University of TEHRAN

Probabilistic Graphical Models

Problem Set 5

Hesam Montazeri
Fahimeh Palizban
Naser Elmi

Farvardin 10, 1398
(March 30, 2019)

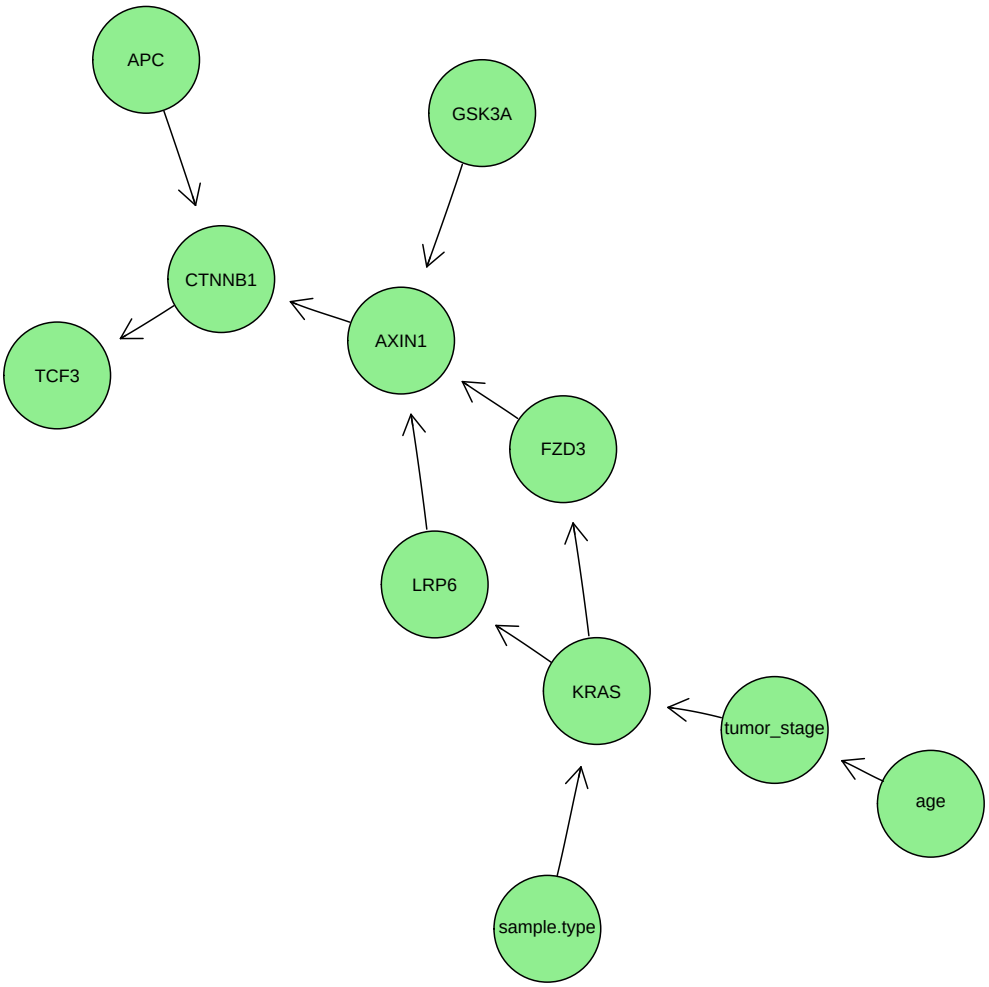
Problem 1: Reading Summary
Write a one-page summary of the previous lecture (max: exceptionally 2 pages for this exercise). Accompany your report by an audio file (max: 10 minutes) in which you explain in your words important topics of the lecture, particularly:

- Parameter estimation and structure learning
- Fully observed data; partially observed data
- Definition of likelihood function; why does it have a different notation than $P(D : \theta)$?
- MLE for Bernoulli distribution (*)
- MLE for Gaussian distribution (*)
- MLE for Bayesian networks $X \rightarrow Y$ (*)
- Global likelihood decomposition of Bayesian networks (*); explain how you maximize local likelihood functions in case of table CPDs.
- Data fragmentation; overfitting
- MLE for Gaussian Bayesian networks (no need to discuss the proof)
- Referring to the slide 26, solve problems 2 and 3 (no need to explain your solutions in the audio file).

* Write down all formulas in your written summary and explain in detail each step of the derivation. In your audio file, only mention the main points of the derivations.

Problem 2: Maximum likelihood estimation of multinomial distribution
Suppose X is a multinomial random variable that takes values $x^1, x^2, \ldots x^K$. The multinomial distribution has K parameters $\Theta = (\theta_1, \ldots, \theta_K)$ such that $P(x : \theta) = \theta_k$ if $x = x_k$ subject to $\sum_k \theta_k = 1$. Let the observed data is $(M[1], \ldots, M[K])$ where $M[k]$ is the number of times the value x^k appears in the data. Prove the ML estimate for θ_k is $\frac{M[k]}{M}$ where $M = \sum_k M[k]$ (hint: use Lagrange multiplier).

Problem 3: WNT Signaling pathway in colorectal cancer-parameter estimation
In this exercise, we extend the Bayesian network discussed in Problem 3, HW4 with two additional nodes namely *age* and *tumor_stage*.



Using the dataset *HW5.csv*, your task is to implement ML parameter estimation method in R for the nodes *age*, *tumor_stage*, *KRAS*, *APC* and *FZD3*. Compare your results with *bn.fit* method (with *method*="mle") in the *bnlearn* package.

Submit your solutions to naser.elmi@ut.ac.ir by Farvardin 18, 1398.