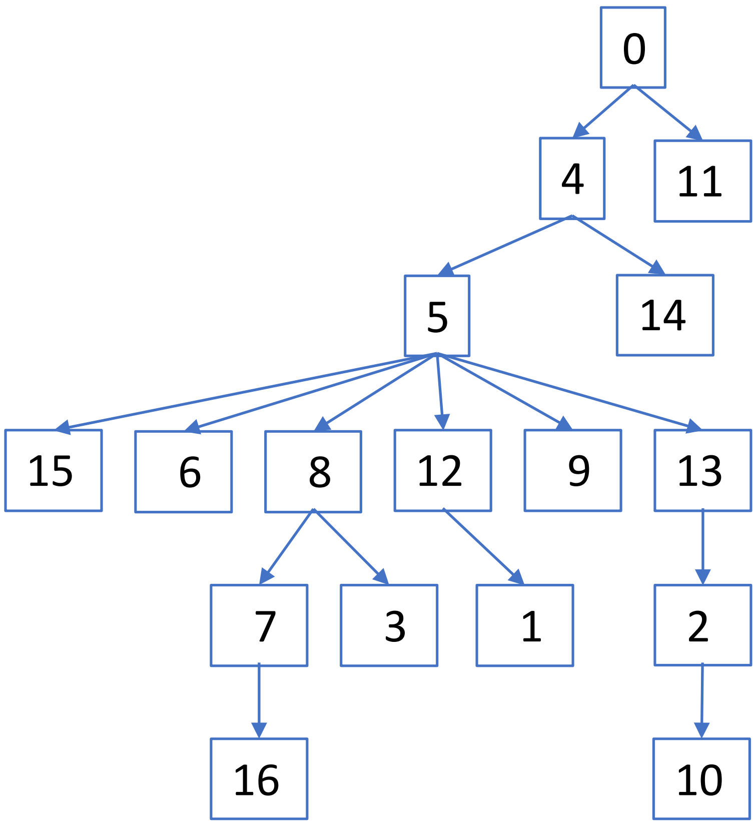
CS 6375
Prof. Nick Ruozzi
Problem Set 5
Student: Tri M. Cao
Date: November 25, 2017

Part 1: Structure Learning

Note: variable number starts at 0, i.e. the first column in the data is variable 0.



Part 2: EM Algorithm

Log likelihood = -3286.8483

The parameters are listed below:

Missingness probabalities

$b_0 = 0.0$	$b_1 = 0.0276$	$b_2 = 0.1103$	$b_3 = 0.0253$	$b_4 = 0.0253$	$b_5 = 0.0345$
$b_6 = 0.0253$	$b_7 = 0.0322$	$b_8 = 0.0345$	$b_9 = 0.0506$	$b_{10} = 0.0161$	$b_{11} = 0.0483$
$b_{12} = 0.0713$	$b_{13} = 0.0575$	$b_{14} = 0.0391$	$b_{15} = 0.0644$	$b_{16} = 0.2391$	

$P(x_0)$

$P(x_0 = 1) = 0.3862$	$P(x_0 = 0) = 0.6138$

$P(x_1|x_{12})$ *I only write the probabilities for the case $P(x_1 = 1)$ since $P(x_1 = 0) = 1 – P(x_1 = 1)$*

$P(x_1 = 1 \mid x_{12} = 0) = 0.6273$	$P(x_1 = 1 \mid x_{12} = 1) = 0.1926$

$P(x_2|x_{13})$

$P(x_2 = 1 \mid x_{13} = 0) = 0.3783$	$P(x_2 = 1 \mid x_{13} = 1) = 0.6236$

$P(x_3|x_8)$

$P(x_3 = 1 \mid x_8 = 0) = 0.1963$	$P(x_3 = 1 \mid x_8 = 1) = 0.8998$

$P(x_4|x_0)$

$P(x_4 = 1 \mid x_0 = 0) = 0.0563$	$P(x_4 = 1 \mid x_0 = 1) = 0.9825$

$P(x_5|x_4)$

$P(x_5 = 1 \mid x_4 = 0) = 0.1845$	$P(x_5 = 1 \mid x_4 = 1) = 0.9549$

$P(x_6|x_5)$

$P(x_6 = 1 \mid x_5 = 0) = 0.3331$	$P(x_6 = 1 \mid x_5 = 1) = 0.9399$

$P(x_7|x_8)$

$P(x_7 = 1 \mid x_8 = 0) = 0.1425$	$P(x_7 = 1 \mid x_8 = 1) = 0.8873$

$P(x_8|x_5)$

$P(x_8 = 1 \mid x_5 = 0) = 0.9904$	$P(x_8 = 1 \mid x_5 = 1) = 0.1475$

$P(x_9|x_5)$

$P(x_9 = 1 \mid x_5 = 0) = 0.9161$	$P(x_9 = 1 \mid x_5 = 1) = 0.0995$

$P(x_{10}|x_2)$

$P(x_{10} = 1 \mid x_2 = 0) = 0.5680$	$P(x_{10} = 1 \mid x_2 = 1) = 0.4432$

$P(x_{11}|x_0)$

$P(x_{11} = 1 \mid x_0 = 0) = 0.5059$	$P(x_{11} = 1 \mid x_0 = 1) = 0.1321$

$P(x_{12}|x_5)$

$P(x_{12} = 1 \mid x_5 = 0) = 0.0978$	$P(x_{12} = 1 \mid x_5 = 1) = 0.7440$

$P(x_{13}|x_5)$

$P(x_{13} = 1 \mid x_5 = 0) = 0.1719$	$P(x_{13} = 1 \mid x_5 = 1) = 0.8474$

$P(x_{14}|x_4)$

$P(x_{14} = 1 \mid x_4 = 0) = 0.3175$	$P(x_{14} = 1 \mid x_4 = 1) = 0.9820$

$P(x_{15}|x_5)$

$P(x_{15} = 1 \mid x_5 = 0) = 0.7198$	$P(x_{15} = 1 \mid x_5 = 1) = 0.1363$

$P(x_{16}|x_7)$

$P(x_{16} = 1 \mid x_7 = 0) = 0.6135$	$P(x_{16} = 1 \mid x_7 = 1) = 0.9932$

Q2b: How is the EM algorithm affected by the initialization? Give an example.
For this dataset, the EM algorithm is not affected by different initializations. I have tried many different initializations, but the EM algorithm always converges to the same parameters and log-likelihood.

Q2c: Explain how you could generate new samples (with missing entries) from the learned parameters.
To generate a new sample, we use the Bayesian network. Start at the root node, we generate each variable following the directed arrow (generate the parent node before generating the child node).
For node 0, we generate variable x_0 using the probability $P(x_0 = 1) = (1 – b_0) * P(x_0 = 1)$. Note that $b_0 = 0$, therefore x_0 is never missing, and $x_0 = 1$ with probability $P(x_0 = 1)$.
Next, we generate a value for x_4, since x_4 is a child of x_0. There is a probability b_4 that x_4 will be missing. And $x_4 = 1$ with probability $(1 – b_4) * P(x_4 = 1 \mid x_0)$, $x_4 = 0$ with probability $(1 – b_4) * (1 – P(x_4 = 1 \mid x_0)$. Note that the value of x_0 is known since we have already generated x_0.
Then we continue to generate other variables.

Q2d: In what ways is the EM algorithm in this setting like data imputation?
The E-step in EM algorithm, like data imputation, tries to complete the data before finding the model parameters that maximize the likelihood. However, unlike data imputation, EM algorithm considers every possible value of the missing variable. This might be important because a variable might rarely have a value, but that value would affect other variables

significantly. That’s why EM algorithm is guaranteed to find a local maximum, while data imputation is not.

Q2e: How do you think the results would change if we replaced the missing completely at random assumption with the missing at random assumption?

The missing mechanism does not affect the EM algorithm, hence changing the missing mechanism will not affect the results. The only thing that change is the sample generating process will be more complicated with many more missingness parameters.