Data Mining Homework Set 2, 2020

Cursus: BETA-INFOMDM Data Mining (INFOMDM)

Aantal vragen: 5

Data Mining Homework Set 2, 2020

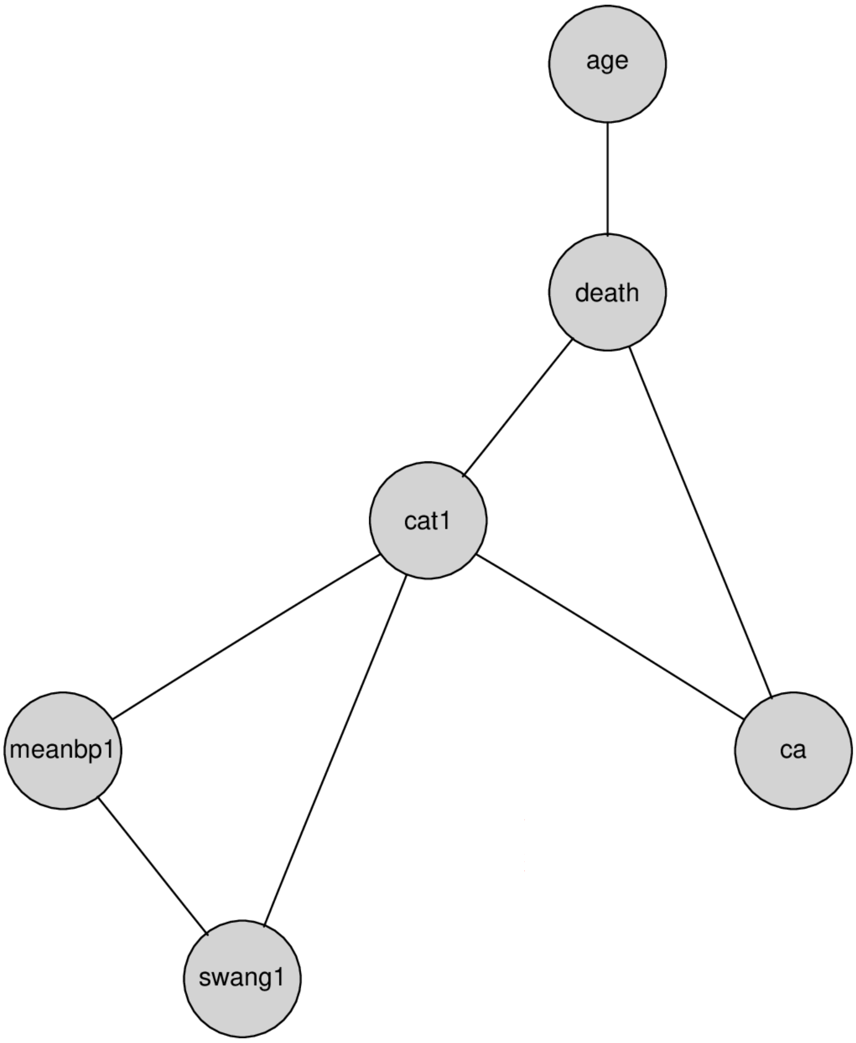
Cursus: Data Mining (INFOMDM)

This is Homework Set 2 of Data Mining

Aantal vragen: 5

1 Consider the graphical log-linear model with the following independence graph:

2 pt.



Which of the following (conditional) independences hold in this model?

a. age ⊥ cat1 — No, because there is a path connecting them in the graph.

b. swang1 ⊥ death | cat1 — Yes, because every path from swang1 to death passes through cat1. We also say that cat1 blocks every path between swang1 and death.

c. age ⊥ swang1 — No, because there is a path connecting them in the graph.

d. death ⊥ {meanbp1, swang1} | {age, cat1} — Yes, this is the local Markov property. death is independent of all remaining variables given the variables that are directly connected to death by and edge. The set {cat1, age} is also called the Markov blanket of death.

e. death ⊥ ca — No, because there is a path connecting them in the graph. In fact, ca and death are directly connected.

f. death ⊥ {meanbp1, swang1} | cat1 — Yes, because every path from {meanbp1,swang1} to death passes through cat1. We also say that cat1 blocks every path between swang1 and death.

g. swang1 ⊥ ca | meanbp1 — No, because there is a path connecting them in the graph, namely through cat1.

2 Consider the following table of counts on binary variables x and y:

n(x,y)	y=0	y=1
x=0	80	20
x=1	40	60

Suppose we fit the independence model x ⊥ y to this data. Give the fitted counts for:

(x=0,y=0):

a. ....60 (0,5 pt.)

(x=1,y=0):

b. ....60............. (0,5 pt.)

(x=0,y=1):

c. ....40............. (0,5 pt.)

(x=1,y=1):

d. ....40............. (0,5 pt.)

3 Consider a graphical model M on three binary variables A,B, and C, with independence graph G=(K,E) with K = {A,B,C} and E = {{B,C}}.

The observed counts are given in the following table:

A	B	C	n(A,B,C)
1	1	1	40
1	1	0	10
1	0	1	5
1	0	0	50
0	1	1	30
0	1	0	5
0	0	1	20
0	0	0	40

Answer the following questions (do not round your answer):

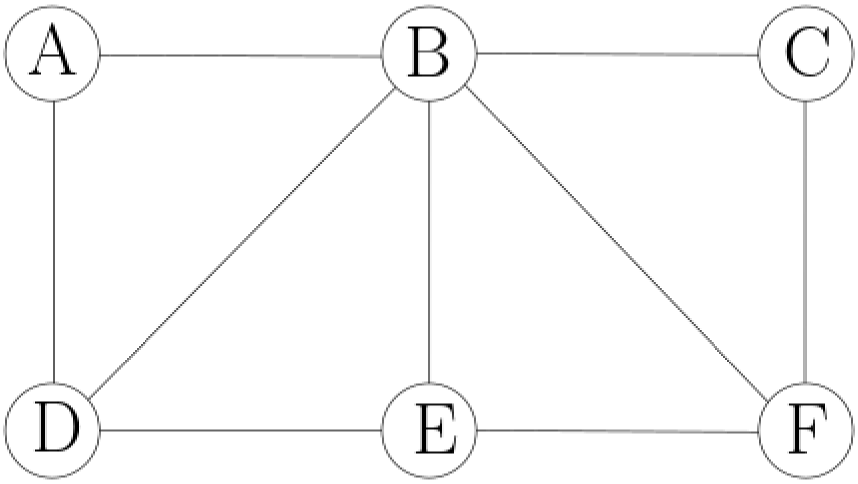
The fitted count $\hat{n}(1,1,1)$ according to model M is: **a.** ..(1 pt.) 36,75

The fitted count $\hat{n}(0,1,0)$ according to model M is: **b.** ..(1 pt.) 7,125

4 We are performing a hill-climbing search in the space of decomposable models. Neighboring models are obtained by either adding an edge to the current model, or removing an edge from the current model.

2 pt.

The current model is given in the following figure:



Which of the following operations produce a valid neighbor? (0 or more answers may be correct)

a. Remove the edge between B and D

b. Add an edge between A and E

c. Add an edge between A and F

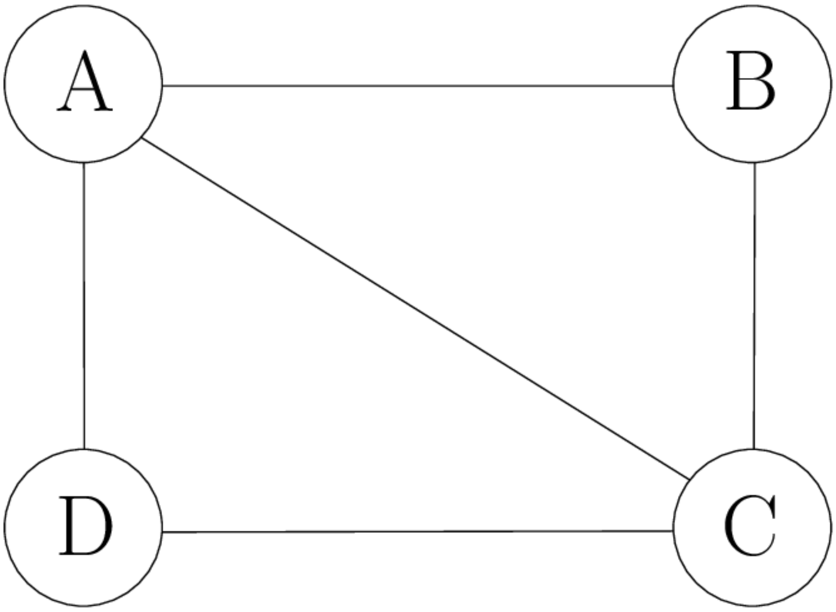
d. Remove the edge between B and F

e. Add an edge between C and D

f. Remove the edge between A and D

g. Remove the edge between B and E

5 Consider the graphical log-linear model M_1 on binary variables A,B,C, and D, with independence graph:



1 pt. **a.** The formula for the maximum likelihood fitted counts of M_1 is given by:

a. $$\frac{n(A,B,C)n(A,C,D)}{n(A)n(C)}$$

b. $$\frac{n(A,B,C)n(A,C,D)n(A,C)}{n(A)n(C)}$$

c. $$\frac{n(A,B,C)n(A,C,D)}{n(A,C)}$$

d. $$\frac{n(A,B)n(B,C)n(A,C)n(C,D)n(A,D)}{n(A)n(B)n(C)n(D)}$$

Consider the model M_0 obtained by removing the edge between A and C from M_1. How many parameters (u-terms) are eliminated by this change?

The number of eliminated u-terms is: **b.** ..(1 pt.) 4

Thank you, goodbye!