Project 1. Solution

Santiago Castro Dau, June Monge, Rachita Kumar, Sarus

2/25/2022

```
## Warning: package 'BiDAG' was built under R version 4.1.2
```

Problem 1: Conditional independence and BNs

a)

The condition $A \perp B|C$ holds for this Bayesian network a). Proof:

$$P(A,B,C) = P(A|C)P(B|C)P(C)$$

Consider

$$\begin{aligned} P(A,B|C) &= \frac{P(A,B,C)}{P(C)} \\ &= P(A|C)P(B|C) \\ &\implies A \perp B|C \end{aligned}$$

Hence Proved.

b)

The condition $A \perp B$ holds for this Bayesian network b). Proof:

$$P(A,B,C) = P(A)P(B)P(C|A,B)$$

$$P(C|A,B) = \frac{P(A,B,C)}{P(A,B)}$$

Substituting

$$\begin{aligned} P(A,B) &= P(A)P(B) \\ &\implies A \perp B \end{aligned}$$

Hence Proved.

Problem 2: Markov blanket

Problem 3: Learning Bayesian networks from protein data

a)

```r
## Set seed
set.seed(2022)

## Read data
data <- vroom(file = "sachs.data.txt", show_col_types = FALSE)

## Visualize data dimensions.
dim(data)
```

```
## [1] 853  11
```

- $N = 853$
- $n = 11$

```r
## Split data
data = data %>% rowid_to_column(var = 'row_id')
train = data %>% sample_frac(0.8, replace = FALSE)
test = anti_join(data, train, by = "row_id")

## Initializing socre object
score_object = scoreparameters(scoretype = 'bge', train)
```