Report for AI3603: Homework III

Chen Jingtao

November 30th, 2025

Contents

1 Code Implementation of BayesianNetworks

Fill in the blank at *BayesianNetworks.py*, complete the functool that is needed for analyzing data using bayesian network.

In *jointFactors()*, we consider two kinds of conditions. When two factors have no common columns, we consider the "cross" method in merge(); and when they have common columns, we treat the joint as a conditional inference.

In *marginalizeFactor()*, we use groupby().sum() method to merge those with same not-hiddenvar-columns.

In *evidenceUpdateNet()*, for each factor inside the net, we select the row that suit the evidence.

In *inference()*, we firstly use *evidenceUpdateNet()* to select out the evident rows, and then do *jointFactors()* among all factors, finally do *marginalizeFactor()* to conclude the probs we get.

2 Written Part

In this section we analyze the given RiskFactorsData.

The detail implementation is written in *Risk_analyze.ipynb*.xw

2.1 Problem 1

The size of the bayes net is 1048.

2.2 Problem 2

(a)If you have bad habits, the probabilities that you have each health problem is as follows:

- diabetes: 0.1796845567127721

- stroke: 0.053783054506654104

- attack: 0.08560619240891772

- angina: 0.0950935453205552

If you have good habits, the probabilities that you have each health problem is as follows:

- diabetes: 0.07440604734997436

- stroke: 0.02939015003013203

- attack: 0.036246688444733055

- angina: 0.03480791596138738

(b)If you have bad health, the probabilities that you have each health problem is as follows:

- diabetes: 0.11442040761053557

- stroke: 0.08373575429785589

- attack: 0.14107269332303135

- angina: 0.1610649668405125

If you have good health, the probabilities that you have each health problem is as follows:

- diabetes: 0.056980800262824795

- stroke: 0.014633237021157612

- attack: 0.015994034858793926

- angina: 0.013123310653369375

2.3 Problem 3

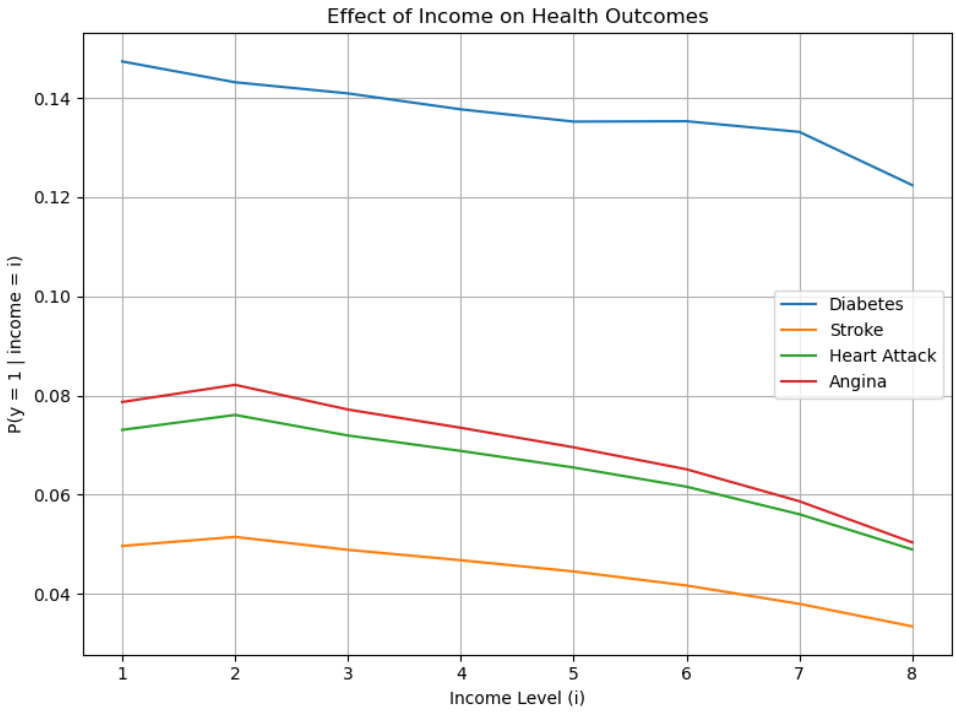

Figure 1: Pronlem 3

From the plot, I can tell two facts:

(1) With higher income level, probabilities of all four kinds of diseases go down, so better income may contribute to less disease.

(2) Over the four kinds of diseases among all levels of income, diabetes occur the most frequently, whilc stroke occur the less frequently.

2.4 Problem 4

(a)If you have bad habits, the probabilities that you have each health problem is as follows:

- diabetes: 0.2466907421997436

- stroke: 0.08088181221862281

- attack: 0.13563231125420805

- angina: 0.13804114406095389

If you have good habits, the probabilities that you have each health problem is as follows:

- diabetes: 0.05576155616484895

- stroke: 0.019589809785240563

- attack: 0.021058123344538895

- angina: 0.023596047799734423

(b)If you have bad health, the probabilities that you have each health problem is as follows:

- diabetes: 0.12013390316425754

- stroke: 0.08374005218091235

- attack: 0.140240442199526

- angina: 0.16041693839228313

If you have good health, the probabilities that you have each health problem is as follows:

- diabetes: 0.05517615493861087

- stroke: 0.014719065353818117

- attack: 0.016014050599026382

- angina: 0.013126217424413758

The effect: we can see in (b) it have not many change, but in (a) we can see the probs go bigger for bad-habits state and go smaller for good-habits, which means the affect on health outcome from habits becomes bigger.

It's valid, since given the state of either habits or health, we may find that still four kinds of health outcome is conditionally independent. By the result we can also see that the probs-trend is similar to what we get in question(2).

2.5 Problem 5

In the q-4-net, we get

$$P(stroke = 1|diabetes = 1) = 0.04486338128501227 \tag{1}$$

$$P(stroke = 1|diabetes = 3) = 0.04032301820310427 \tag{2}$$

In the q-5-net, we get

$$P(stroke = 1|diabetes = 1) = 0.0774910969238895 \tag{3}$$

$$P(stroke = 1|diabetes = 3) = 0.03483211249742304 \tag{4}$$

The effect: the first prob go bigger and the second one go smaller. In another words, if one has diabetes, he/she is more possible to have stroke; and if one hasn't, he/she is also less possible to have stroke.

This is not valid, since the health outcome become dependent over probabilities.