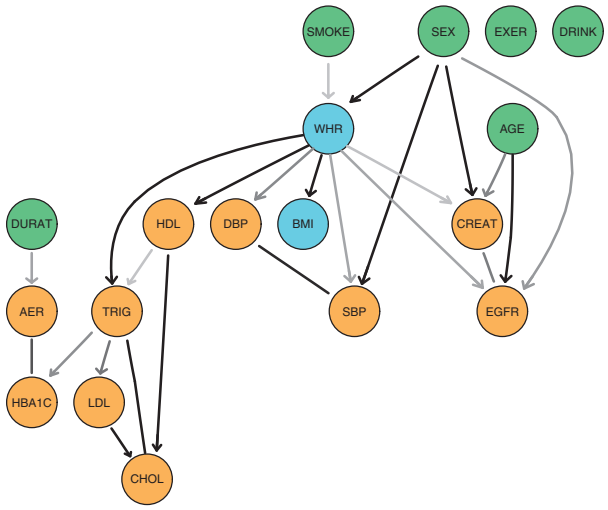
Learning the topology of a Bayesian Network
from a database of cases using the K2 algorithm

A Bayesian belief-network [1] structure is a directed acyclic graph in which nodes represent domain variables and arcs between nodes represent probabilistic dependencies [2]. Given a database of records, it is interesting to construct a probabilistic network which can provide insights into probabilistic dependencies existing among the variables in the database. Such network can be further used to classify future behaviour of the modelled system [2]. Although researchers have made substantial advances in developing the theory and application of belief networks, the actual construction of these networks often remains a difficult, time consuming task. An efficient method for determining the relative probabilities of different belief-network structures, given a database of cases and a set of explicit assumptions is described in [2] and [3].



The K2 algorithm [3] can be used to learn the topology of a Bayes network [2], i.e. of finding the most probable belief-network structure, given a database.

Part 1 After having studied the problem in the suggested literature ([2]-[3]), Implement the algorithm in R and check its performances with the test data set given in [3].

Part 2 Implement and test the K2 algorithm with the test data sets ([3]). Investigate if it is possible to code it inside the `bnstruct` R package [4]-[5].

Bibliography

[1] M. Scutari and J. B. Denis, *Bayesian Networks*, CRC Press, 2022, Taylor and Francis Group.

[2] G. F. Cooper and E. Herskovits, *A Bayesian Method for the Induction of Probabilistic Networks from Data*, Machine Learning **9**, (1992) 309

[3] C. Ruiz, *Illustration of the K2 Algorithm for learning Bayes Net Structures*, `http://web.cs.wpi.edu/~cs539/s11/Projects/k2_algorithm.pdf`

[4] A. Franzin et al., *`bnstruct`: an R package for Bayesian Network structure learning in the presence of missing data*, Bioinformatics **33(8)** (2017) 1250

[5] F. Sambo and A. Franzin, *`bnstruct`: an R package for Bayesian Network Structure Learning with missing data*, December 12, 2016