Probabilistic Graphical Models

Problem Set 4

Hesam Montazeri
Fahimeh Palizban
Naser Elmi

Esfand 13, 1397
(March 4, 2019)

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

- Tabular CPDs

- Deterministic CPDs; its independence properties

- Tree CPDs; context-specific independencies ($\perp_c$); Rule CPDs

- Generalized linear models

- Linear Gaussian models

- Multivariate Gaussians, covariance form, information form, independencies in Gaussians

- Operations on Gaussians: marginal distribution; conditional distributions. What distribution do you obtain for each operation?

- Different types of dependencies in hybrid models

- Conditional Gaussian Bayesian network

Problem 2: Gaussian Bayesian networks
Consider the following Bayesian network. Suppose all random variables are continuous.
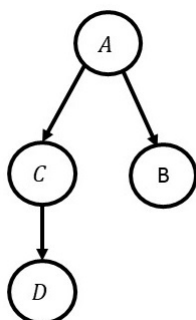
(a) Write down linear Gaussian models for all variables. How many parameters in total does the BN have?

(b) Now set some arbitrary values to the parameters of conditional Gaussians. Compute joint distribution $p(A, B, C, D)$ from the linear Gaussian models.

(c) Compute the distribution of $B \mid C = 2$. What is $p(B > 0 \mid C = 2)$?

(d) Use the *bnlearn* package to compute $p(B > 0 \mid C = 2)$.

Problem 3: WNT Signaling pathway in colorectal cancer

WNT signaling is a key pathway in regulating development and stemness and has been described as a hallmark of colorectal cancers. The aim of this exercise is to use conditional Gaussian Bayesian network and investigate whether gene expression data of selected genes in this pathway can discriminate normal tissue and tumor samples.

To begin, use the R-script HW4.r and build the below conditional Gaussian Bayesian network for a subset of TCGA colorectal cancer dataset (HW4train.csv.)
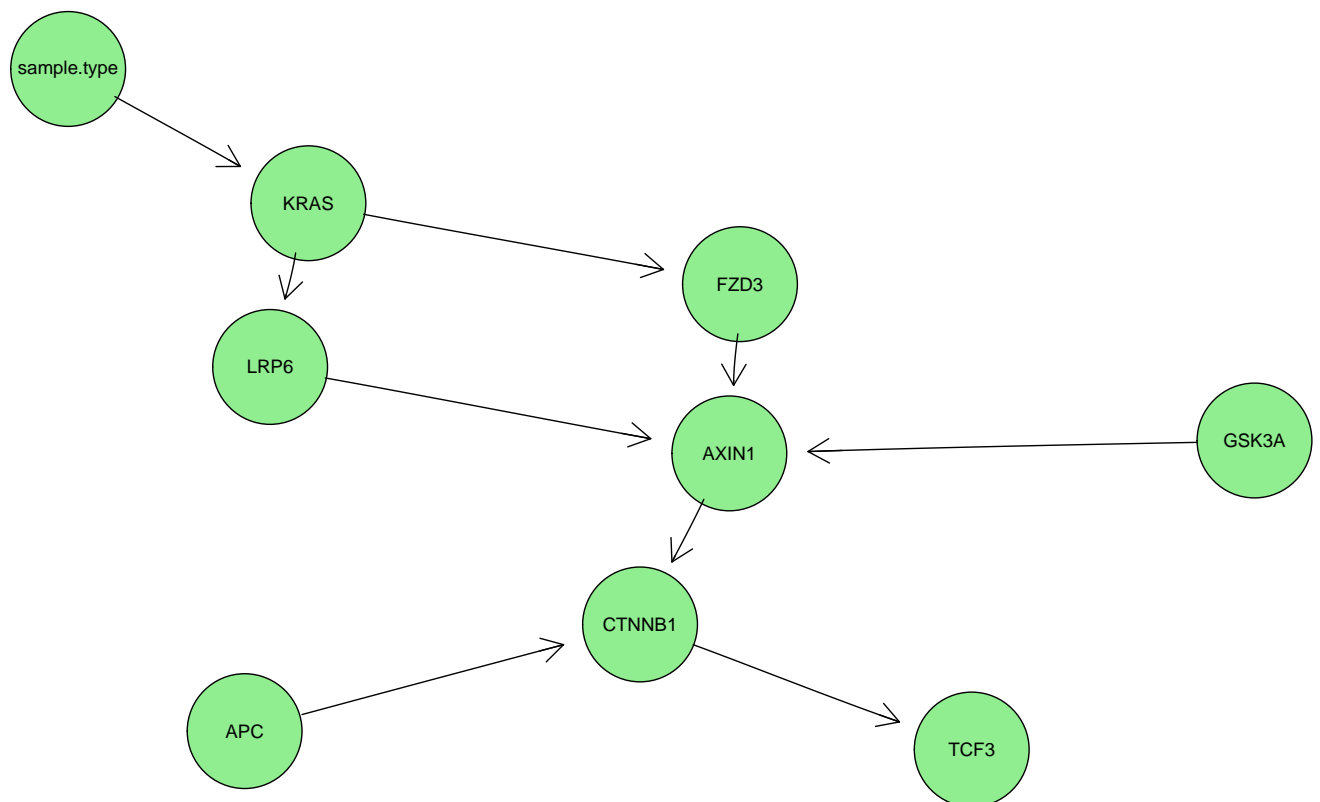
(a) Perform the parameter estimation. Write down the conditional probabilities for sample.type, KRAS | sample.type, and CTNNB1 | APC, AXIN1.

(b) Compute probability of sample.type=NT for all observations in the test data HW4test.csv. Report your results in a .csv file.

(c) Now modify the network such that sample.type is the parent of all other nodes. Plot the network and repeat part b.

Only the first 10 observations in the test data are normal tissue samples. What is your conclusion?