CSE 643: Artificial Intelligence - Assignment 3

Akshat Gupta

November 13, 2023

Theory

1. Consider the following Random Variables

   - **T** $\rightarrow$ Person has travelled
   - **D** $\rightarrow$ Person has caught a disease other than Corona
   - **C** $\rightarrow$ Person has caught Corona disease
   - **M** $\rightarrow$ Person has a mild case of disease
   - **S** $\rightarrow$ Person has a severe case of disease
   - **W** $\rightarrow$ Person died

   When the RV is written in the lowercase, it indicates that RV is assigned the value $True$. Following is the representation of the sentences in probabilistic distribution form:

   (a) $P(t \wedge (d \vee c)) = 0.825$
   (b) $P((m \wedge c)|t) = 0.15$
   $P((s \wedge c)|t) = 0.22$
   (c) $P(d|t) = 0.485$
   (d) $P(w|(d \wedge t)) = 0.24$
   (e) $P(\neg t \wedge s \wedge c) = 0.025$
   (f) $P(s|\neg t) = 0.025$
   (g) $P(w \wedge c) = 0.059$
   (h) $P(m \vee s) = 0.7$
   (i) $P(t|s) = 0.8$
   (j) $P(c) = 0.5$

   The following observations can be made from the given propositions and the defined Random Variables:

   - C and D are mutually exclusive by definition $\Rightarrow P(c \vee d) = P(c) + P(d)$
   - S and M are mutually exclusive by definition $\Rightarrow P(m \vee s) = P(m) + P(s) = 0.7$

   The set of axioms that the distribution should satisfy are the Kolmogorov axioms:

   $$\sum P(x) = 1$$
   $$0 \leq P(x) \leq 1$$
   $$P(a \vee b) = P(a) + P(b) - P(a \wedge b)$$

2. Consider the doors to be labelled $D_1, D_2,$ and D_3, let W be the event of winning the key and L be the event of losing a life:

(a) Without loss of generality, say the initial choice was door D_2. Now, the person reveals one door that leads to loss of life; without loss of generality, say that door was D_1:

$$P(D_2 = W|D_1) = \frac{P(D_1|D_2 = W) \cdot P(D_2 = W)}{P(D_1)}$$

$$\text{Here } P(D_1) = \sum_{i=1}^{3} P(D_1, D_i = W)$$

$$\Rightarrow P(D_2 = W|D_1) = \frac{1/2 \text{ x } 1/3}{1/3 \text{ x } 0 + 1/3 \text{ x } 1 + 1/3 \text{ x } 1/2} = \frac{1}{3}$$

Clearly, the key must be behind either D_2 or D_3, since D_1 has been revealed.

$$\Rightarrow P(D_2 = W|D_1) + P(D_3 = W|D_1) = 1$$

$$\Rightarrow P(D_3 = W|D_1) = \frac{2}{3}$$

Thus, it is more favourable to switch the choice to D_3 in this case.

(b) In the case where the person can reveal the loss of life with probability $\frac{1}{3}$ and winning the key with probability $\frac{2}{3}$, let R be the event that D_1 is revealed:

$$P(D_2 = W|D_1, R) = P(D_2 = W|D_1 = L, R) \cdot P(D_1 = L) + P(D_2 = W|D_1 = W, R) \cdot P(D_1 = W)$$

$$P(D_2 = W|D_1, R) = \frac{P(D_1 = L, R|D_2 = W) \cdot P(D_2 = W)}{P(D_1 = L, R)} \cdot P(D_1 = L)$$

$$P(D_2 = W|D_1, R) = \frac{1/3 \text{ x } 1 \text{ x } 1/2}{1/3 \text{ x } 1} \cdot \frac{1}{3} = \frac{1}{6}$$

In the other case where $D_2 = L$:

$$P(D_2 = L|D_1, R) = P(D_2 = L|D_1 = L, R) \cdot P(D_1 = L) + P(D_2 = L|D_1 = W, R) \cdot P(D_1 = W)$$

$$P(D_2 = L|D_1, R) = \frac{P(D_1 = L, R|D_2 = L) \cdot P(D_2 = L)}{P(D_1 = L, R)} \cdot P(D_1 = L) + \frac{P(D_1 = W, R|D_2 = L) \cdot P(D_2 = L)}{P(D_1 = W, R)} \cdot P(D_1 = W)$$

$$P(D_2 = L|D_1, R) = \frac{1/2 \text{ x } 1/3 \text{ x } 2/3}{2/3 \text{ x } 1/3} \cdot \frac{2}{3} + \frac{1/2 \text{ x } 2/3 \text{ x } 2/3}{1/3 \text{ x } 2/3} \cdot \frac{1}{3}$$

$$P(D_2 = L|D_1, R) = \frac{2}{3}$$

The probability of winning when switching is still higher than the probability of winning when not switching. Hence, switching is still a better option.

(c) If choosing to switch and the man has revealed the door that leads to L, the conditional probability of winning will be $P(D_2 = L|D_1 = L, R)$. This, as was calculated in the previous parts, is $= \frac{1}{2}$.

(d) Assuming that the conditions of the problem where the man reveals the door which indicates L with probability $\frac{1}{3}$ holds, and that W assumes a value 1 and L assumes a value 0, and X models the prize:

$$E[X] = \sum_{x \in \{W, L\}} xP(x)$$

$$\Rightarrow E[X] = 1 \cdot P(W) + 0 \cdot P(L)$$

$$\Rightarrow E[X] = P(W)$$

$$\Rightarrow E[X] = P(W|\text{Switch}) \cdot P(\text{Switch}) + P(W|\text{No Switch}) \cdot P(\text{No Switch})$$

Assuming $P(\text{Switch}) = \frac{2}{3}$ and $P(\text{No Switch}) = \frac{1}{3}$ from (a):

$$\Rightarrow E[X] = \frac{2}{3} \cdot \frac{2}{3} + \frac{1}{6} \cdot \frac{1}{3}$$

$$\Rightarrow E[X] = 0.5$$

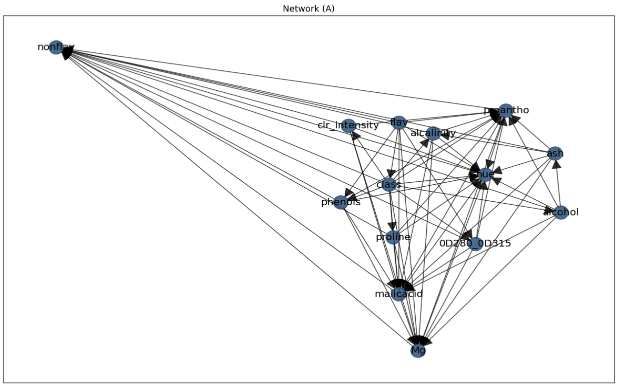

Figure 1: Network (A)

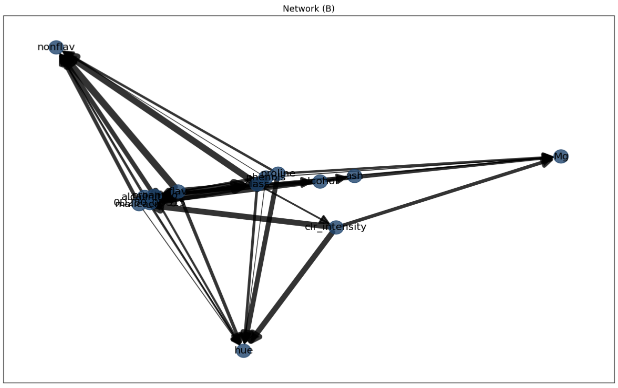

Figure 2: Network (B)

Computational

(a) The dataset was discretized using `sklearn.preprocessing.KBinsDiscretizer` where the number of bins (`n_bins`) were chosen based on the type of distribution each variable shows. No pre-processing other than discretization is needed for this task since the data is being used to build a Bayesian Network (BN) where the data values are not mathematically operated on; it is instead their frequency of occurrence which is of more significance and is the point of operation in the mathematical respect.

(b) Following BNs were constructed from the given data:

- **Network (A)** - The raw data was fed to the `bnlearn.structure_learning.fit()` function and was used to create a BN (see Figure 1). The probabilistic distributions were determined by fitting the model obtained from the previous step to the `bnlearn.parameter_learning.fit()` function and using the `bnlearn.print_CPD()` function to print all the respective probabilities.
- **Network (B)** - Network **(A)** was pruned using the `bnlearn.independence_test()` function which uses the `chi_sqaure` test, with a `p_value` of 0.01, to determine which edges to exclude from the input model (See Figure 2). The thickness of the edges in the network indicate the edge weights, the thicker the edge, the higher the edge weight.
- A new network was then constructed using the `bnlearn.structure_learning.fit()` function with the `methodtype='tan'` i.e. the Tree-augmented Naive Bayes method and the `scoretype=k2` i.e. the K2 scoring technique. It was a less complex and thus more indicative network (See Figure 3).
- **Feature Selection** - Using `sklearn.feature_selection.SelectFromModel` and `LinearSVC` modules, feature selection was performed on the original data, leading to the dimensionality of the features being reduced to 7. The parameters for the `LinearSVC` mdoel were:

$$\begin{aligned} C &= 0.05 \\ penalty &= l2 \\ dual &= False \end{aligned}$$

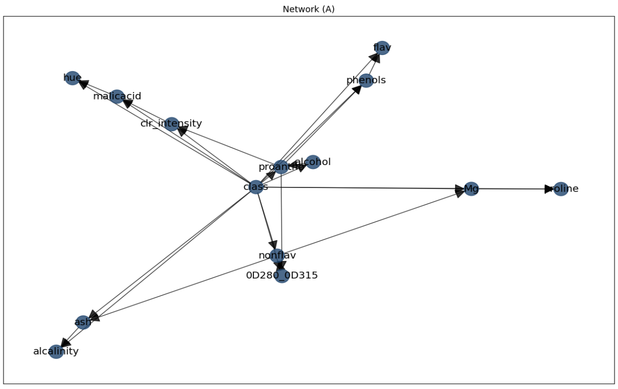

Figure 3: Improved Network (A)

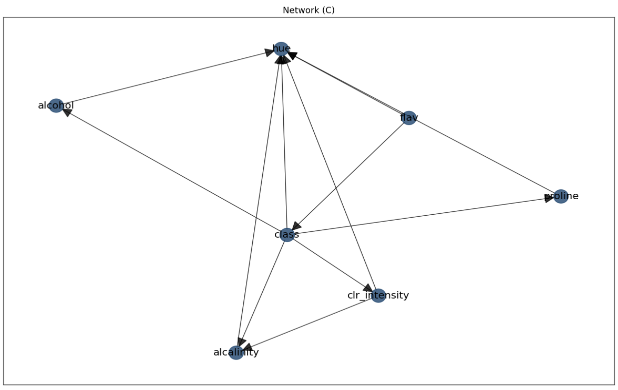

Figure 4: Network (C)

Subsequently, a BN was constructed using `HillClimbingSearch` and K2 scoring on the reduced dataset (See Figure 4). Using the `bnlearn.parameter_learning.fit()` function, the corresponding probability distribution was learned with the `methodtype=bayes` and the `scoretype=k2`. The only parent of *class* was *Flav* (as is visible from the network **(C)**) and is thus the only other attribute significant for the class prediction. The probabilistic distribution $P(class|F1)$ can be visualized from Figure 5.

- Four random test samples were generated to test the accuracy of the 4 constructed networks on each of these samples. The results for the predictions of the classes were not always consistent for the 4 networks for each of these samples. The Improved Network **(A)** and the Network **(C)** predicted a different class for the 1$^{\text{st}}$ sample than the other networks. Network **(C)** consistently predicted a different class for each of the samples thus indicating that it does not perform well when compared to the other networks. This is expected since Network **(C)** is looking at a much smaller domain of attributes to make predictions for the class for the sample as compared to the other networks which can potentially model much more complex relations between the attributes and the target variable.
- The certainty of predictions in these networks, however, varied between the 4 of them to say that Network **(C)** performed significantly worse than the other networks since it is not able to predict with good enough certainty which class the test samples belong to.
- The above observation can be verified using the below performance metric for the classification problem:

Model	Accuracy
Network **(A)**	94.44%
Network **(B)**	94.44%
Improved Network **(A)**	91.66%
Network **(C)**	36.11%

Table 1: Performance Measure

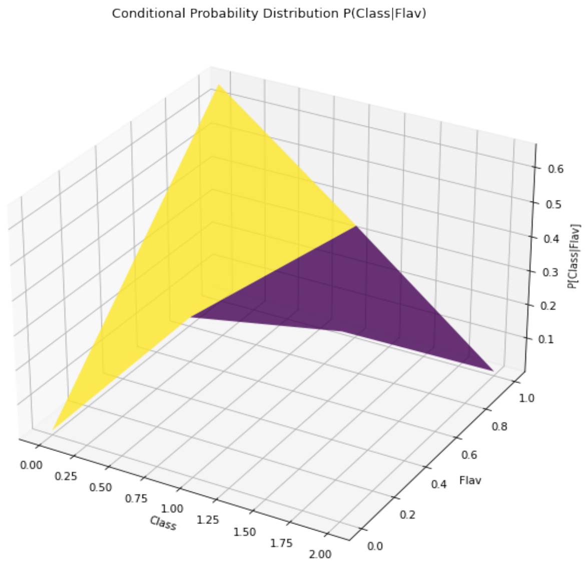

Figure 5: $P(Class|F1)$