Learning Probabilistic Models for Cardiovascular Data

Kate Sanders and Robert Long

July 2017

16.5 million Americans over the age of 20 currently suffer from Coronary Heart Disease, or CHD. It is the leading cause of death in the United states. We asses the issue of learning probabilistic models from a cardiovascular study. Specifically, we examine the Coronary Artery Risk Development in Young Adults (CARDIA) study and strived to learn a joint distribution over several risk factors and Coronary Artery Calcification (CAC), a key indicator that the risk of CHD is sub-clinically present in an individual.

In the artificial intelligence community, Bayesian Networks (BN) are deemed as popular tools for effectively modeling the joint distribution across several related variables. We went about learning these BNs on CARDIA data and indicate that the resulting BN encapsulates some interesting correlations between risk factors. The fascinating associations that we observed were, sex and BMI influencing glucose and sex and glucose influencing CAC. The extent of our research has the potential to lead to developing individualized treatment plans for improving population health.

"Although, the socioeconomic relationships we observed such as, race and education influencing smoking, were not directly related to CHD it was still very intriguing to analyze the trends portrayed in the Conditional Probability tables (CPT). For example, we saw that as higher education is completed the probability of a person being a nonsmoker increases," said co-author Robert Long.

Here is a link to our video presentation
https://www.youtube.com/watch?v=_WLAiKQiZws&t=3s

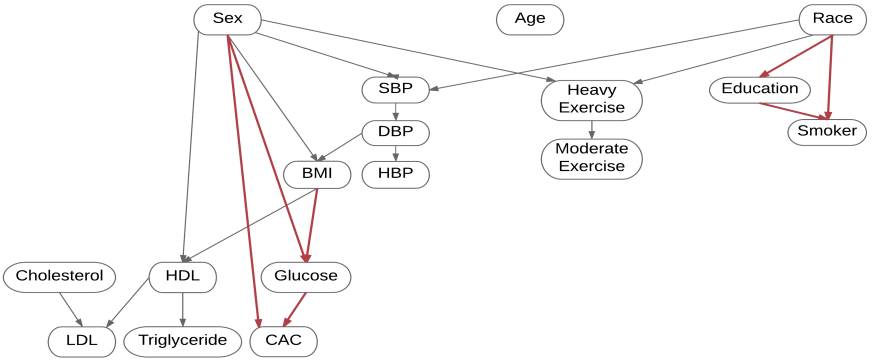

Figure 1: Intersection of models learned using the AIC, BIC, and BDe scoring metrics in the Hill-Climbing algorithm.

1 Contact Information

Kate Sanders
Email: SandersKM@hendrix.edu
Telephone: 870-613-4511

Robert Long
Email: robmlong95@gmail.com
Telephone: 219-713-6044