BoltzMM: an R package for maximum pseudolikelihood estimation of fully-visible Boltzmann machines

Andrew T. Jones\(^1\), Jessica J. Bagnall\(^2\), and Hien D. Nguyen\(^2\)

\(^1\) School of Mathematics and Physics, University of Queensland, St. Lucia 4072, Queensland Australia
\(^2\) Department of Mathematics and Statistics, La Trobe University, Bundoora 3086, Victoria Australia

Summary

The BoltzMM package (Version 0.1.3; https://CRAN.R-project.org/package=BoltzMM) for the R statistical programming environment (R Core Team, 2018) provides a complete suite of functions for application and estimation of fully-visible Boltzmann Machine (FVBM) models. The FVBM was first considered in Hyvarinen (2006), and can be described as follows.

Let \( X \in \mathbb{X} \) (where \( \mathbb{X} = \{-1, +1\}^d \) is the set of spin-binary strings of dimension \( d \)) be a random variable with probability mass function (PMF) of the form

\[
f(x; \theta) = \mathbb{P}(X = x) = \exp \left( \frac{1}{2} x^\top M x + b^\top x \right) / z(\theta),
\]

where

\[
z(\theta) = \sum_{\xi \in \mathbb{X}} \exp \left( \frac{1}{2} \xi^\top M \xi + b^\top \xi \right),
\]

is a normalizing constant, \( b \in \mathbb{R}^d \), and \( M \in \mathbb{R}^{d \times d} \) is a symmetric matrix with zero diagonal. We put the unique elements of the bias vector \( b \) and the interaction matrix \( M \) into the parameter vector \( \theta \). Interpretations of the parameter vector elements of an FVBM can be found in Section 2 of Bagnall, Jones, Karavarsamis, & Nguyen (2018).

The FVBM model can be viewed as a multivariate extension of the classical Bernoulli distribution, on spin-binary random variable strings (i.e, random variables \( X \in \mathbb{X} \), rather \( X \in \{0, 1\}^d \)), and can be demonstrated to be equivalent to the logistic multivariate binary model that was proposed by Cox (1972). It can also be considered as a fully-connected binary graphical model (see Bagnall et al., 2018 for more details). Simply put, the FVBM is a model that allows for the characterization of interactions between multiple binary random variables, simultaneously.

Let \( X_1, \ldots, X_n \in \mathbb{X} \) be a sample of \( n \) independent and identically distributed (IID) observations from an FVBM with unknown parameter vector \( \theta_0 \). In Hyvarinen (2006), it was proved that \( \theta_0 \) could be consistently estimated from \( X_1, \ldots, X_n \in \mathbb{X} \) via the so-called maximum pseudolikelihood estimator (MPLE) of the type described in Lindsay (1988). Furthermore, in H. D. Nguyen & Wood (2016a), a minorization-maximization (MM; see Hunter & Lange, 2004; H. D. Nguyen, 2017) was proposed for the computation of the MPLE. The asymptotic normality of the MPLE was proved in H. D. Nguyen & Wood (2016b), which allows for the use of the MPLE of the FVBM as both a point estimator and a method for constructing useful hypothesis tests regarding relationships between binary random variables.
The BoltzMM package makes the technical results above accessible. The most fundamental of the functions of BoltzMM are pfvbm and rfvbm, which allow for the computation of the probability of any binary string, given valid input bias vector and interaction matrix, and the random generation of binary strings, given input parameter elements, respectively. Since the support $X$ is finite, it is often desirable to compute the probabilities of all possible outcomes, which can be achieved using the function allpfvbm. It is also often desirable to compute the marginal probability of any particular string index

$$f_i(x_i; \theta) = P(X_i = x_i),$$

where $i = 1, \ldots, d$, $X_i \in \{-1, 1\}$, and $X^T = (X_1, \ldots, X_d)$, for any parameter vector. This can be achieved using the function marginpfvbm.

The function fitfvbm implements the MM algorithm of H. D. Nguyen & Wood (2016a) in order to compute the MPLE of an unknown FVBM parameter vector $\theta_0$, from an $n$ observation realization of the sample $X_1, \ldots, X_n$ (i.e., $x_1, \ldots, x_n$). That is, fitfvbm takes in the data as a $n \times d$ dimensional matrix, where each row is a spin-binary string of length $d$. Using the fitfvbm function output, and the asymptotic normality results of H. D. Nguyen & Wood (2016b), standard errors can be computed using the function fvbmsderr, with the functions fvbmcov, fvbmlHess, and fvbmpartiald providing intermediate calculations required by fvbmsderr.

As described in H. D. Nguyen & Wood (2016b) and demonstrated in Bagnall et al. (2018), hypothesis testing can be conducted using Wald-type hypothesis tests. These Wald-type tests can be constructed for testing of hypotheses of the form

$$H_0: \theta_{0k} = \theta_k, \text{ versus } H_1: \theta_{0k} \neq \theta_k,$$

where $\theta_{0k}$ is the $k$th element of the unknown parameter vector $\theta_0$, and $\theta_k$ is some hypothesized value that it may take. The $p$-values and $z$-scores for these hypothesis tests can be computed using the outputs from fitfvbm and fvbmsderr, via the function fvbmtests.

All of the functions have been documented, and the documentation and examples for each function can be accessed via the help function in R. The analysis of data arising from the voting patterns of the Senate of the 45th Australian Parliament in 2016, using the BoltzMM package, can be found in Bagnall et al. (2018). The report describes the manner in which the FVBM can be used to conduct inference regarding multivariate binary data. The data that were analysed in the report are included in the package and can be accessed via the command data(senate).

The BoltzMM package is programmed in native R, with computationally intensive subroutines programmed in C and integrated via the Rcpp and RcppArmadillo packages of Eddelbuettel (2013). The package is therefore suitable for analysis of large data sets. On https://cran.r-project.org, the only other package that provides facilities for estimation of Boltzmann machine (BM) models, of any kind, is the deepnet package of Rong (2014). Here, facilities for estimation of the so-called restricted Boltzmann machine (RBM) are included, via the functions rbm.down, rbm.train, and rbm.up.

The RBM is an alternative simplification, to the FVBM, of the BM models first considered by Ackley, Hinton, & Sejnowski (1985). The BM is a parametric probabilistic model that can closely approximate any PMF on $X$ (see, e.g., Le Roux & Bengio, 2008). Unfortunately, due to the intractability of its requirement for hidden (or unobserved) variables, it is often difficult to efficiently and meaningfully conduct parameter estimation when both dimension $d$ and sample size $n$ are large. Like the BM models, estimation of RBM models is also intractable, due to the requirement for hidden random variables. These hidden variables also prevent the establishment of asymptotics for the RBM parameter estimators, which makes the statistical inference of such models limited. Thus we view the functions from deepnet as being complementary but not overlapping with BoltzMM.
Users can obtain the latest build of BoltzMM on GitHub. The latest stable build can be obtained from CRAN (https://CRAN.R-project.org/package=BoltzMM), and an archival build can be obtained from Zenodo. Algorithm derivations and theoretical results regarding the implemented functions can be found in H. D. Nguyen & Wood (2016b) and H. D. Nguyen & Wood (2016a). An example application of the BoltzMM package for the analysis of real data is described in Bagnall et al. (2018). Thorough descriptions of the package functions appear in the manual, which can be accessed at https://cran.r-project.org/web/packages/BoltzMM/BoltzMM.pdf. Bug reports and other feedback can be directed to the GitHub issues page.

Acknowledgements

Jessica Bagnall is funded by a Research Training Program Stipend scholarship from La Trobe University. Hien Nguyen is funded under Australian Research Council (ARC) grant numbers DE170101134 and DP180101192.

References


