Bayesian Networks Lab

Andrea Passerini and Luca Erculiani

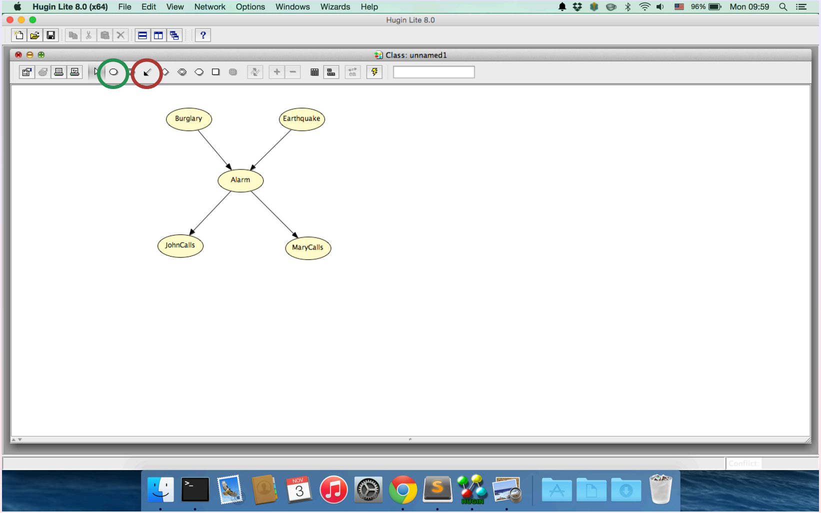
Machine Learning

The software

HuginLite

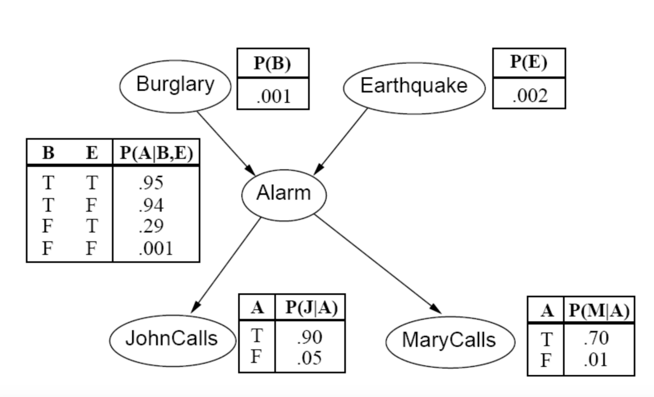
- Trial version of the Hugin family of software for Bayesian Networks
- The free trial version is limited to handle max. 50 states and learn from max. 500 cases
- It is prohibited to use the free Hugin Lite for any other purpose than the demonstration of capabilities and proof of concept
- Freely available at `http://www.hugin.com/index.php/hugin-lite/`

Defining Nodes and Links

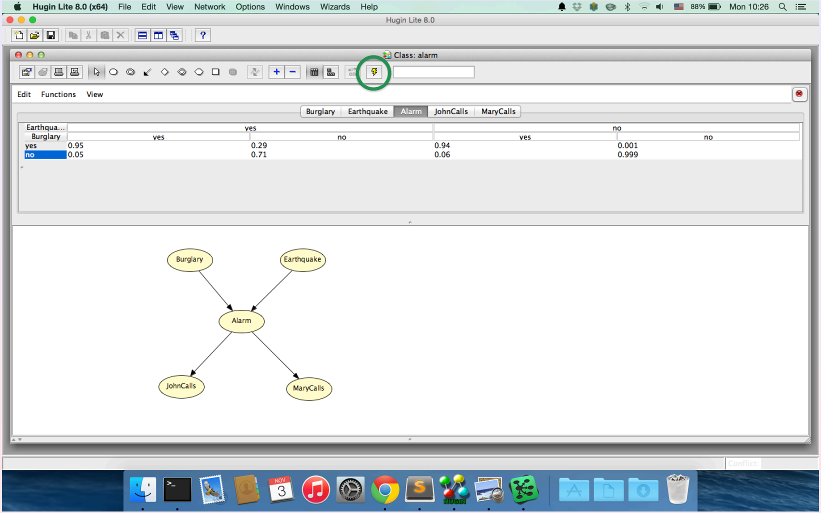


Defining the States

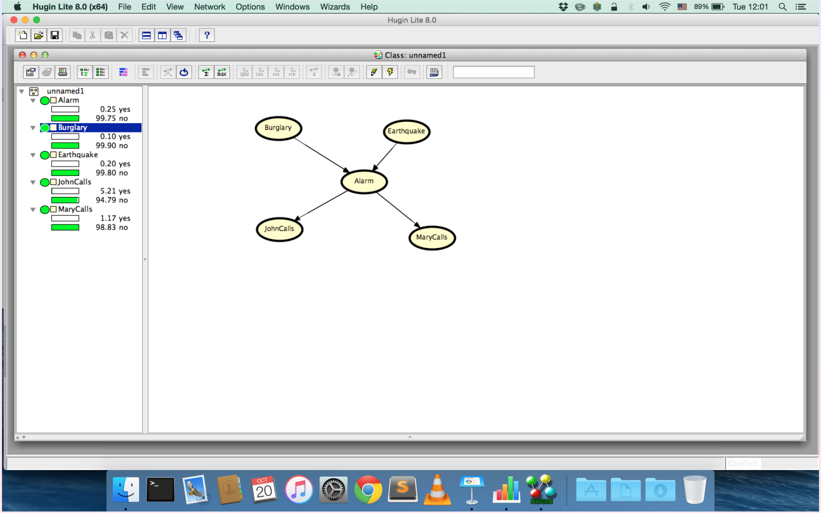
- Open CPT by clicking on a node holding the CRTL key
- Rename states, insert probability for each configuration



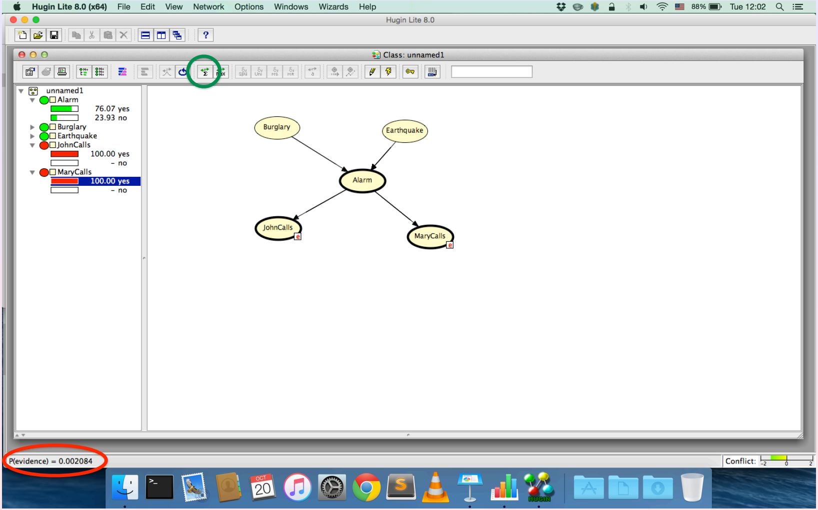
Compiling the Network



Running the Network



P(evidence)



Hugin Lite 8.0 (x64) File Edit View Network Options Windows Wizards Help
Hugin Lite 8.0
Class: unnamed1
unnamed1
Alarm
76.07 yes
23.93 no
Burglary
Earthquake
JohnCalls
100.00 yes
– no
MaryCalls
100.00 yes
– no
Burglary
Earthquake
Alarm
JohnCalls
MaryCalls
P(evidence) = 0.002084
Conflict:

Computing the probability of a combination of states

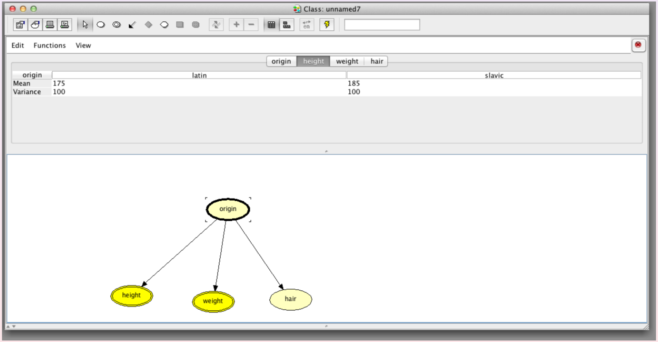
- We want to compute $P(alarm = "yes", johncalls = "yes" | burglary = "yes")$
- Exploting that $P(A, B) = P(A|B)P(B)$

$$P(alarm = "yes", johncalls = "yes" | burglary = "yes") = \\ = \frac{P(alarm = "yes", johncalls = "yes", burglary = "yes")}{P(burglary = "yes")}$$

$$P(alarm = "yes", johncalls = "yes" | burglary = "yes") = \\ = \frac{0.000846}{0.001} = 0.846$$

Hybrid Networks

- Continuous nodes with mean and variance (Gaussian distributions)
- Continous nodes can be children of discrete ones, not viceversa



Learning from Data

Learning Wizard

1. Select Wizards, Learning Wizard
2. Load the training file (`small_asia.dat`)
3. In structure constraints import model information (from `ChestClinic.net`)
4. Select a learning algorithm
5. RUN the learning algorithm
6. Compile the learned network

Warning

- Without priors, some configurations get zero probability
- Add priors (experience) before running the learning (e.g. prior of 1 to each configuration)

Using Learned Network

Analysis Wizard

1. Select Wizards, Analysis Wizard
2. Sample 100 new examples according to the learned network
3. Check them in Data Source
4. Analyze the quality of the generated data in Data Accuracy
5. Clear the Data Source and Load the test file (`test_asia_small.dat`)
6. Analyze the performance of classification of the learned network

Assignment

1. Consider the data file `leukemia.dat`
2. Each example contains 5 genes (active/inactive) and a label (AML/ALL)
3. Randomly split the file in train and test (80% train, 20% test)
4. Learn Bayesian network on train with different learning algorithms:
   - NPC
   - Greedy search-and-score
   - Fixed Naive Bayes structure (NOTE: this is NOT tree-augmented Naive Bayes, see slides on Naive Bayes)
5. Test the learned Bayesian networks on test
6. Write a short report (2-3 pages) summarizing the methodology used and the results obtained.

Assignment

- After completing the assignment submit it via email
- Send an email to mllab@unitn.it
- Subject: HuginSubmit2019
- Attachment: id_name_surname.zip containing:
  - the report (named report.pdf)
  - the training and test sets (named leukemia_train.dat and leukemia_test.dat)
  - the learned networks (named npc.net greedy.net nb.net)

NOTE

- No group work
- This assignment is mandatory in order to take the oral exam
- The assignment should be sent at least a week before the date of the oral exam