University of TEHRAN

Probabilistic Graphical Models

Problem Set 8

Hesam Montazeri
Fahimeh Palizban
Naser Elmi

Ordibehesht 4, 1398
(April 24, 2019)

Problem 1: Reading Summary

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

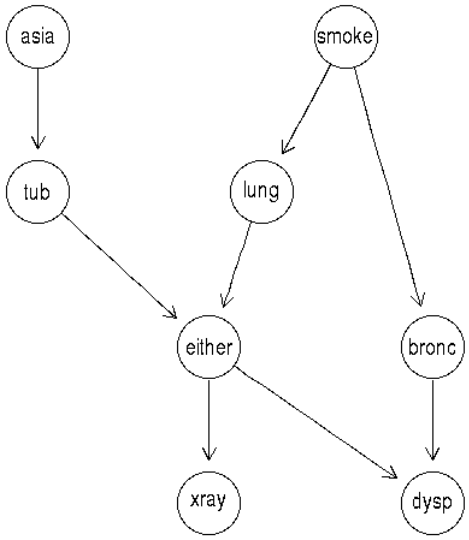
- Likelihood score; a simple example of the score
- Decomposition of the likelihood score and its proof (explain the steps of the proof)
- Limitations of the likelihood score
- Bayesian score; marginal likelihood; an example of marginal likelihood for a single random variable
- Bayesian information criterion (BIC); its consistency property
- Score equivalence
- Structure and parameter priors
- Structure search; learning tree-structured networks
- Briefly explain Kruskal's and Prim's algorithms.

* 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: Tuberculosis, Lung cancer and Bronchitis

Lauritzen and Spiegelhalter introduced a synthetic example on diagnosis of tuberculosis, lung cancer or bronchitis according to several factors [1]. The underlying network is assumed as the below figure and the dataset is available at *bnlearn* named *asia*. The authors motivated the example as

> Shortness-of-breath (dyspnoea) may be due to tuberculosis, lung cancer or bronchitis, or none of them, or more than one of them. A recent visit to Asia increases the chances of tuberculosis, while smoking is known to be a risk factor for both lung cancer and bronchitis. The results of a single chest X-ray do not discriminate between lung cancer and tuberculosis, as neither does the presence or absence of dyspnoea.



Apply different learning algorithms for Bayesian networks for this dataset, particularly:

(a) Compare the best tree-structured Bayesian network (Chow-Liu), empty network, a random network, the true network, learned networks by constraint-based learning algorithm (PC algorithm) and greedy hill climbing in terms of BIC and likelihood scores (for undirected graphs, you need to orient them by *pdag2dag*).

(b) Now consider the random network generated in the previous step. Generate all neighboring networks by edge addition, removal, or reversal and report the best network of each operation separately.

You may use utility functions available at *bnlearn*. Set a random seed in the beginning of your code so your results are reproducible.

Problem 3: Bayesian score with a K2 prior
From the textbook, exercise 18.10.

Submit your solutions to naser.elmi@ut.ac.ir and fahimehpalizban@ut.ac.ir by Ordibehesht 9, 1398.

References

[1] Lauritzen, S.L. and Spiegelhalter, D.J., 1988. *Local computations with probabilities on graphical structures and their application to expert systems.* Journal of the Royal Statistical Society: Series B (Methodological), 50(2), pp.157-194.