Illustration of the K2 Algorithm for Learning Bayes Net Structures

Prof. Carolina Ruiz
Department of Computer Science, WPI
ruiz@cs.wpi.edu
http://www.cs.wpi.edu/∼ruiz

The purpose of this handout is to illustrate the use of the K2 algorithm to learn the topology of a Bayes Net. The algorithm is taken from [aEH93].

Consider the dataset given in [aEH93]:

case	x_1	x_2	x_3
1	1	0	0
2	1	1	1
3	0	0	1
4	1	1	1
5	0	0	0
6	0	1	1
7	1	1	1
8	0	0	0
9	1	1	1
10	0	0	0

Assume that x_1 is the classification target

The K2 algorithm taken from [aEH93] is included below. This algorithm heuristically searches for the most probable belief–network structure given a database of cases.

```
procedure K2;
{Input: A set of n nodes, an ordering on the nodes, an upper bound u on the
        number of parents a node may have, and a database D containing m cases.}
{Output: For each node, a printout of the parents of the node.}
for i:= 1 to n do
    π_i := ∅;
    P_old := f(i, π_i); {This function is computed using Equation 20.}
    OKToProceed := true;
    While OKToProceed and |π_i| < u do
        let z be the node in Pred(x_i) - π_i that maximizes f(i, π_i ∪ {z});
        P_new := f(i, π_i ∪ {z});
        if P_new > P_old then
            P_old := P_new;
            π_i := π_i ∪ {z};
        else OKToProceed := false;
    end {while};
    write('Node: ', x_i, ' Parent of x_i: ',π_i);
end {for};
end {K2};
```

Equation 20 is included below:

$$f(i, \pi_i) = \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} \prod_{k=1}^{r_i} \alpha_{ijk}!$$

where:

π_i : set of parents of node x_i

q_i = $|\phi_i|$

ϕ_i : list of all possible instantiations of the parents of x_i in database D. That is, if $p_1, \ldots, p_s$ are the parents of x_i then ϕ_i is the Cartesian product $\{v_1^{p_1}, \ldots, v_{r_{p_1}}^{p_1}\}$x...x$\{v_1^{p_s}, \ldots, v_{r_{p_s}}^{p_s}\}$ of all the possible values of attributes p_1 through p_s.

r_i = $|V_i|$

V_i : list of all possible values of the attribute x_i

α_{ijk} : number of cases (i.e. instances) in D in which the attribute x_i is instantiated with its k^{th} value, and the parents of x_i in π_i are instantiated with the j^{th} instantiation in ϕ_i.

N_{ij} = $\sum_{k=1}^{r_i} \alpha_{ijk}$. That is, the number of instances in the database in which the parents of x_i in π_i are instantiated with the j^{th} instantiation in ϕ_i.

The informal intuition here is that $f(i, \pi_i)$ is the probability of the database D given that the parents of x_i are π_i.

Below, we follow the K2 algorithm over the database above.

Inputs:

- The set of $n = 3$ nodes $\{x_1, x_2, x_3\}$,
- the ordering on the nodes x_1, x_2, x_3. We assume that x_1 is the classification target. As such the Weka system would place it first on the node ordering so that it can be the parent of each of the predicting attributes.
- the upper bound $u = 2$ on the number of parents a node may have, and
- the database D above containing $m = 10$ cases.

K2 Algorithm.

$\underline{i = 1}$**:** Note that for $i = 1$, the attribute under consideration is x_1. Here, $r_1 = 2$ since x_1 has two possible values {0,1}.

1. $\pi_1 := \emptyset$
2. $P_{old} := f(1, \emptyset) = \prod_{j=1}^{q_1} \frac{(r_1-1)!}{(N_{1j}+r_1-1)!} \prod_{k=1}^{r_1} \alpha_{1jk}!$

   Let's compute the necessary values for this formula.

   - Since $\pi_1 = \emptyset$ then $q_1 = 0$. Note that the product ranges from $j = 1$ to $j = 0$. The standard convention for a product with an upper limit smaller than the lower limit is that the value of the product is equal to 1. However, this convention wouldn't work here since regardless of the value of i, $f(i, \emptyset) = 1$. Since this value denotes a probability, this would imply that the best option for each node is to have no parents.

     Hence, j will be ignored in the formula above, but not i or k. The example on page 13 of [aEH93] shows that this is the authors' intended interpretation of this formula.
   - $\alpha_{1\_1} = 5$: # of cases with $x_1 = 0$ (cases 3,5,6,8,10)
   - $\alpha_{1\_2} = 5$: # of cases with $x_1 = 1$ (cases 1,2,4,7,9)
   - $N_{1\_} = \alpha_{1\_1} + \alpha_{1\_2} = 10$

   Hence,

   $P_{old} := f(1, \emptyset) = \frac{(r_1-1)!}{(N_{1\_}+r_1-1)!} \prod_{k=1}^{r_1} \alpha_{1\_k}! = \frac{(2-1)!}{(N_{1\_}+2-1)!} \prod_{k=1}^{2} \alpha_{1\_k}! = \frac{1}{11!} * 5! * 5! = 1/2772$
3. Since $Pred(x_1) = \emptyset$, then the iteration for $i = 1$ ends here with $\pi_1 = \emptyset$.

$\underline{i = 2}$**:** Note that for $i = 2$, the attribute under consideration is x_2. Here, $r_2 = 2$ since x_2 has two possible values {0,1}.

1. $\pi_2 := \emptyset$
2. $P_{old} := f(2, \emptyset) = \prod_{j=1}^{q_2} \frac{(r_2-1)!}{(N_{2j}+r_2-1)!} \prod_{k=1}^{r_2} \alpha_{2jk}!$

   Let's compute the necessary values for this formula.

   - Since $\pi_2 = \emptyset$ then $q_2 = 0$. Note that the product ranges from $j = 1$ to $j = 0$. The standard convention for a product with an upper limit smaller than the lower limit is that the value of the product is equal to 1. However, this convention wouldn't work here since regardless of the value of i, $f(i, \emptyset) = 1$. Since this value denotes a probability, this would imply that the best option for each node is to have no parents.

     Hence, j will be ignored in the formula above, but not i or k. The example on page 13 of [aEH93] shows that this is the authors' intended interpretation of this formula.
   - $\alpha_{2\_1} = 5$: # of cases with $x_2 = 0$ (cases 1,3,5,8,10)
   - $\alpha_{2\_2} = 5$: # of cases with $x_2 = 1$ (cases 2,4,6,7,9)
   - $N_{2\_} = \alpha_{2\_1} + \alpha_{2\_2} = 10$

   Hence,

   $P_{old} := f(2, \emptyset) = \frac{(r_2-1)!}{(N_{2\_}+r_2-1)!} \prod_{k=1}^{r_2} \alpha_{2\_k}! = \frac{(2-1)!}{(N_{2\_}+2-1)!} \prod_{k=1}^{2} \alpha_{2\_k}! = \frac{1}{11!} * 5! * 5! = 1/2772$
3. Since $Pred(x_1) = \{x_1\}$, then the only iteration for $i = 2$ goes with $z = x_1$.

   $P_{new} := f(2, \pi_2 \cup \{x_1\}) = f(2, \{x_1\}) = \prod_{j=1}^{q_2} \frac{(r_2-1)!}{(N_{2j}+r_2-1)!} \prod_{k=1}^{r_2} \alpha_{2jk}!$

   - ϕ_2: list of unique π-instantiations of $\{x_1\}$ in $D = ((x_1 = 0), (x_1 = 1))$
   - $q_2 = |\phi_2| = 2$
   - $\alpha_{211} = 4$: # of cases with $x_1 = 0$ and $x_2 = 0$ (cases 3,5,8,10)
   - $\alpha_{212} = 1$: # of cases with $x_1 = 0$ and $x_2 = 1$ (case 6)
   - $\alpha_{221} = 1$: # of cases with $x_1 = 1$ and $x_2 = 0$ (case 1)
   - $\alpha_{222} = 4$: # of cases with $x_1 = 1$ and $x_2 = 1$ (case 2,4,7,9)
   - $N_{21} = \alpha_{211} + \alpha_{212} = 5$
   - $N_{22} = \alpha_{221} + \alpha_{222} = 5$

   $P_{new} = \prod_{j=1}^{q_2} \frac{(r_2-1)!}{(N_{2j}+r_2-1)!} \prod_{k=1}^{2} \alpha_{2jk}! = \frac{(2-1)!}{(5+2-1)!} * \prod_{k=1}^{2} \alpha_{21k}! * \frac{(2-1)!}{(5+2-1)!} * \prod_{k=1}^{2} \alpha_{22k}!$

   $= \frac{1}{6!} * \alpha_{211}! * \alpha_{212}! * \frac{1}{6!} * \alpha_{221}! * \alpha_{222}! = \frac{1}{6!} * 4! * 1! * \frac{1}{6!} * 1! * 4! = \frac{1}{6*5} * \frac{1}{6*5} = 1/900$
4. Since $P_{new} = 1/900 > P_{old} = 1/2772$ then the iteration for $i = 2$ ends with $\pi_2 = \{x_1\}$.

$\underline{i = 3}$**:** Note that for $i = 3$, the attribute under consideration is x_3. Here, $r_3 = 2$ since x_3 has two possible values {0,1}.

1. $\pi_3 := \emptyset$
2. $P_{old} := f(3, \emptyset) = \prod_{j=1}^{q_3} \frac{(r_3-1)!}{(N_{3j}+r_3-1)!} \prod_{k=1}^{r_3} \alpha_{3jk}!$

   Let's compute the necessary values for this formula.

   - Since $\pi_3 = \emptyset$ then $q_3 = 0$. Note that the product ranges from $j = 1$ to $j = 0$. The standard convention for a product with an upper limit smaller than the lower limit is that the value of the product is equal to 1. However, this convention wouldn't work here since regardless of the value of i, $f(i, \emptyset) = 1$. Since this value denotes a probability, this would imply that the best option for each node is to have no parents.

     Hence, j will be ignored in the formula above, but not i or k. The example on page 13 of [aEH93] shows that this is the authors' intended interpretation of this formula.
   - $\alpha_{3\_1} = 4$: # of cases with $x_3 = 0$ (cases 1,5,8,10)
   - $\alpha_{3\_2} = 6$: # of cases with $x_3 = 1$ (cases 2,3,4,6,7,9)
   - $N_{3\_} = \alpha_{3\_1} + \alpha_{3\_2} = 10$

   Hence,

   $P_{old} := f(3, \emptyset) = \frac{(r_3-1)!}{(N_{3\_}+r_3-1)!} \prod_{k=1}^{r_3} \alpha_{3\_k}! = \frac{(2-1)!}{(N_{3\_}+2-1)!} \prod_{k=1}^{2} \alpha_{3\_k}! = \frac{1}{11!} * 4! * 6! = 1/2310$
3. Note that $Pred(x_3) = \{x_1, x_2\}$. Initially, $\pi_3 = \emptyset$. We need to compute $argmax(f(3, \pi_3 \cup \{x_1\}), f(3, \pi_3 \cup \{x_2\}))$.

   - $f(3, \pi_3 \cup \{x_1\}) = f(3, \{x_1\}) = \prod_{j=1}^{q_3} \frac{(r_3-1)!}{(N_{3j}+r_3-1)!} \prod_{k=1}^{r_3} \alpha_{3jk}!$
     - ϕ_3: list of unique π-instantiations of $\{x_1\}$ in $D = ((x_1 = 0), (x_1 = 1))$
     - $q_3 = |\phi_3| = 2$
     - $\alpha_{311} = 3$: # of cases with $x_1 = 0$ and $x_3 = 0$ (cases 5,8,10)
     - $\alpha_{312} = 2$: # of cases with $x_1 = 0$ and $x_3 = 1$ (case 3,6)
     - $\alpha_{321} = 1$: # of cases with $x_1 = 1$ and $x_3 = 0$ (case 1)
     - $\alpha_{322} = 4$: # of cases with $x_1 = 1$ and $x_3 = 1$ (case 2,4,7,9)
     - $N_{31} = \alpha_{311} + \alpha_{312} = 5$
     - $N_{32} = \alpha_{321} + \alpha_{322} = 5$

     $f(3, \{x_1\}) = \prod_{j=1}^{q_3} \frac{(r_3-1)!}{(N_{3j}+r_3-1)!} \prod_{k=1}^{2} \alpha_{3jk}! = \frac{(2-1)!}{(5+2-1)!} * \prod_{k=1}^{2} \alpha_{31k}! * \frac{(2-1)!}{(5+2-1)!} * \prod_{k=1}^{2} \alpha_{32k}!$
     $= \frac{1}{6!} * \alpha_{311}! * \alpha_{312}! * \frac{1}{6!} * \alpha_{321}! * \alpha_{322}! = \frac{1}{6!} * 3! * 2! * \frac{1}{6!} * 1! * 4! = \frac{1}{6*5*2} * \frac{1}{6*5} = 1/1800$
   - $f(3, \pi_3 \cup \{x_2\}) =) = \prod_{j=1}^{q_3} \frac{(r_3-1)!}{(N_{3j}+r_3-1)!} \prod_{k=1}^{r_3} \alpha_{3jk}!$
     - ϕ_3: list of unique π-instantiations of $\{x_2\}$ in $D = ((x_2 = 0), (x_2 = 1))$
     - $q_3 = |\phi_3| = 2$
     - $\alpha_{311} = 4$: # of cases with $x_2 = 0$ and $x_3 = 0$ (cases 1,5,8,10)
     - $\alpha_{312} = 1$: # of cases with $x_2 = 0$ and $x_3 = 1$ (case 3)

- $\alpha_{321} = 0$: # of cases with $x_2 = 1$ and $x_3 = 0$ (no case)
- $\alpha_{322} = 5$: # of cases with $x_2 = 1$ and $x_3 = 1$ (case 2,4,6,7,9)
- $N_{31} = \alpha_{311} + \alpha_{312} = 5$
- $N_{32} = \alpha_{321} + \alpha_{322} = 5$

$f(3, \{x_2\}) = \prod_{j=1}^{q_3} \frac{(r_3-1)!}{(N_{3j}+r_3-1)!} \prod_{k=1}^{2} \alpha_{3jk}! = \frac{(2-1)!}{(5+2-1)!} * \prod_{k=1}^{2} \alpha_{31k}! * \frac{(2-1)!}{(5+2-1)!} * \prod_{k=1}^{2} \alpha_{32k}!$
$= \frac{1}{6!} * \alpha_{311}! * \alpha_{312}! * \frac{1}{6!} * \alpha_{321}! * \alpha_{322}! = \frac{1}{6!} * 4! * 1! * \frac{1}{6!} * 0! * 5! = \frac{1}{6*5} * \frac{1}{6} = 1/180$
Here we assume that $0! = 1$

4. Since $f(3, \{x_2\}) = 1/180 > f(3, \{x_1\}) = 1/1800$ then $z = x_2$. Also, since $f(3, \{x_2\}) = 1/180 > P_{old} = f(3, \emptyset) = 1/2310$, then $\pi_3 = \{x_2\}$, $P_{old} := P_{new} = 1/180$.

5. Now, the next iteration of the algorithm for $i = 3$, considers adding the remaining predecessor of x_3, namely x_1, to the parents of x_3.
   $f(3, \pi_3 \cup \{x_1\}) = f(3, \{x_1, x_2\}) = \prod_{j=1}^{q_3} \frac{(r_3-1)!}{(N_{3j}+r_3-1)!} \prod_{k=1}^{r_3} \alpha_{3jk}!$

   - ϕ_3: list of unique π-instantiations of $\{x_1, x_2\}$ in $D = ((x_1 = 0, x_2 = 0), (x_1 = 0, x_2 = 1), (x_1 = 1, x_2 = 0), (x_1 = 1, x_2 = 1))$
   - $q_3 = |\phi_3| = 4$
   - $\alpha_{311} = 3$: # of cases with $x_1 = 0, x_2 = 0$ and $x_3 = 0$ (cases 5,8,10)
   - $\alpha_{312} = 1$: # of cases with $x_1 = 0, x_2 = 0$ and $x_3 = 1$ (case 3)
   - $\alpha_{321} = 0$: # of cases with $x_1 = 0, x_2 = 1$ and $x_3 = 0$ (no case)
   - $\alpha_{322} = 1$: # of cases with $x_1 = 0, x_2 = 1$ and $x_3 = 1$ (case 6)
   - $\alpha_{331} = 1$: # of cases with $x_1 = 1, x_2 = 0$ and $x_3 = 0$ (case 1)
   - $\alpha_{332} = 0$: # of cases with $x_1 = 1, x_2 = 0$ and $x_3 = 1$ (no case)
   - $\alpha_{341} = 0$: # of cases with $x_1 = 1, x_2 = 1$ and $x_3 = 0$ (no case)
   - $\alpha_{342} = 4$: # of cases with $x_1 = 1, x_2 = 1$ and $x_3 = 1$ (case 2,4,7,9)
   - $N_{31} = \alpha_{311} + \alpha_{312} = 4$
   - $N_{32} = \alpha_{321} + \alpha_{322} = 1$
   - $N_{33} = \alpha_{331} + \alpha_{332} = 1$
   - $N_{34} = \alpha_{341} + \alpha_{342} = 4$

   $f(3, \{x_1, x_2\}) = \prod_{j=1}^{q_3} \frac{(r_3-1)!}{(N_{3j}+r_3-1)!} \prod_{k=1}^{2} \alpha_{3jk}!$
   $= \frac{(2-1)!}{(4+2-1)!} * \prod_{k=1}^{2} \alpha_{31k}! * \frac{(2-1)!}{(1+2-1)!} * \prod_{k=1}^{2} \alpha_{32k}! * \frac{(2-1)!}{(1+2-1)!} * \prod_{k=1}^{2} \alpha_{33k}! * \frac{(2-1)!}{(4+2-1)!} * \prod_{k=1}^{2} \alpha_{34k}!$
   $= \frac{1}{5!} * \alpha_{311}! * \alpha_{312}! * \frac{1}{2!} * \alpha_{321}! * \alpha_{322}! * \frac{1}{2!} * \alpha_{331}! * \alpha_{332}! * \frac{1}{5!} * \alpha_{341}! * \alpha_{342}!$
   $= \frac{1}{5!} * 3! * 1! * \frac{1}{2!} * 0! * 1! * \frac{1}{2!} * 1! * 0! * \frac{1}{5!} * 0! * 4! = \frac{1}{5*4} * \frac{1}{2} * \frac{1}{2} * \frac{1}{5} = 1/400$

6. Since $P_{new} = 1/400 < P_{old} = 1/180$ then the iteration for $i = 3$ ends with $\pi_3 = \{x_2\}$.

Outputs: For each node, a printout of the parents of the node.

Node: x_1, Parent of $x_1 : \pi_1 = \emptyset$
Node: x_2, Parent of $x_2 : \pi_2 = \{x_1\}$
Node: x_3, Parent of $x_3 : \pi_3 = \{x_2\}$

This concludes the run of K2 over the database D. The learned topology is

$$x_1 \longrightarrow x_2 \longrightarrow x_3$$

References

[aEH93] Gregory F. Cooper Edward Herskovits. A bayesian method for the induction of probabilistic networks from data. Technical Report KSL-91-02, Knowledge Systems Laboratory. Medical Computer Science. Stanford University School of Medicine, Stanford, CA 94305-5479, Updated Nov. 1993. Available at: http://smi-web.stanford.edu/pubs/SMI_Abstracts/SMI-91-0355.html.