

# VE593: Problem Solving with AI Techniques

## Project II

Simona Emilova Doneva (J11803099007)

November 17, 2018

### 1 Introduction

The general objectives of the project are to get familiar with the workings of a Bayesian Network and apply it to solve real-world problems. The tasks are split into Structure Learning, Parameter Learning and Inference. The corresponding algorithms are first shortly introduced. After that, their implemented solution is tested on two data sets.

### 2 Structure and Parameter Learning

The goal of this part is to learn the connections between given variables based on a data set of observations where those variables co-occur.

#### 2.1 Implementation

The implemented algorithm for this task is the K2 Algorithm. It further takes as input a scoring function, which guides the selection of parents for a given node. The two scoring functions that were used are the K2 original score, and the Bayesian Information Criterion (BIC) score. Furthermore, the K2 score was adapted to its log-version as follows <sup>1</sup>:

$$K2Score(i, \pi_i) = \sum_{i=1}^n \sum_{j=1}^{q_i} (\log((r_i - 1)!) - \log((N_{ij} + r_i - 1)!) + \sum_{k=1}^{r_i} \log(N_{ijk})!)$$

The outputs of this stage is an adjacency graph that exposes how the variables are connected (rows are interpreted as the parent). Furthermore a total score is computed that represents the sum over achieved by the given structure.

#### 2.2 Experiments and Results Evaluation

There are two hyper-parameters that could be experimented with. Firstly, the number of parents allowed per node. Secondly, the number of samples used for training the network.

The impact of training samples on the time was inspected by keeping a fixed score algorithm K2Score with a K equal to 2, and increasing the number of considered observations. Figure 1 shows the results plotted for the wine data set.

---

<sup>1</sup>Please refer to [http://www.lx.it.pt/~asmc/pub/talks/09-TA/ta\\_pres.pdf](http://www.lx.it.pt/~asmc/pub/talks/09-TA/ta_pres.pdf) for details.

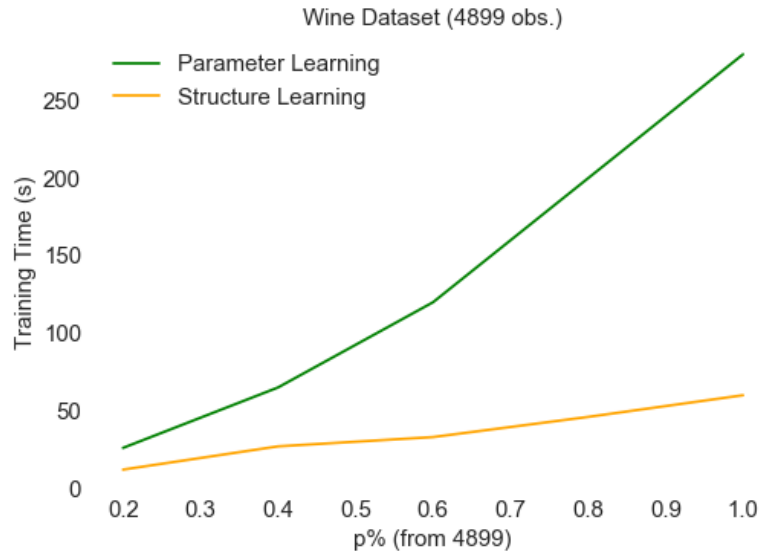


Figure 1: Average time performance on the wine data set with different train set size.

One can observe the general trend that Parameter Learning takes more time, which also increases faster compared to Structure Learning. As the time for all cases was below 5 minutes and it is generally agreed that a 70/30 train-test split is a good choice, further experiments were done in such settings. Similar behaviour was observed for the protein data set and therefore the same proportion was kept there.

As next the impact of K was examined. Final results on both data sets can be seen in Figure 2, and ... for protein.

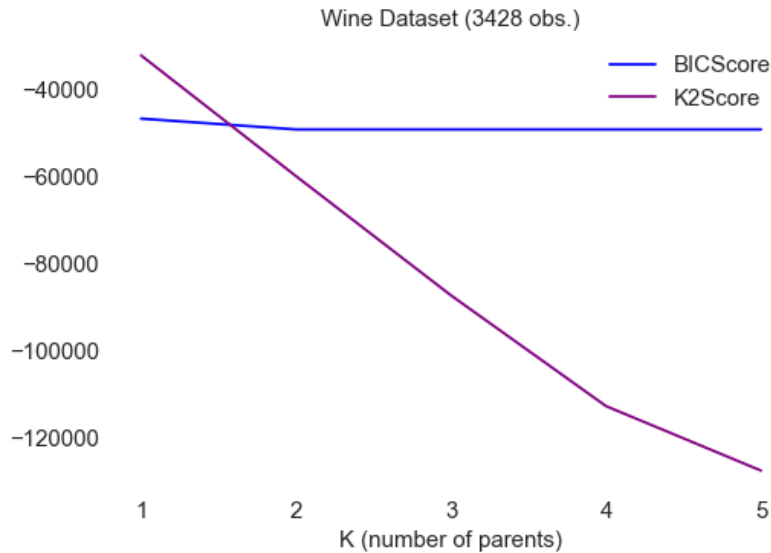


Figure 2: Total Score with BIC and K2 on fixed wine train data set.

It can be observed that the values that the BIC score stabilized around the value -49046, while K2Score kept decreasing with respect to an increasing K. It was further kept track of the average number of parents per node. For BIC the number to which the algorithm converged was 0.5, while for K2Score this was 2.5. Therefore as preferred K for this data set, 2 was selected.

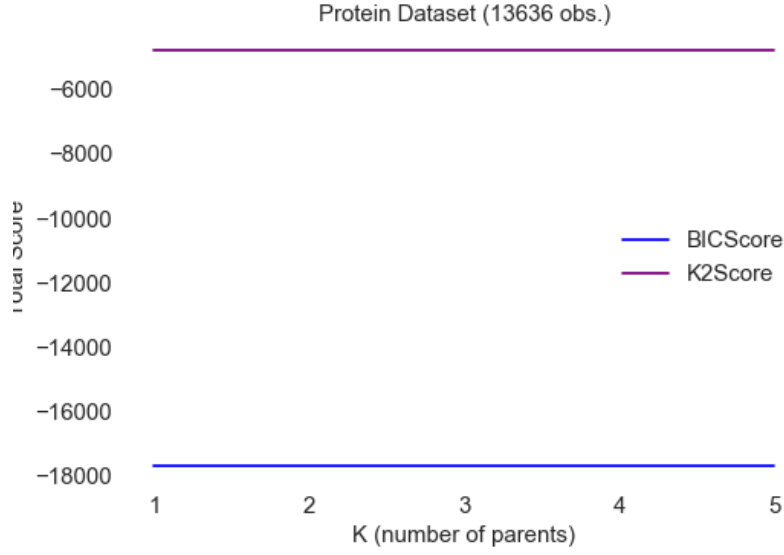


Figure 3: Total Score with BIC and K2 on fixed protein train data set.

The behaviour of the algorithm for this data set is very different that from the previous one. One can see that here independently of the number of allowed parents the scores flattened around a single value. Furthermore in both cases, the average number of parents was below one. Note however that for this data set, the number of variables (6) is less than for the previous one (12).

### 2.3 Final Graphs and Total Score Results

The final adjacency graph for both data sets are presented below, along with the final score achieved. In all cases K was set to two, and a sample of 0.7 from the datasets was used.

$$M_{k2} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$M_{bic} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Table 1: AM for wine data set using K2Score

Table 2: AM for wine data set using BIC Score

One can see that for the wine data set there is a general overlap between the derived relationships. Note however that using BIC leads to less parents, which is also expected as this metric tends to prefer less complex structures.

$$M_{k2} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad M_{bic} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Table 3: AM for protein data set using K2Score

Table 4: AM for protein data set using BIC Score

For the protein data set both algorithms selected only one relationship. Furthermore in Table 5 an overview of the scores achieved by the two score functions for both data sets is provided.

	Wine data set	Protein data set
K2 Score	- 59,792	- 4,779
BIC Score	- 49,046	- 17,681

Table 5: Total Score Results Overview.

### 3 Inference

#### 3.1 Implementation

Inference for this project was performed with the Variable Elimination (VE) method. Figure 4 provides an overview over the implemented logic.

```

Data: Query Variable Index, Observed Variables, Model
Result: Query Probability
conditionedFactors = [CPT list for each variable, conditioned
on the observed indices];
foreach var in model.adjMatrix where var != query and var !=
observed do
    relevantFactors < - getAllFactorsContaining(var);
    conditionedFactors.remove(relevantFactors);
    joinedFactors = pointwiseProduct(relevantFactors);
    reducedFactors = sumout(joinedFactors, var);
end
return normalize(pointwiseProduct(conditionedFactors))

```

Figure 4: Pseudocode Inference with VE.

#### 3.2 Experiments and Results Evaluation

The quality of the trained Bayesian Network can be evaluated in two directions. Firstly, how well it can be used to predict a target variable given a new observation. Secondly, to find a subset of variables that contribute to the prediction quality the most.

Table 6 gives an overview for the performed experiments. One can see that the obtained accuracy

on the protein data set is slightly above 0.5, which is close to random guessing. This could be explained with the previous findings that there is very little connection between the given variables. Therefore making predictions for this task is similar to always choosing the most probable class.

	Wine data set	Protein data set
Accuracy Score	n.a.	0.532
Prominent Variables	n.a.	n.a.

Table 6: Accuracy Scores over the test data sets.

From this end, it was not possible to obtain a set of useful predictors for the target class. Unfortunately, the testing of the algorithm on the wine data set was not successful due to the higher complexity of the possible combinations, and the time constraint given for the project. This is left for future research.