AU 332 Artificial Intelligence: Principles and Techniques

By: Xiaocheng Wang(517030910326)

HW#: 4

November 28, 2019

I. INTRODUCTION

II. CODING PART

In this part I will implement code for computing exact inferences in Bayesian networks of discrete random variables using variable elimination. Here are five functions to be completed.

A. joinFactors(factor1, factor2)

Should return a factor table that is the join of factor 1 and 2. I use pd.merge to merge two dataframes, and update the probabilies with the product of corresponding probabilies of older dataframes. I specifically handle with the case when the two factors share no common column. In this case, I suppose they are independent, so the joint distribution is the product.

B. marginalizeFactor(factorTable, hiddenVar)

This function should return a factor table that marginalizes margVar out of it. Assume that hiddenVar is on the left side of the conditional. Therefore, we can directly use pd.groupby to leave out the hidden variable and sum them up by mean() method.

C. marginalizeNetworkVariables(bayesNet, hiddenVar)

This function takes a Bayesian network, bayesNet, and marginalizes out a list of variables hiddenVar. For every hidden variable, we find all the related conditional distribution tables in the network, and replace them with a new factor that maginalizes the hidden variable. The new factor is contructed by join all distributions in the network which contains the hidden variable, and sum them up on the hidden variable axis.

D. evidenceUpdateNet(bayesNet, evidenceVars, evidenceVals)

This function takes a Bayesian network, bayesNet, and sets the list of variables, evidenceVars, to the corresponding list of values, evidenceVals. For distributions in the bayes network, we only choose rows from them which have corresponding values for the evidence variable.

E. inference(bayesNet, hiddenVar, evidenceVars, evidenceVals)

This function takes in a Bayesian network and returns a single joint probability table resulting from the given set of evidence variables and marginalizing a set of hidden variables. The final probabilities should be normalized.

III. WRITTEN PART

1. (a).The size (in terms of the number of probabilities needed) of this network is the sum of the sizes of each table in the network. In the given bayes network, this number is 504.

   (b). The full joint distribution size is 32768.

2. (a) If you have bad habits (smoke and don't exercise):

	probs	exercise	smoke	diabetes
0	0.150516	2	1	1
1	0.008965	2	1	2
2	0.822423	2	1	3
3	0.018096	2	1	4

(a)diabetes

	probs	exercise	smoke	attack
0	0.07433	2	1	1
1	0.92567	2	1	2

(b)heart

	probs	exercise	smoke	stroke
0	0.049264	2	1	1
1	0.950736	2	1	2

(c)stroke

	probs	exercise	smoke	angina
0	0.080448	2	1	1
1	0.919552	2	1	2

(d)angina

FIG. 1: Healthy problems with bad habits.

If you have good habits (don't smoke and do exercise):

	probs	exercise	smoke	diabetes
0	0.127119	1	2	1
1	0.008865	1	2	2
2	0.847693	1	2	3
3	0.016323	1	2	4

(a)diabetes

	probs	exercise	smoke	attack
0	0.052798	1	2	1
1	0.947202	1	2	2

(b)heart

	probs	exercise	smoke	stroke
0	0.03611	1	2	1
1	0.96389	1	2	2

(c)stroke

	probs	exercise	smoke	angina
0	0.054755	1	2	1
1	0.945245	1	2	2

(d)angina

FIG. 2: Healthy problems with good habits.

(b) If you have poor health (high blood pressure, high cholesterol, and overweight):

	probs	bmi	diabetes	cholesterol	bp
0	0.115423	3	1	1	1
1	0.007662	3	2	1	1
2	0.860873	3	3	1	1
3	0.016043	3	4	1	1

(a)diabetes

	probs	cholesterol	bp	bmi	attack
0	0.140784	1	1	3	1
1	0.859216	1	1	3	2

(b)heart attack

	probs	cholesterol	bp	bmi	stroke
0	0.082686	1	1	3	1
1	0.917314	1	1	3	2

(c)stroke

	probs	cholesterol	bp	bmi	angina
0	0.161608	1	1	3	1
1	0.838392	1	1	3	2

(d)angina

FIG. 3: Healthy problems with poor health.

If you have good health (low blood pressure, low cholesterol, and normal weight):

	probs	bmi	diabetes	cholesterol	bp
0	0.057710	2	1	2	3
1	0.009543	2	2	2	3
2	0.922194	2	3	2	3
3	0.010553	2	4	2	3

(a)diabetes

	probs	cholesterol	bp	bmi	attack
0	0.016161	2	3	2	1
1	0.983839	2	3	2	2

(b)heart attack

	probs	cholesterol	bp	bmi	stroke
0	0.01446	2	3	2	1
1	0.98554	2	3	2	2

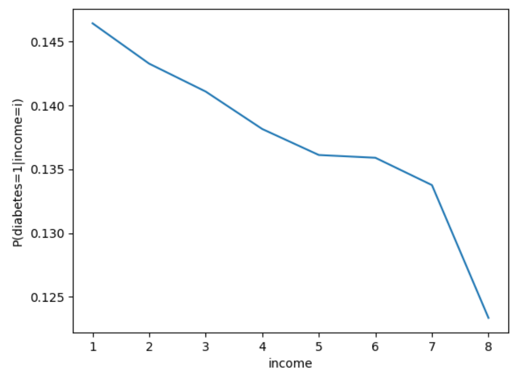
(c)stroke

	probs	cholesterol	bp	bmi	angina
0	0.013326	2	3	2	1
1	0.986674	2	3	2	2

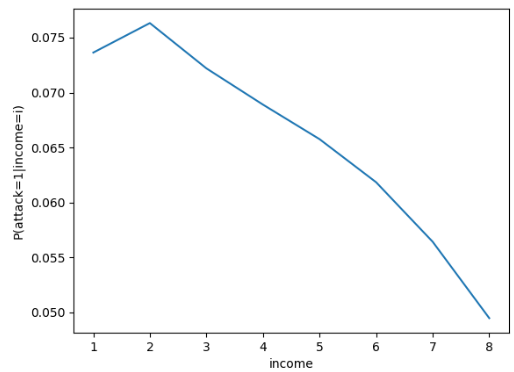
(d)angina

FIG. 4: Healthy problems with good health.

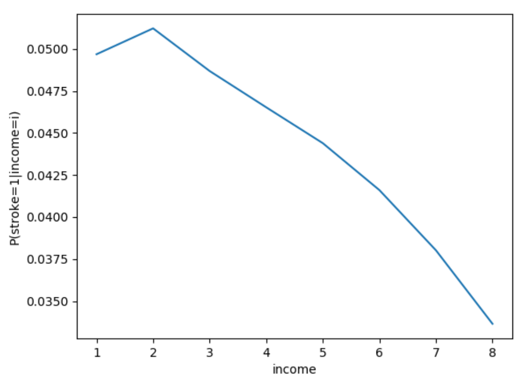
3. The plots are shown below:



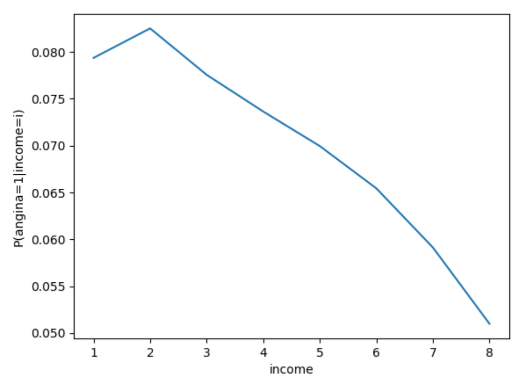
(a)diabetes



(b)heart attack



(c)stroke



(d)angina

FIG. 5: relationship between income and health problems

According to the relationships between income and health problems, we can conclude that the more money one earns, the more health he loses.

4. This makes assumption that smoking and exercise have no influence on health problems. Add edges from smoke and exercise to the health outcomes respectively, and redo the queries in Q2, and see the results below. When smoking and don't exercise:

	probs	exercise	smoke	diabetes
8	0.210945	2	1	1
9	0.006915	2	1	2
10	0.760693	2	1	3
11	0.021447	2	1	4

(a)diabetes

	probs	exercise	smoke	attack
0	0.121166	2	1	1
1	0.878834	2	1	2

(b)heart attack

	probs	exercise	smoke	stroke
0	0.078035	2	1	1
1	0.921965	2	1	2

(c)stroke

	probs	exercise	smoke	angina
0	0.119007	2	1	1
1	0.880993	2	1	2

(d)angina

FIG. 6: New network, bad habits.

When no smoking and exercise:

	probs	exercise	smoke	diabetes
0	0.098552	1	2	1
1	0.009884	1	2	2
2	0.877576	1	2	3
3	0.013988	1	2	4

(a)diabetes

	probs	exercise	smoke	attack
0	0.031015	1	2	1
1	0.968985	1	2	2

(b)heart attack

	probs	exercise	smoke	stroke
0	0.024311	1	2	1
1	0.975689	1	2	2

(c)stroke

	probs	exercise	smoke	angina
0	0.0368	1	2	1
1	0.9632	1	2	2

(d)angina

FIG. 7: New network, good habits.

Compared with the results in Q2, smoking and exercise make the probabilites changes dramatically under the new bayes network. Therefore, the assumption of the first graph is invalid.

5. This assumes that the four health problems are independent, and do not influence each other. In the new network, we evaluate $P(stroke = 1|diabetes = 1)$ and $P(stroke = 1|diabetes = 3)$, which is shown below. It's illustrated that diabetes does have a positive correlation relationship with stroke. So the assumption is invalid.

	probs	diabetes	stroke
0	0.044164	1	1
1	0.955836	1	2

(a)$P(stroke|diabetes = 1)$

	probs	diabetes	stroke
0	0.040478	3	1
1	0.959522	3	2

(b)$P(stroke|diabetes = 3)$

FIG. 8: No edge from diabetes to stroke.

	probs	diabetes	stroke
0	0.076198	1	1
1	0.923802	1	2

(a)$P(stroke|diabetes = 1)$

	probs	diabetes	stroke
0	0.035015	3	1
1	0.964985	3	2

(b)$P(stroke|diabetes = 3)$

FIG. 9: With edge from diabetes to stroke.