once one too large and twice one too small (not that it makes a difference in...

This post is visible to everyone.

+ Expand discussion

estimate datatype

discussion posted 3 days ago by **jbtl**

Is $\hat{\theta}_{2|1}$ a scalar or some 2-dimensional structure like list of lists or map of maps?

This post is visible to everyone.

Add a response:

0 responses

Preview

Submit

Collapse discussion

1