Bayesian Networks in Data Science

Tairan Ye

March 1, 2018

1 Introduction

A Bayesian network, Bayesian network, belief network, Bayesian model or probabilistic directed acyclic graphical model is a probabilistic graphical model, a type of statistical model, that represents a set of variables and their conditional dependencies via a directed acyclic graph (DAG). For example, a Bayesian network could represent the probabilistic relationships between diseases and symptoms. Given symptoms, the network can be used to compute the probabilities of presence of various diseases.

1.1 Statistical Introduction

Given data x and parameter θ, a simple Bayesian analysis starts with a prior probability or prior, $p(x|\theta)$ to compute a posterior probability $p(\theta|x) \propto p(x|\theta)p(\theta)$. Most likely, the prior on θ depends in turn on other parameters φ that are not mentioned in the likelihood. So, the prior $p(\theta)$ must be replaced by a likelihood $p(\theta \mid \varphi)$, and a prior $p(\varphi)$ on the newly introduced parameters φ is required, resulting in a posterior probability:

$$p(\theta, \varphi|x) \propto p(x|\theta)p(\theta|\varphi)p(\varphi) \tag{1}$$

which is the simplest example of a hierarchical Bayes model.
The process will be repeated – for example, the parameters φ may depend in turn on additional parameters ψ, which will require their own prior. Eventually the process must terminate, with priors that do not depend on any other unmentioned parameters.

1.2 Major Application

1. *Inferring Unobserved Variables*
   Since a Bayesian network is a complete model for the variables and their relationships, it can be used to answer probabilistic queries. The most common approximate inference algorithms are importance sampling, stochastic MCMC simulation, mini-bucket elimination, loopy belief propagation, generalized belief propagation, and variational methods.

2. *Parameter Learning*
   In order to fully specify the Bayesian network and thus fully represent the joint probability distribution, it is necessary to specify for each node X the probability distribution for X conditional upon X's parents.

3. *Structure Learning*
   A particularly fast method for exact BN learning is to cast the problem as an optimization problem, and solve it using integer programming.

2 Data

The dataset is named "Breast Cancer Wisconsin (Diagnostic)", from the "UCI Machine Learning Repository". Features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass. They describe characteristics of the cell nuclei present in the image. Separating plane described above was obtained using Multisurface Method-Tree (MSM-T) [K. P. Bennett, "Decision Tree Construction Via Linear Programming." Proceedings of the 4th Midwest Artificial Intelligence and Cognitive Science Society, pp. 97-101, 1992], a classification method which uses linear programming to construct a decision tree. Relevant features were selected using an exhaustive search in the space of 1-4 features and 1-3 separating planes. The actual linear program used to obtain the separating plane in the 3-dimensional space is that described in: [K. P. Bennett and O. L. Mangasarian: "Robust Linear Programming Discrimination of Two Linearly Inseparable Sets", Optimization Methods and Software 1, 1992, 23-34].

The more detail will be updated before next class on 03/06/2018. I apologize for the delay.