Machine Learning

Bayesian Network

Marchelo Bragagnini
cesarbrma91@gmail.com
@MarchBragagnini

Universidad Católica San Pablo

Centro de Investigación
e Innovación en
Ciencia Computación

Outline

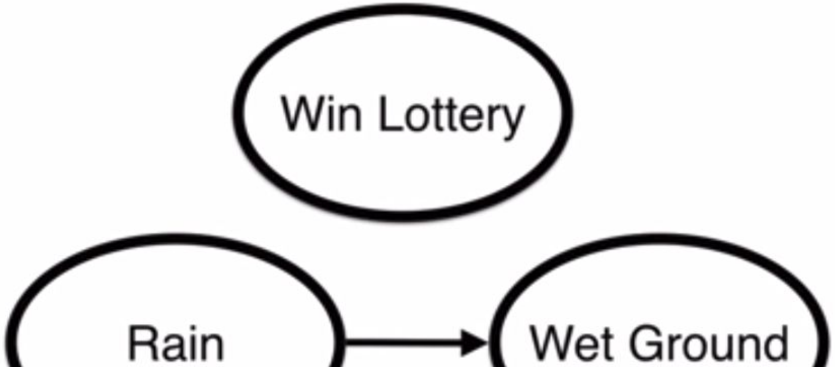
1. Bayes’ Theorem
2. Bayes Network
3. Conditional independencies in BN
4. Hybrid Bayesian Network
5. Exact Inference in Bayesian Networks
6. Approximate Inference in Bayesian Networks
7. Summary Bayesian Networks
8. Exercises
9. Weka
10. References

Bayes' Theorem

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$

- Comments:
  - Bayes' rule tells us how to 'invert' conditional probabilities, i.e. to find P(B|A) from P(A|B).

Bayes Network

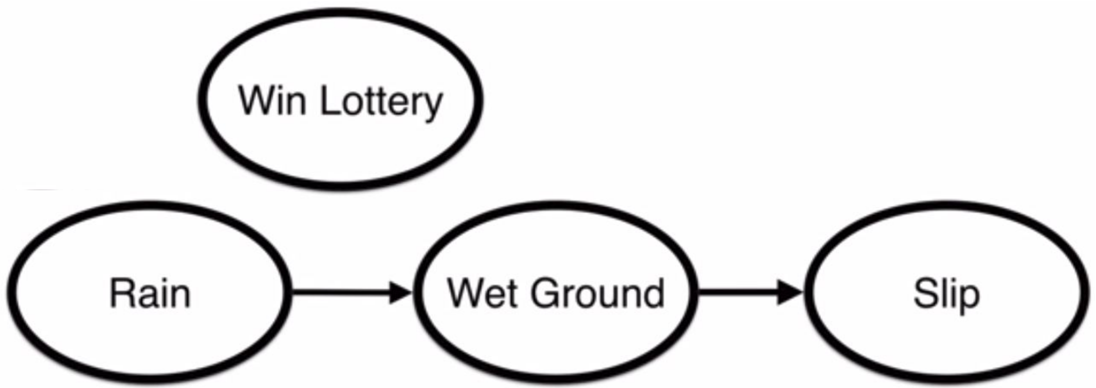


$$P(L, R, W, S) = P(L)P(R)P(W \mid R)P(S$$

$$P(L, R, W) = P(L)P(R)P(W \mid R)$$

[ ] https://www.youtube.com/watch?v=TuGDMj43ehw

Bayes Network



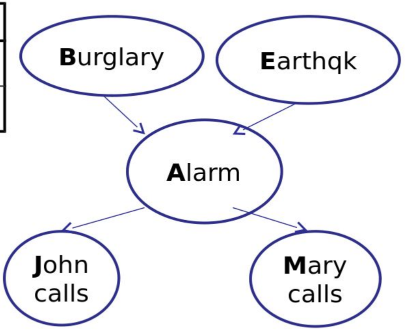
$$P(L, R, W, S) = P(L)P(R)P(W \mid R)P(S \mid W)$$

$$P(L, R, W) = P(L)P(R)P(W \mid R)$$

[ ] https://www.youtube.com/watch?v=TuGDMj43ehw

Bayes Network

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998

B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

A	J	P(J|A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

$$P(+b,-e,+a,-j,+m) =$$

$$P(+b)P(-e)P(+a|+b,-e)P(-j|+a)P(+m|+a)$$

[ ] UC Berkeley CS188 Intro to AI -- Course Materials

Conditional independencies in BN

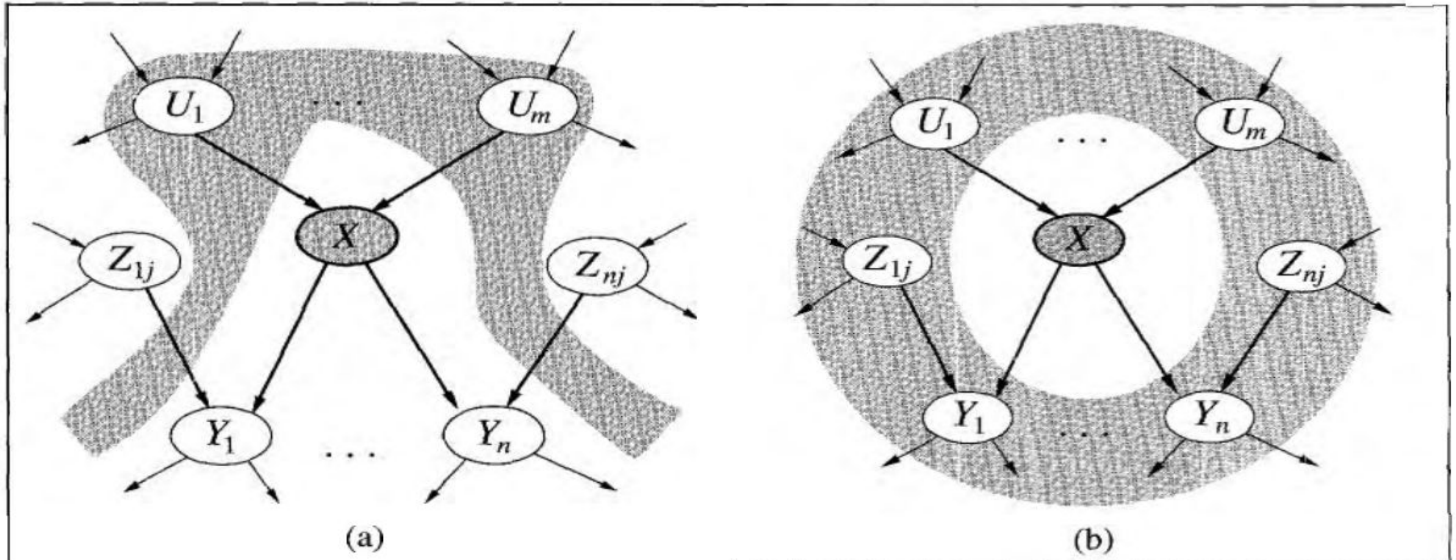
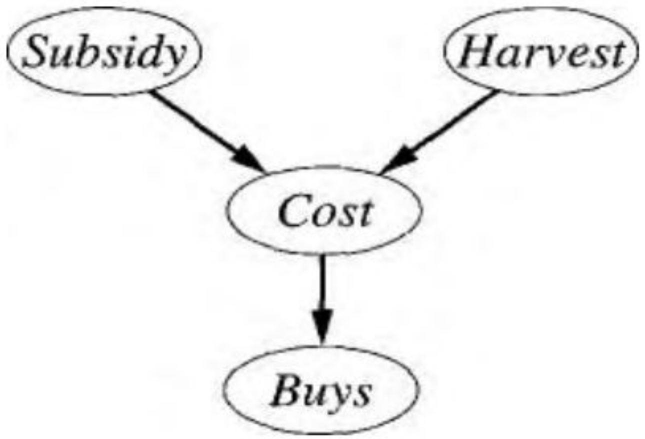


Figure 14.4 (a) A node X is conditionally independent of its non-descendants (e.g., the Z_{ij}s) given its parents (the U_is shown in the gray area). (b) A node X is conditionally independent of all other nodes in the network given its Markov blanket (the gray area).

[ ] Artificial Intelligence A Modern Approach, 2nd edition, Russel & Norvig

Hybrid Bayesian Network

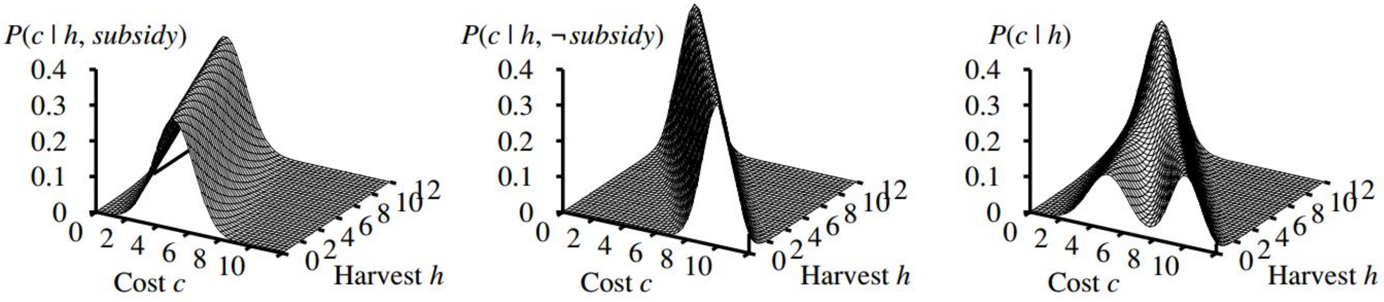
A network with both discrete and continuous variables is called a hybrid Bayesian network. To specify a hybrid network, we have to specify two new kinds of distributions: the conditional distribution for a continuous variable given discrete or continuous parents; and the conditional distribution for a discrete variable given continuous parents.



A simple network with discrete variables(Subsidy and Buys)and continuous variables (Harvest and Cost).

[ ] Artificial Intelligence A Modern Approach, 2nd edition, Russel & Norvig

Hybrid Bayesian Network



(a) (b) (c)

The graphs in (a) and (b) show the probability distribution over Cost as a function of Harvest size, with Subsidy true and false, respectively. Graph (c) shows the distribution P(Cost | Harvest), obtained by summing over the two subsidy cases.

$$P(c \mid h, subsidy) \;=\; N(a_t h + b_t, \sigma_t^2)(c) = \frac{1}{\sigma_t\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{c-(a_t h+b_t)}{\sigma_t}\right)^2}$$

$$P(c \mid h, \neg subsidy) \;=\; N(a_f h + b_f, \sigma_f^2)(c) = \frac{1}{\sigma_f\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{c-(a_f h+b_f)}{\sigma_f}\right)^2}$$

[ ] Artificial Intelligence A Modern Approach, 2nd edition, Russel & Norvig

Exact Inference in Bayesian Networks

Considering the following equation:

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha\, \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

Calculate

$$\mathbf{P}(Burglary \mid JohnCalls = true, MaryCalls = true)$$

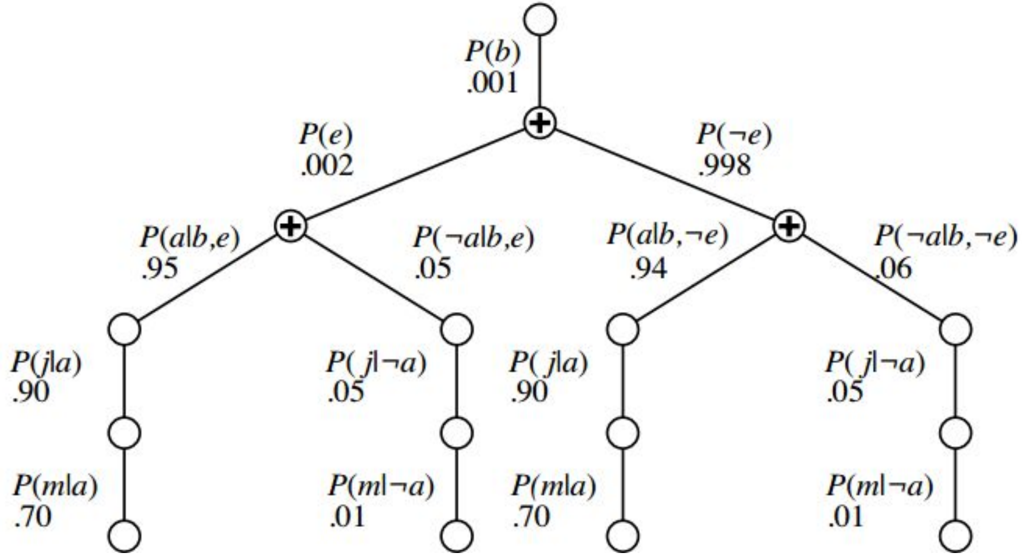
$$\mathbf{P}(B \mid j, m) = \alpha\, \mathbf{P}(B, j, m) = \alpha \sum_{e} \sum_{a} \mathbf{P}(B, j, m, e, a,)$$

[ ] Artificial Intelligence A Modern Approach, 2nd edition, Russel & Norvig

Exact Inference in Bayesian Networks

$$P(b \mid j, m) = \alpha \sum_e \sum_a P(b)P(e)P(a \mid b, e)P(j \mid a)P(m \mid a)$$

$$P(b \mid j, m) = \alpha P(b) \sum_e P(e) \sum_a P(a \mid b, e)P(j \mid a)P(m \mid a)$$

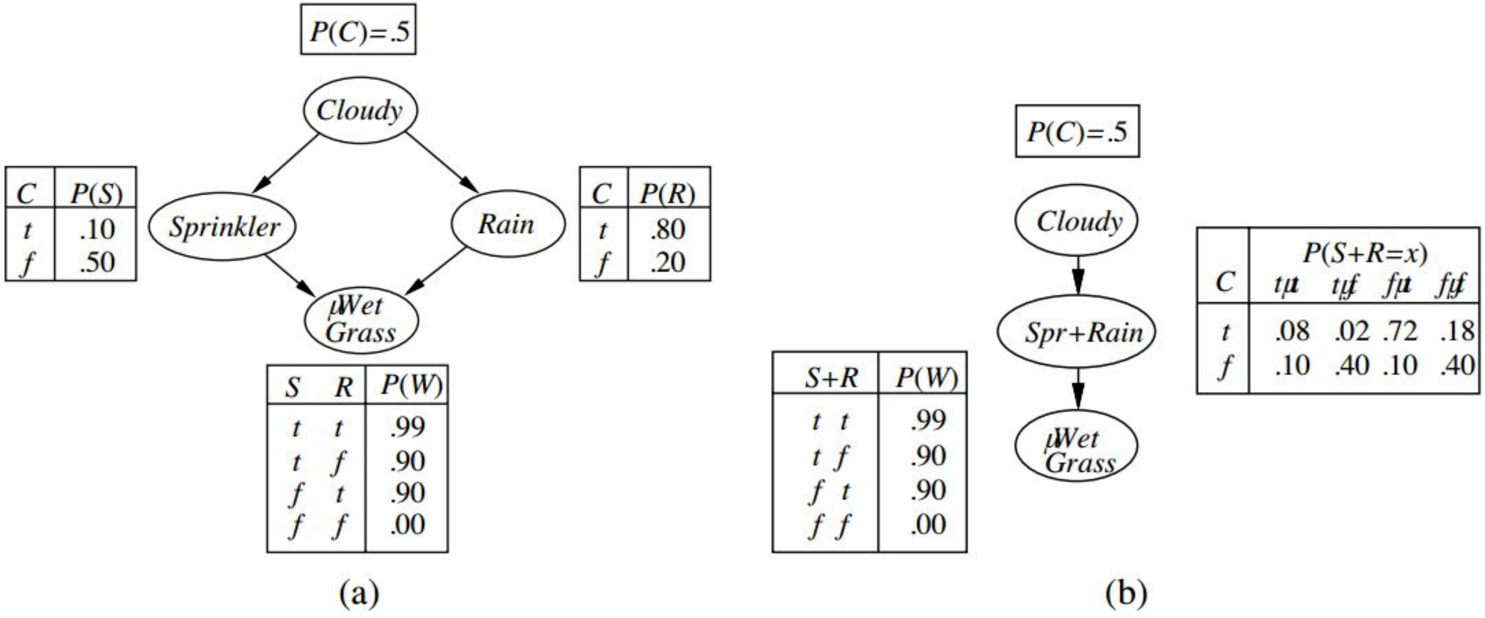


[ ] Artificial Intelligence A Modern Approach, 2nd edition, Russel & Norvig

Exact Inference in Bayesian Networks

Complexity: NP Hard (generally)
One method: For example, in a polytree network, one would need to issue O(n) queries CLUSTERING costing O(n) each, for a total of 0 (n 2) time. Using clustering algorithms (also known as JOIN TREE join tree algorithms), the time can be reduced to O(n).

Clustering:



[ ] Artificial Intelligence A Modern Approach, 2nd edition, Russel & Norvig

Approximate Inference in Bayesian Networks

1. Monte Carlo algorithms.
   - Markov Chain Monte Carlo
     - Gibbs Sampler

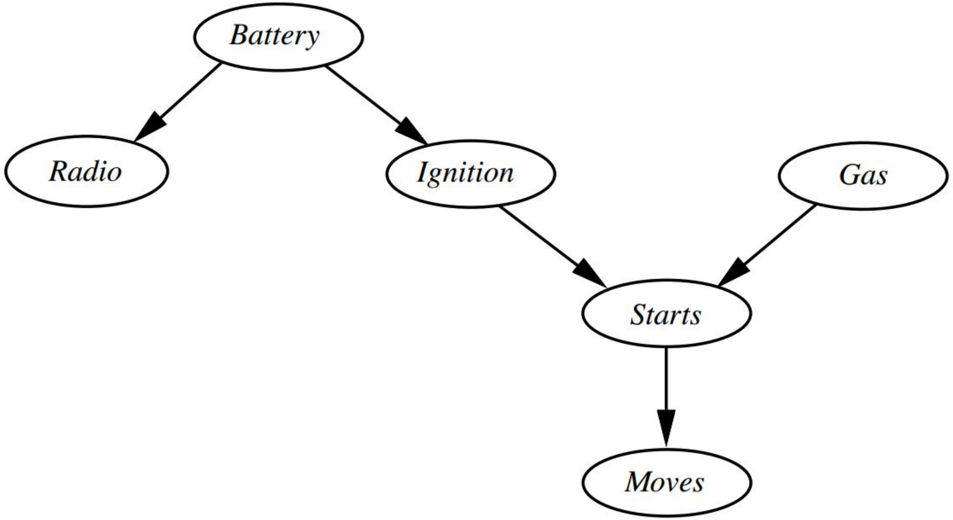
2. Likelihood Weighting

[ ] Artificial Intelligence A Modern Approach, 2nd edition, Russel & Norvig

Exercises

14.8 Consider the network for car diagnosis shown in Figure 14.21.

a. Extend the network with the Boolean variables $IcyWeather$ and $StarterMotor$.



[ ] Artificial Intelligence A Modern Approach, 3ra edition, Russel & Norvig

Exercises

14.12 Three soccer teams A, B, and C, play each other once. Each match is between two teams, and can be won, drawn, or lost. Each team has a fixed, unknown degree of quality—an integer ranging from 0 to 3—and the outcome of a match depends probabilistically on the difference in quality between the two teams.

a. Construct a relational probability model to describe this domain, and suggest numerical values for all the necessary probability distributions.

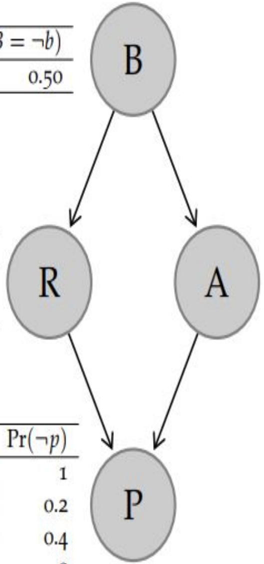
b. Construct the equivalent Bayesian network.

c. Suppose that in the first two matches A beats B and draws with C. Using an exact inference algorithm of your choice, compute the posterior distribution for the outcome of the third match.

d. Suppose there are n teams in the league and we have the results for all but the last match. How does the complexity of predicting the last game vary with n?

[ ] Artificial Intelligence A Modern Approach, 2da edition, Russel & Norvig

Exercises



$\Pr(B=b)$	$\Pr(B=\neg b)$
0.50	0.50

B	$\Pr(r)$	$\Pr(\neg r)$
T	0.3	0.7
F	0.8	0.2

B	$\Pr(a)$	$\Pr(\neg a)$
T	0.1	0.9
F	0.5	0.5

R	A	$\Pr(p)$	$\Pr(\neg p)$
T	T	0	1
T	F	0.8	0.2
F	T	0.6	0.4
F	F	1	0

Exercise 2 – Bayesian Networks – Inference

Figure shows a graphical model with conditional probabilities tables about whether or not you will panic at an exam based on whether or not the course was boring ("B"), which was the key factor you used to decide whether or not to attend lectures ("A") and revise doing the exercises after each lecture ("R").
You should use the model to make exact *inference* and answer the following queries:

- what is the probability that you will panic or not before the exam given that you attended the lectures and revised after each lecture?
- what is the probability that you will panic or not before the exam?
- your teacher saw you panicking at the exam and he wants to work out from the model the reason for that. Was it because you did not come to the lecture or because you did not revise?

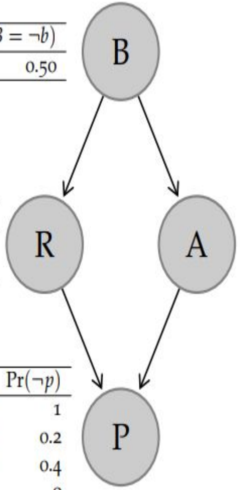
[ ] DM825 - Introduction to Machine Learning, Sheet 12, Spring 2013

Exercises

Solution

what is the probability that you will panic or not before the exam? **Solution**

$\Pr(B=b)$	$\Pr(B=\neg b)$
0.50	0.50



B	$\Pr(r)$	$\Pr(\neg r)$
T	0.3	0.7
F	0.8	0.2

B	$\Pr(a)$	$\Pr(\neg a)$
T	0.1	0.9
F	0.5	0.5

R	A	$\Pr(p)$	$\Pr(\neg p)$
T	T	0	1
T	F	0.8	0.2
F	T	0.6	0.4
F	F	1	0

$$
\begin{aligned}
\Pr(p) &= \sum_{b,r,a} \Pr(b,r,a,p) \\
&= \sum_{b,r,a} \Pr(b)\Pr(r|b)\Pr(a|b)\Pr(p|r,a) \\
&= \sum_{b} \Pr(b) \sum_{r,a} \Pr(r|b)\Pr(a|b)\Pr(p|r,a) \\
&= 0.5 \cdot (0.3 \cdot 0.1 \cdot 0 + 0.3 \cdot 0.9 \cdot 0.8 + 0.7 \cdot 0.1 \cdot 0.6 + 0.7 \cdot 0.9 \cdot 1) \\
&\quad +0.5 \cdot (0.8 \cdot 0.5 \cdot 0 + \cdot 0.8 \cdot 0.5 \cdot 0.8 + 0.2 \cdot 0.5 \cdot 0.6 + 0.2 \cdot 0.5 \cdot 1) \\
&= 0.684
\end{aligned}
$$

[ ] DM825 - Introduction to Machine Learning, Sheet 12, Spring 2013

Summary Bayesian Networks

1. A Bayesian network is a directed acyclic graph whose nodes correspond to random variables; each node has a conditional distribution for the node, given its parents.
2. Bayesian networks provide a concise way to represent conditional independence relationships in the domain.
3. A Bayesian network specifies a full joint distribution; each joint entry is defined as the product of the corresponding entries in the local conditional distributions. A Bayesian network is often exponentially smaller than the full joint distribution.
4. Stochastic approximation techniques such as likelihood weighting and Markov chain Monte Carlo can give reasonable estimates of the true posterior probabilities in a network and can cope with much larger networks than can exact algorithms.

[ ] Artificial Intelligence A Modern Approach, 2nd edition, Russel & Norvig

Weka

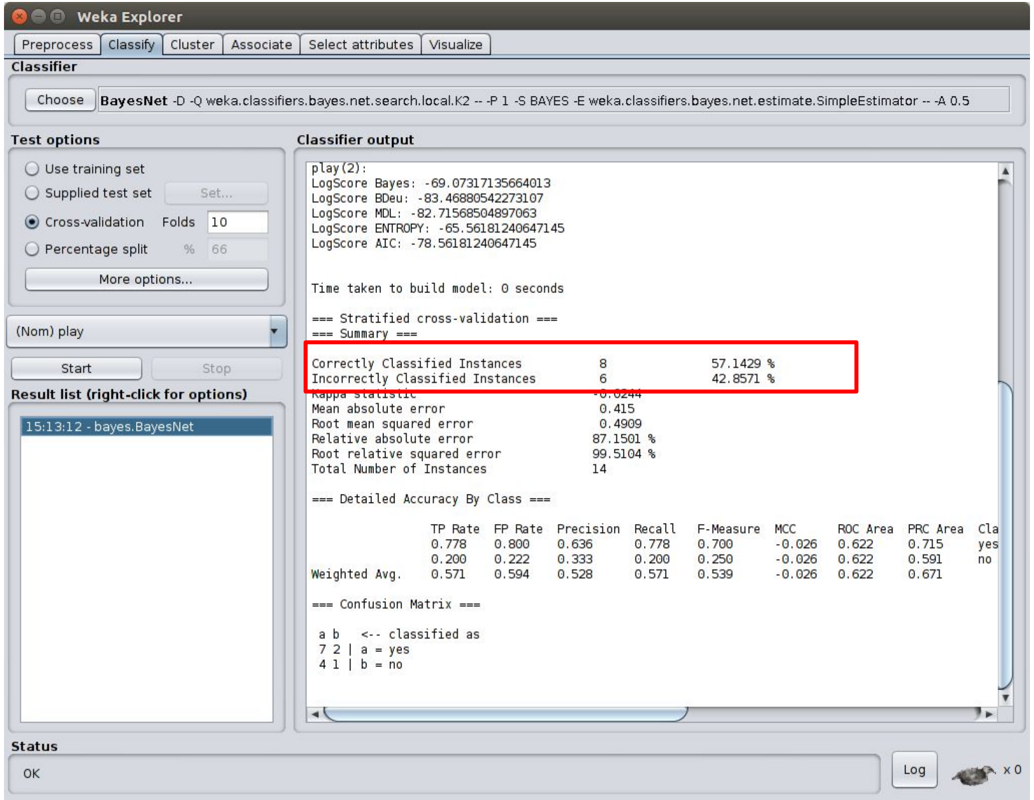
Download & commands

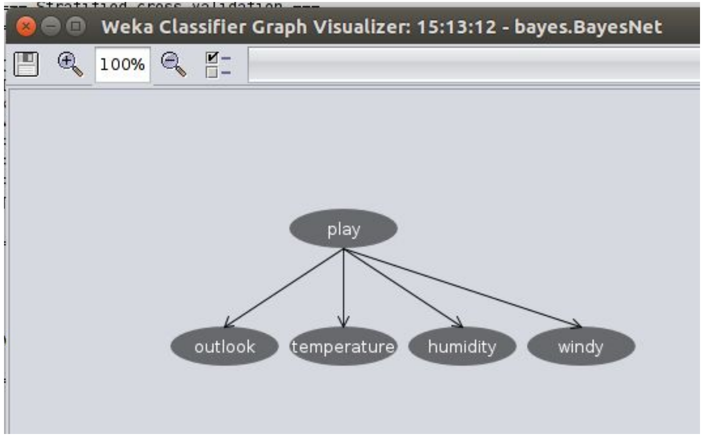
- https://www.cs.waikato.ac.nz/ml/weka/downloading.html

```
$ cd weka*
$ java -jar weka.jar
$ curl https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data --output iris.csv
```

Weka

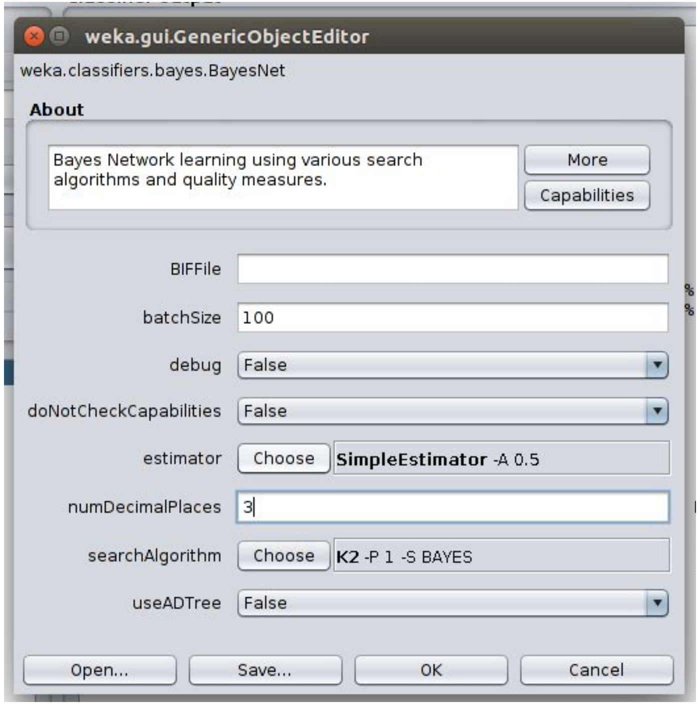
1 Open file
2 Choose Bayes classifier

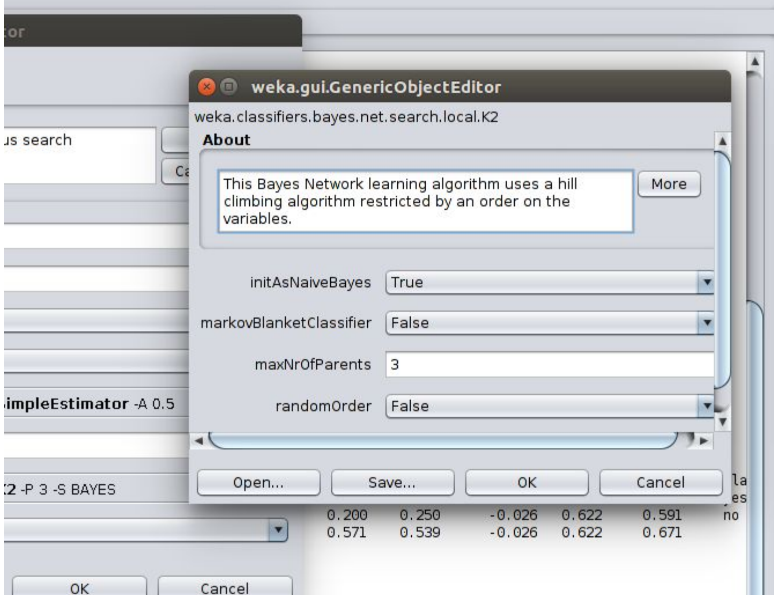




[ ] http://web.ydu.edu.tw/~alan9956/docu/refer/BayesWEKA.pdf

Weka



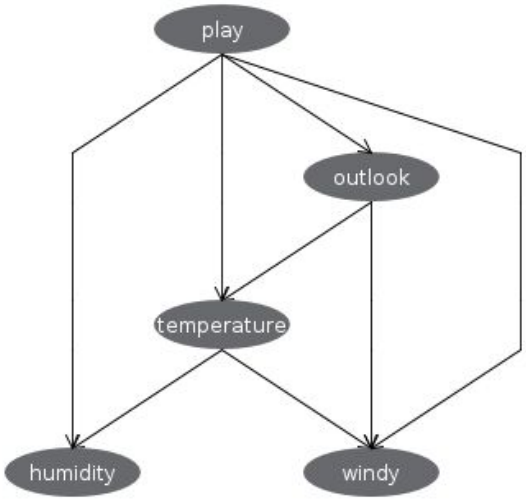


[ ] http://web.ydu.edu.tw/~alan9956/docu/refer/BayesWEKA.pdf

Weka

Classifier output

```
play(2):
LogScore Bayes: -65.48679330235674
LogScore BDeu: -169.62470568231376
LogScore MDL: -133.5327267657704
LogScore ENTROPY: -79.4320515086576
LogScore AIC: -120.43205150865761


Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances           9               64.2857 %
Incorrectly Classified Instances         5               35.7143 %
Kappa statistic                          0.186
Mean absolute error                      0.4359
Root mean squared error                  0.5002
Relative absolute error                 91.5426 %
Root relative squared error            101.3956 %
Total Number of Instances               14

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Cla
                 0.778    0.600    0.700      0.778    0.737      0.189    0.578     0.728     yes
                 0.400    0.222    0.500      0.400    0.444      0.189    0.578     0.632     no
Weighted Avg.    0.643    0.465    0.629      0.643    0.632      0.189    0.578     0.694

=== Confusion Matrix ===

 a b   <-- classified as
 7 2 | a = yes
 3 2 | b = no
```



[ ] http://web.ydu.edu.tw/~alan9956/docu/refer/BayesWEKA.pdf

References

- https://machinelearningmastery.com/load-csv-machine-learning-data-weka/
- https://www.youtube.com/watch?v=tpH905jiBZ0
- http://web.ydu.edu.tw/~alan9956/docu/refer/BayesWEKA.pdf
- https://www.youtube.com/watch?v=TuGDMj43ehw
- Artificial Intelligence: A Modern Approach http://aima.cs.berkeley.edu/
- CS 5804: Introduction to Artificial Intelligence http://courses.cs.vt.edu/cs4804/Fall16/
- UC Berkeley CS188 Intro to AI -- Course Materials http://ai.berkeley.edu/lecture_slides.html
- JavaBayes - https://www.cs.cmu.edu/~javabayes/Home/node3.html
- DM825 - Introduction to Machine Learning Sheet 12, Spring 2013

Machine Learning

Bayesian Network

Marchelo Bragagnini
cesarbrma91@gmail.com
@MarchBragagnini

Universidad Católica San Pablo

Centro de Investigación
e Innovación en
Ciencia Computación