Probabilistic Graphical Models Project 3: BN Structure Learning Under Incomplete Data

Matthew Caulfield

ECSE-6810

November 17, 2020

1 Introduction

The goal of this project is to learn the structure of a 5 node Bayesian network given a set of incomplete data samples. Each node in this Bayesian network can have states 1 or 2 and an in determinant number of parents. When constructing a Bayesian network the links between nodes and the conditional probability distribution that is the best fit with the the data must be found. Because the data is incomplete the missing data samples must also be estimated and the structure is the one that finds the conditional probability distribution that best fits the estimated data. To find the most likely structure of the network given a set of data with some missing samples, the structural Expectation Maximization algorithm is used.

2 Theory

The structural Expectation Maximization algorithm is used to find the structure of the network given the incomplete data. This algorithm from Professor Ji's textbook is shown in algorithm 1[1].

Algorithm 1: Structure EM algorithm

initialize BN structure to G^0 and θ^0
while *not converging* **do**
  for m = *1 to* M **do**
    if *missing data* $z^m \in x^m$ **then**
      for j=*1 to* $K^{|z^m|}$ **do**
        $w_{m,j} = p(z_j^m | y^m, \theta^t)$
      end
    end
  end
  $E_q(BIC(G)) = \sum_{m=1}^{M} \sum_{j=1}^{K^{|z^m|}} w_{m,j} logp(y^m, z_j^m | \theta^t, G) - \frac{logM}{2} Dim(G)$
  $G^{t+1}, \theta^{t+1} = \max_G E_q(BIC(G))$ using hill-climbing method
  Output: G^{t+1}
end

This algorithm can be broken into two parts the expectation step and the maximization step. In algorithm one the expectation step is the step that iterates over the data to find the weights of the missing data based on the conditional probabilities of the current structure. The maximization step finds the best structure based on the BIC score and using the hill climbing algorithm shown in algorithm 2 also from Professor Ji's book [1].

Algorithm 2: Hill Climbing algorithm

initialize BN structure to G^0
order nodes $X_1, X_2, ...X_N$ **while** *Score not converging* **do**
  for n = *1 to* N **do**
    Add, remove, or change direction of a link of X_n to maximize the expected Bic Score from equation 1.
  end
end

This algorithm calculates the expected BIC score of a Bayesian Network using equation 1 .

$$E_q(BIC(G)) = \sum_{m=1}^{M} \sum_{j=1}^{K^{|z^m|}} w_{m,j} logp(y^m, z_j^m | \theta^t, G) - \frac{logM}{2} Dim(G) \tag{1}$$

This algorithm has some flaws for instance it may find a local maximum of the score and which means that the final network is not necessarily the best network for the data. It can also make a network that does not fit the DAG constraint. To make sure the final network is a DAG equation 2 is used.

$$tr(e^W) - d = 0 \tag{2}$$

Where W is a $d \times d$ binary matrix such that $W_{ij} = 1$ means that the ith and jth node are linked together.

3 Results

Unfortunately, after running one iteration of the structure EM algorithm, the Matlab code caused a memory overflow. The resulting structure from the one iteration is shown in figure 1. The BIC

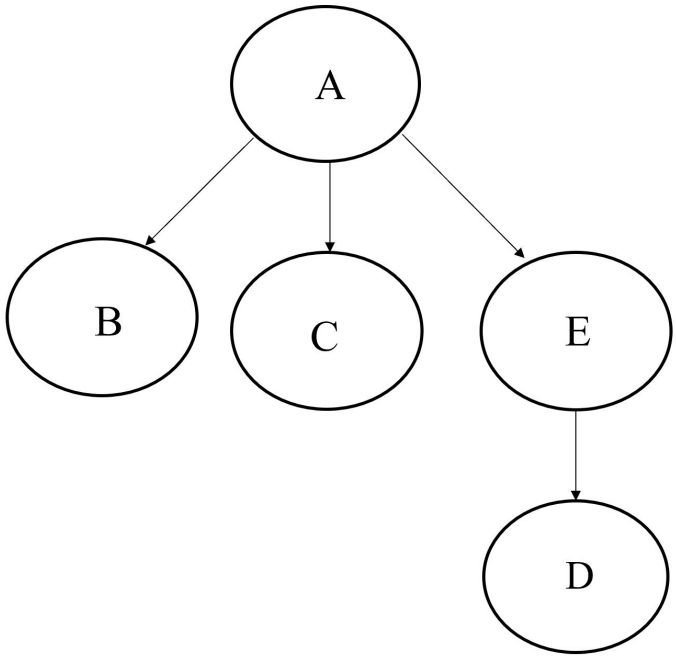

Figure 1: Final network

score of the initial unconnected network with the parameters to a uniform distribution was -34.8169. While the BIC score of the network found after the first iteration was -29.8864. As you can see the BIC score improved greatly. If the project was able to run more than one iteration you would continue to see the BIC score decrease until it converged to a maxima. Using equation 2 we found that this network was a DAG.

4 Works Cited

[1] Ji, Q. (2020). Chapter 3. In Probabilistic graphical models for computer vision (pp. 79-90). London: Academic Press.