School of Engineering and Applied Science (SEAS), Ahmedabad University

Probabilistic Graphical Models (CSE 516)

Project Abstract

Deadline : 23-10-2022 11:59 PM

Group Name: r-ads-g5
Team Members:

- Kalp Ranpura(AU1920134)
- Kshitij Shekhar(AU2020165)
- Nihar Jani(AU2040205)

Algorithm: Bayesian Networks

I. Summary

We will be working on creating a Bayesian belief Network trained Cleveland heart disease data set and to know its effectiveness in predicting heart diseases. We will use Bayesian Network modelling to discover the relationship between 14 relevant attributes of the data collected from the UCL repository. The secondary goal is to discover how the dependencies between attributes affect the performance of the model. Bayesian Network modelling has the ability to predict new scenarios and also produce a reliable and transparent graphical modelling to better understand the data. We will be using the structural learning and parameter learning operations of the Bayesian network to build our model. The methodology used would include a number of steps ranging from Data retrieval to handling missing values and the transformation of target class attribute. We will then compare the obtained results of this model with existing classifiers to see how well the model performs comparatively.

II. Abstract

A. Background

Globally, cardiovascular disease is the leading cause of death. Finding useful information from the data that is already available can be aided by data mining from the medical field. Cardiovascular disorders are brought on by heart and blood vessel dysfunction. One-fifth of fatalities have been connected to cardiac causes, and an arrhythmic substrate is the site of the interaction. Using data mining, one can find relevant trends in a vast quantity of is a data. Understanding data and making discoveries from it are the most crucial tasks in data mining. To determine which individual has a higher risk of experiencing a heart attack based on several characteristics, we use Bayesian networks. It belongs to the category of probabilistic graphical models. It is a node and edge-based directional acyclic graph.

Some Basic Definition in Bayesian Belief Network. A path in between vertices A and B in a Bayesian Network is blocked if it passes through a vertex C in a way that either:

- Serial Connection ((A→C→B) or ((A←C←B)) or diverging (A←C→B) and C is conditioned on
- Converging (A→C←B) and neither

B. Motivation

Cardiovascular Diseases are the leading cause of death globally, especially among developed, high-income countries with advanced, standards of living and (ironically) access to high-quality healthcare. Many factors contribute to CVD, and a large majority of them cannot be understood by a layman. One of the many advantages and trickle-down effects of advances in AI and its derivatives(like probabilistic graphical models) is the democratization of data and its applications. While classic ML models like SVMs and Random Forests are able to solve problems like the prediction of heart disease given some features, they are often of limited use as they are often not very explainable. Deep Learning models often produce state-of-the-art results for almost any problem that has enough data(enough is an understatement, they often require thousands of data points to produce state-of-the-art results), but are like black boxes - we don't understand how they come to make the correct predictions. But probabilistic models like Bayesian Networks, while limited in their scope of applications pride themselves on interpretability. The project entails finding the relationships between 14 features of the dataset and performing inference. A Bayes Net would be able to show us the conditional dependencies and independencies among the features of the dataset.

C. Contribution

The paper's authors implemented the Bayes Net in R using the bnlearn library. We will implement the same in pgmpy and attempt to match and/or beat the performance metrics the authors were able to produce.

III. References

[1] Muibideen, M., amp; Prasad, R. (2020, December 17). A fast algorithm for heart disease prediction using Bayesian network model. arXiv.org. Retrieved October 23, 2022, from https://arxiv.org/abs/2012.09429v1