Extension of Bayesian Network Classifiers to Regression Problems*

Antonio Fernández and Antonio Salmerón

Department of Statistics and Applied Mathematics, University of Almería,
Carrera de Sacramento s/n, E-04120 Almería, Spain
{afalvarez,antonio.salmeron}@ual.es

Abstract. In this paper we explore the extension of various Bayesian network classifiers to regression problems where some of the explanatory variables are continuous and some others are discrete. The goal is to compute the posterior distribution of the response variable given the observations, and then use that distribution to give a prediction. The involved distributions are represented as Mixtures of Truncated Exponentials. We test the performance of the proposed models on different datasets commonly used as benchmarks, showing a competitive performace with respect to the state-of-the-art methods.

Keywords: Bayesian networks, Regression, Mixtures of truncated exponentials.

1 Introduction

Bayesian network classifiers [4] constitute one of the most successful applications of Bayesian networks. Recently, Bayesian networks have been proposed as a tool for solving regression problems [2,3,7]. The main advantage with respect to other regression models is that it is not necessary to have a full observation of the explanatory variables to give a prediction for the response variables. Also, the model is usually richer from a semantic point of view. The most recent advances considering the application of Bayesian networks to regression problems, came through the extension of models that had been originally proposed as classifiers, as the so-called *naive Bayes (NB)* [7] and *tree augmented naive Bayes (TAN)* [2]. In both cases, an interesting feature of the proposed solutions is that they can handle problems in which the explanatory variables are discrete or continuous. This is achieved by modelling the joint distribution of the variables involved in the problem, as a mixture of truncated exponentials (MTE) [6].

The fact that models as NB and TAN are appropriate for classification as well as regression is not surprising, as the nature of both problems is similar: the goal is to predict the value of a response or class variable given an observation over the explanatory variables. Therefore, the aim of this paper is to

* This work has been supported by the Spanish Ministry of Education and Science, project TIN2007-67418-C03-02 and by Junta de Andalucía, project P05-TIC-00276.

H. Geffner et al. (Eds.): IBERAMIA 2008, LNAI 5290, pp. 83–92, 2008.

investigate the behaviour of different Bayesian network classifiers when applied to regression problems. In all the cases we will consider problems where some of the independent variables are continuous while some others are discrete, and therefore we will concentrate on the use of MTEs, since this distribution has shown to be an appropiate tool for handling discrete and continuous variables simultaneously. More precisely, the starting point is the NB and TAN models for regression proposed in [2,7].

2 Bayesian Networks and Regression

We will denote random variables by capital letters, and their values by lowercase letters. Random vectors will be denoted by boldfaced characters. We will use $\Omega_{\mathbf{X}}$ to indicate the support of a random vector $\mathbf{X}$. A *Bayesian network* for variables $\mathbf{X} = \{X_1, \ldots, X_n\}$ is a directed acyclic graph where each X_i is assigned to one node, which has associated a conditional distribution given its parents [8]. An arc linking two variables indicates the existence of probabilistic dependence between both of them. A fundamental property of Bayesian networks is that the joint distribution over $\mathbf{X}$ factorises according to the d-separation criterion as $p(x_1, \ldots, x_n) = \prod_{i=1}^{n} p(x_i|pa(x_i))$, where $Pa(X_i)$ is the set of parents of X_i and $pa(x_i)$ is a configuration of values of them. The factorisation induced allows to represent complex distributions by a set of simpler ones, and therefore the number of parameters needed to specify a model is lower in general.

A Bayesian network can be used as a regression model. Assume that Y is the response variable and $X_1, \ldots, X_n$ are the explanatory variables. Then, in order to predict the value for Y for the observations $x_1, \ldots, x_n$, the conditional density $f(y|x_1, \ldots, x_n)$, is computed and a numerical prediction for Y is given using the corresponding mean or the median. As $f(y|x_1, \ldots, x_n)$ is proportional to $f(y) \times f(x_1, \ldots, x_n|y)$, solving the regression problem would require to specify an n dimensional density for $X_1, \ldots, X_n$ given Y. However, using the factorisation encoded by the Bayesian network, this problem is simplified depending on the structure of the network. The extreme case is the NB structure, where all the explanatory variables are considered independent given Y (see Fig. 1(a)).

The strong independence assumption behind NB models is somehow compensated by the reduction of the number of parameters to be estimated from data, since in this case, it holds that $f(y|x_1, \ldots, x_n) \propto f(y) \prod_{i=1}^{n} f(x_i|y)$, which

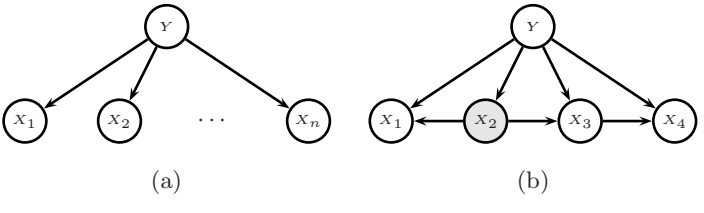

Fig. 1. Structure of a *naive Bayes* model (a) and a TAN model (b)