平成 30 年度 知識工学 第 3 回 レポート課題

2019 年 1 月 22 日

二宮 崇

次の問題 1～6 を解き、その解答を提出せよ。提出するレポートには、所属、学生証番号、名前を一番上に記入せよ。

提出方法：レポートを講義の時に直接提出。

締切り： 2019 年 1 月 29 日（火） 14:30

問題1

次の完全結合確率分布が与えられているとする。次の(a)～(d)を計算せよ。

$Toothache$(歯痛), $Cavity$(虫歯), $Catch$(歯医者の不快なさぐり針が痛い歯にひっかかる)という三つの論理確率変数に対する完全結合分布

$Toothache$	$Cavity$	$Catch$	P
0	0	0	0.576
0	0	1	0.144
0	1	0	0.008
0	1	1	0.072
1	0	0	0.064
1	0	1	0.016
1	1	0	0.012
1	1	1	0.108

(a) $P(toothache)$

(b) $P(Cavity)$に対する確率分布を求めよ。

(c) $P(Toothache|cavity)$に対する確率分布を求めよ。

(d) $P(Cavity|toothache \vee catch)$ に対する確率分布を求めよ。

問題 2　次の(a)～(c)について答えよ。

(a) $P(a \wedge b|c) = P(a|b \wedge c)P(b|c)$を証明せよ

(b) $P(b|a \wedge c) = \frac{P(a|b \wedge c)P(b|c)}{P(a|c)}$を証明せよ。

(c) 次の式(1)～(3)が全て同値であることを証明せよ。(式(1)～(3)がすべて同値であるとは、式(1)⇔式(2), 式(1)⇔式(3), 式(2)⇔式(3)が成り立つこと)

$P(a|b) = P(a)$ … (1)

$P(b|a) = P(b)$ … (2)

$P(a \wedge b) = P(a)P(b)$ … (3)

問題3　肩こりが生じる事前確率が1/20で、髄膜炎が生じる事前確率が1/50,000とする。髄膜炎が起きたときには50%の確率で肩こりが発生することが知られているとする。肩こりが発生したとき、その原因が髄膜炎である確率を求めよ。

問題4　ある文書(例えばテキストだけ取り出したウェブページやe-メールなど)が与えられた時、それらのカテゴリ(政治、スポーツ、社会などのジャンルや、スパムと非スパムなど)を自動的に推論することは文書分類と呼ばれ、単純ベイズがよく用いられる。次の(a)～(b)に答えよ。

(a) 文書は単語列$w_1, \cdots, w_n$からできているとする。あるカテゴリが与えられたとき各単語は独立して出現することを仮定する。カテゴリを原因($Cause$)、各単語を結果($Effect$)とみなした単純ベイズを用いて、あるカテゴリcと単語列$w_1, \cdots, w_n$に対する同時確率$P(c, w_1, \cdots, w_n)$を求めよ。カテゴリの事前確率$P(c)$と、カテゴリcから単語wが出現する確率$P(w|c)$を用いて良い。

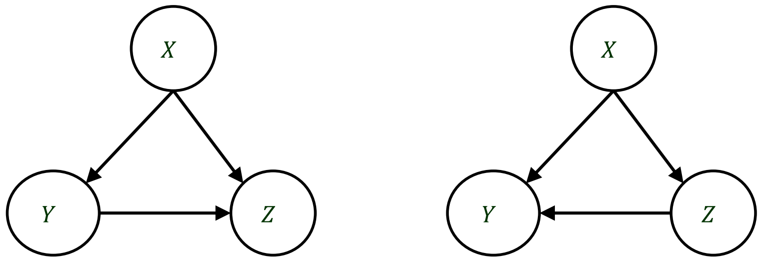
(b) カテゴリ集合をCとする。(a)で求めた式を用いると、ある文書(=単語列$w_1, \cdots, w_n$)が与えられたときのカテゴリは次の式で推定することができる。

$$\hat{c} = \underset{c \in C}{argmax}\, P(c, w_1, \cdots, w_n)$$

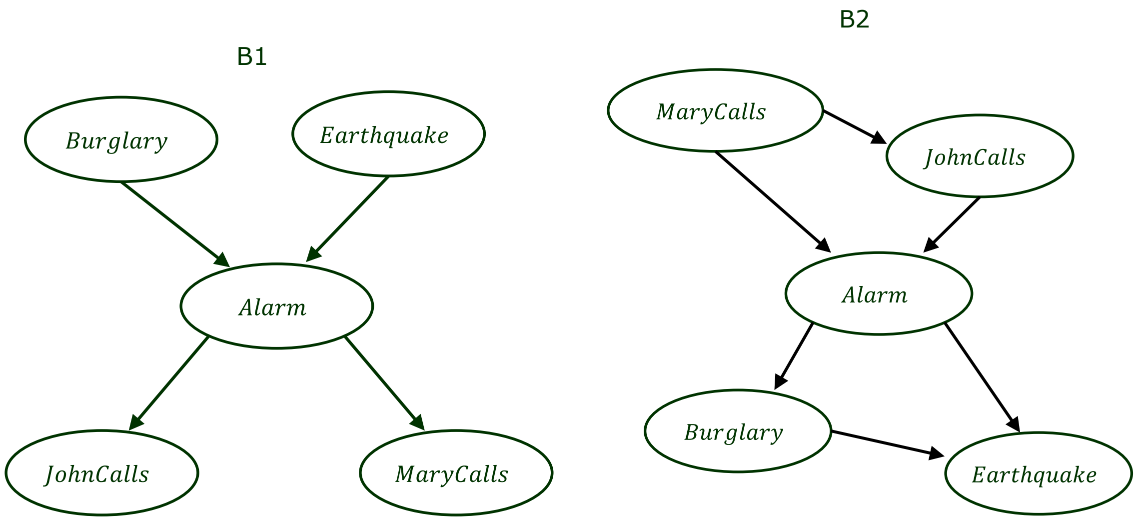
つまり、$P(c)$と$P(w|c)$が既知ならば、各カテゴリの中で$P(c, w_1, \cdots, w_n)$を最大とするカテゴリが予想されるカテゴリとなる。そこで、正解カテゴリ付の文書集合Dが与えられたとして、Dから$P(c)$と$P(w|c)$を求めることを考える。文書集合Dに含まれる文書数を$|D|$、Dに含まれる文書のうちカテゴリcである文書の数を$Count(c)$、カテゴリcに属する全文書中に単語wが出現した回数を$Count(w, c)$、語彙集合をVとしたとき、$P(c)$と$P(w|c)$を求めよ。

問題5　次の(a)～(b)について答えよ。

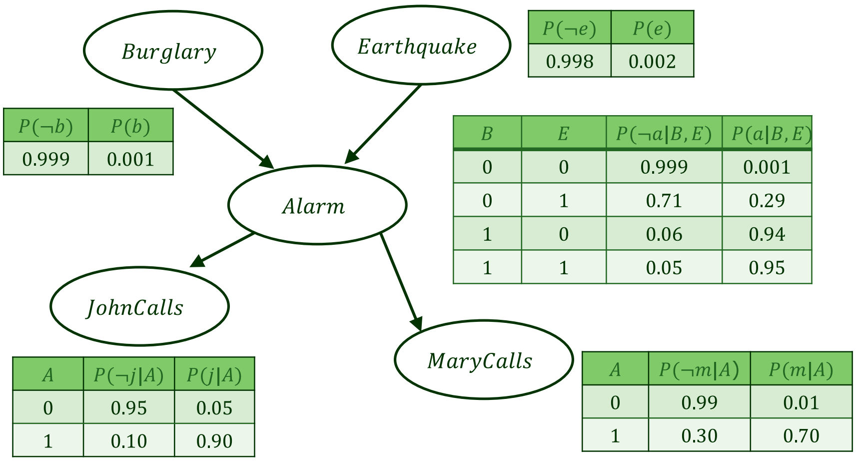
(a) 次の二つのベイジアンネットが等価であることを示せ。



(b) 次の二つのベイジアンネットB1とB2について考える。B1が与えられたとき、B1からB2が導出されることを示せ。



問題６　次の防犯アラームに関するベイジアンネットについて考える。



$P(\neg b)$	$P(b)$
0.999	0.001

$P(\neg e)$	$P(e)$
0.998	0.002

B	E	$P(\neg a\|B,E)$	$P(a\|B,E)$
0	0	0.999	0.001
0	1	0.71	0.29
1	0	0.06	0.94
1	1	0.05	0.95

A	$P(\neg j\|A)$	$P(j\|A)$
0	0.95	0.05
1	0.10	0.90

A	$P(\neg m\|A)$	$P(m\|A)$
0	0.99	0.01
1	0.30	0.70

$Burglary, Earthquake, Alarm, JohnCalls, MaryCalls$は論理確率変数で、それぞれ$B, E, A, J, M$と略記することにする。また、各論理確率変数に対し、$X=1$はx、$X=0$は$\neg x$と表記することにする。次の（a）～（c）の問いに答えよ。

（a） $P(B,E,A,j,m)$を求めよ。（B,E,Aに対する確率分布の表を作成すればよい）

（b） （a）の結果を用いて、$P(B|j,m)$を求めよ。（Bに対する確率分布の表を作成すればよい）

（c） AlarmのCPTをNoisy-ORで表現することを考える。AlarmのCPTに関して次の式だけが与えられたとする。

$$P(\neg a|\neg b, e) = 0.71$$
$$P(\neg a|b, \neg e) = 0.06$$

このとき、次の表を埋めて、Noisy-ORで表現したAlarmのCPTを完成させよ。

B	E	$P(\neg a\|B,E)$	$P(a\|B,E)$
0	0		
0	1	0.71	
1	0	0.06	
1	1		