Bayesian Networks

May 4, 2020

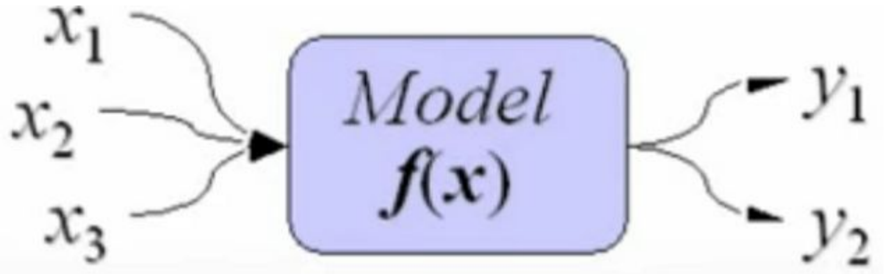
Presenter: Avirup Guha Neogi

b-it Bonn-Aachen International Center for Information Technology

Outline

- Probabilistic Graphical Models and probability preliminaries
- Frequentist and Bayesian Approaches
- Bayesian Networks (BN)
- Case Study - A BN for medical diagnosis of the respiratory system

Models



- Based on the concept of a declarative representation.
- This model encodes our knowledge of how the system works in a computer-readable form.
- Inference drawn by reasoning algorithms.
- Separation of knowledge and reasoning.
- Our focus is on complex systems that involve uncertainty.

Probabilistic Graphical Models (PGM)

- We need to reason not only about what is possible, but also about what is probable.

- Graphical representation helps us to visualise better and using Graph Theory properties.

- These models are statistical models that encode complex joint multivariate probability distributions using graphs.

- Used in inference and learning.

- Two flavours of PGMs are Bayesian Networks and Markov Random Fields.

Benefits of Probabilistic Models

- Modeling Power - Deterministic model is just a special case of probabilistic models.

- The power of doubt - The construction of a probabilistic model requires the systematic examination of all possible values of each variable and of each configuration of the object.

- Usability in the context of partial information - Practical Example: How to decide whether to eat flashy coloured mushrooms

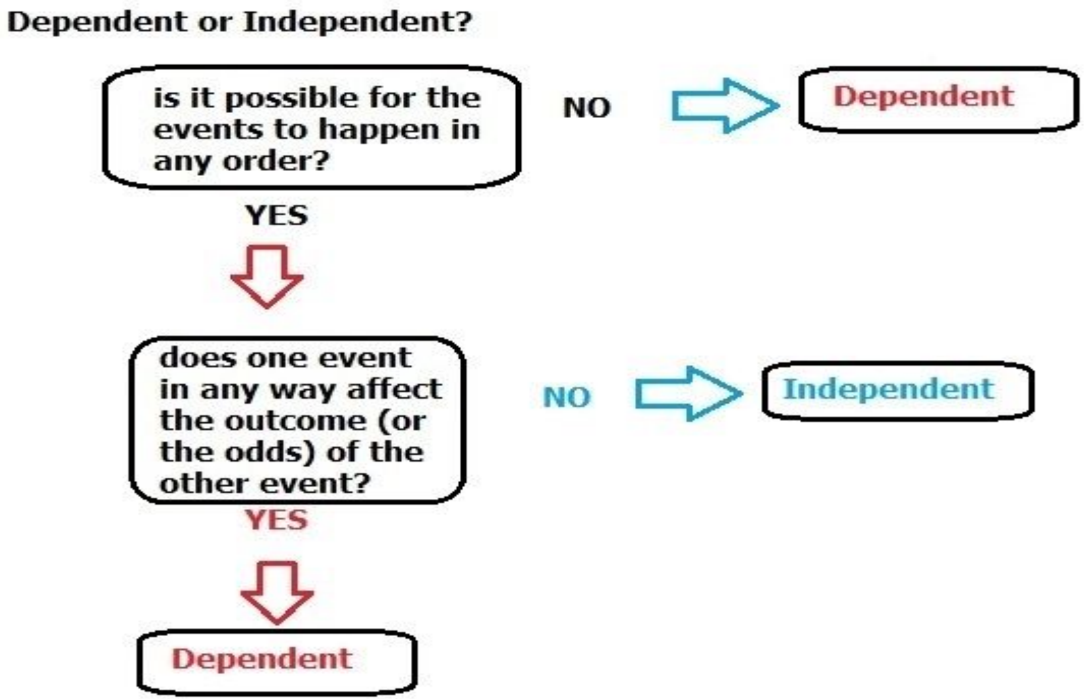
Drawbacks of PGM

- The joint probability distribution IP (X 1 , X 2 , . . . , X n) is rarely employed per se.

- It can be graphically represented only if n = 1 or 2. Even in the in the simplest non-trivial case n = p = 2

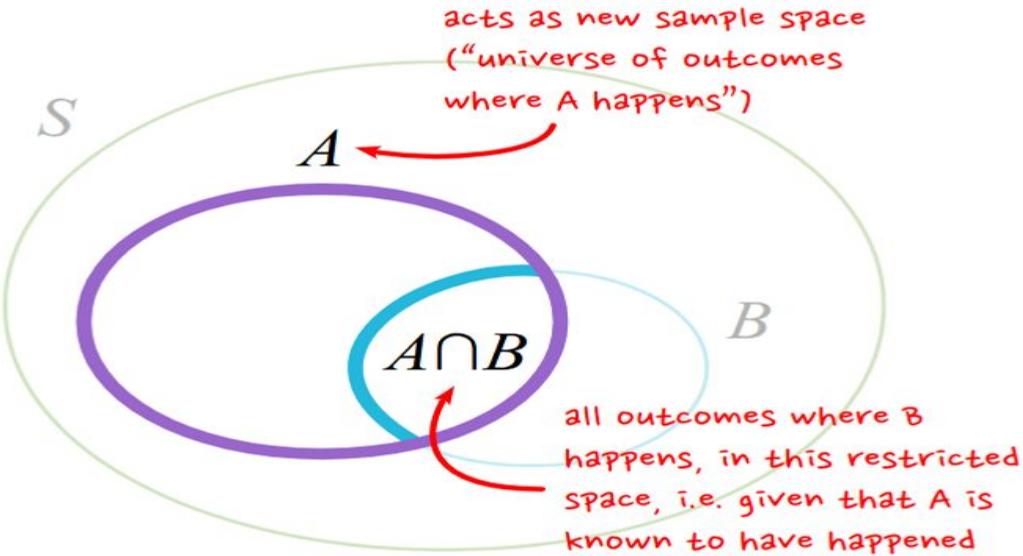
- The joint probability distribution gives rise to a phenomenon of combinatorial explosion. For instance, if each variable takes on p different values ($p \geq 1$), then the joint probability distribution has to be described by the probabilities of p n potential configurations of the object,

Dependent and Independent Events



Conditional Probability

- The conditional probability of B given A: $P(B|A) = P(A \cap B) / P(A)$



Independence of Events

- $P(A \cap B) = P(B|A)\, P(A)$ - From the definition of conditional probability
- Unconditional Independence : $P(A \cap B) = P(B)\, P(A)$, Assuming $P(A) \neq 0$ and iff $P(B|A) = P(B)$
- Conditional Independence : $P(A \cap B \mid C) = P(A|C)\, P(B|C)$, C is the event on which A and B become conditionally independent.

Bayes’ Theorem - The Theory That Would Not Die

- Assuming P(A), P(B) ≠ 0, P(A|B) = P(B|A) P(A) / P(B)

  Since P(A|B) P(B) = P(A⋂B) from the definition of conditional probability

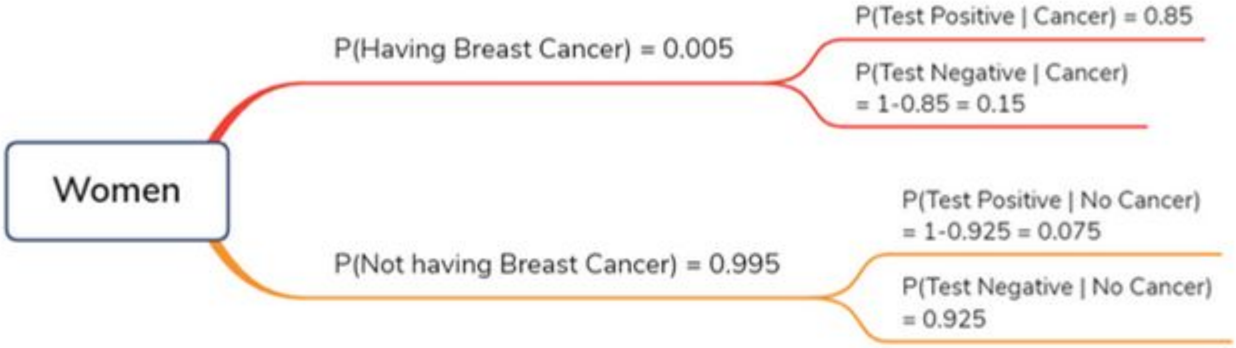
Prior probability of A

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Posterior probability of A, given evidence B

Bayes’ Theorem - A Practical Example

The probability of breast cancer among women in their mid-50s is 0.005. An established test identified people who had breast cancer and those that were healthy. A new mammography test in clinical trials has a probability of 0.85 for detecting cancer correctly. In women without breast cancer, it has a chance of 0.925 for a negative result. If a 55-year-old Caucasian woman tests positive for breast cancer, what is the probability that she, in fact, has breast cancer?



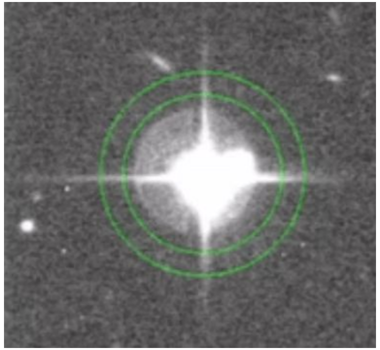
P(Cancer | Test Positive) = P(Cancer) * P(Test Positive | Cancer) / P(Test Positive)
P(Cancer | Test Positive) = 0.005 * 0.85 / (0.005 * 0.85 + 0.995 * 0.075) = 0.054

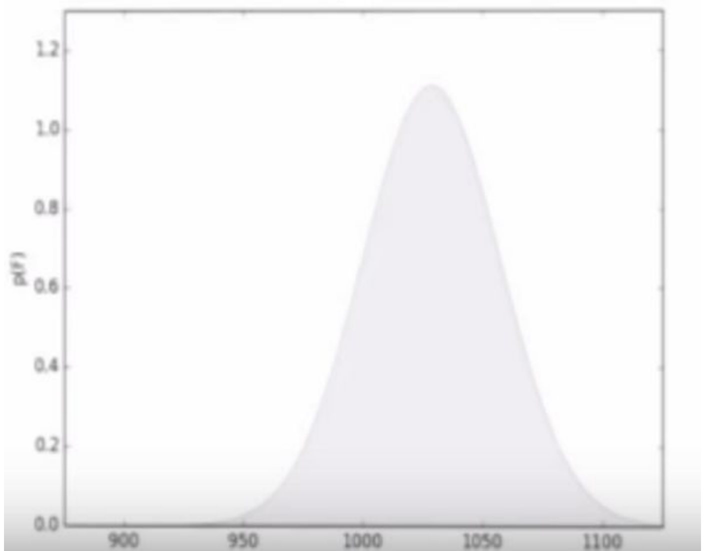
Frequentist Approach - Maximum Likelihood

- Fundamentally related to the frequency of related events

$$P(D_i \mid F_{\rm true}) = \frac{1}{\sqrt{2\pi e_i^2}} \exp\left[\frac{-(F_i - F_{\rm true})^2}{2e_i^2}\right]$$

- Models being <u>fixed</u> and data <u>vary</u> around them



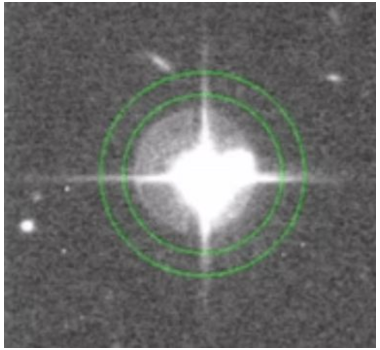


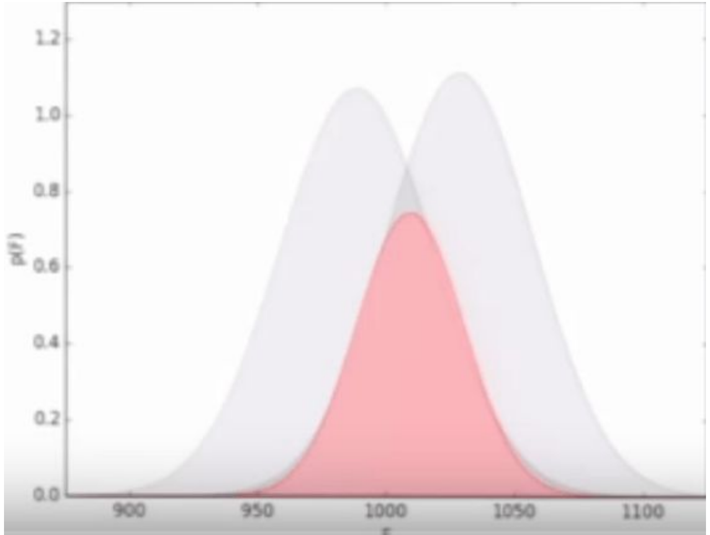
Frequentist Approach - Maximum Likelihood

- Fundamentally related to the frequency of related events

$$\mathcal{L}(D \mid F_{\text{true}}) = \prod_{i=1} P(D_i \mid F_{\text{true}})$$

- Models being <u>fixed</u> and data <u>vary</u> around them



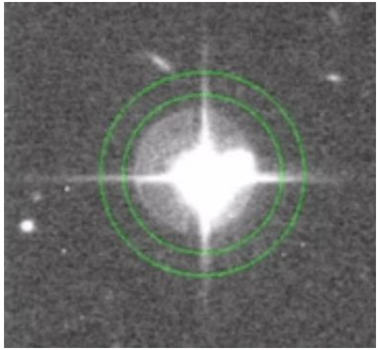


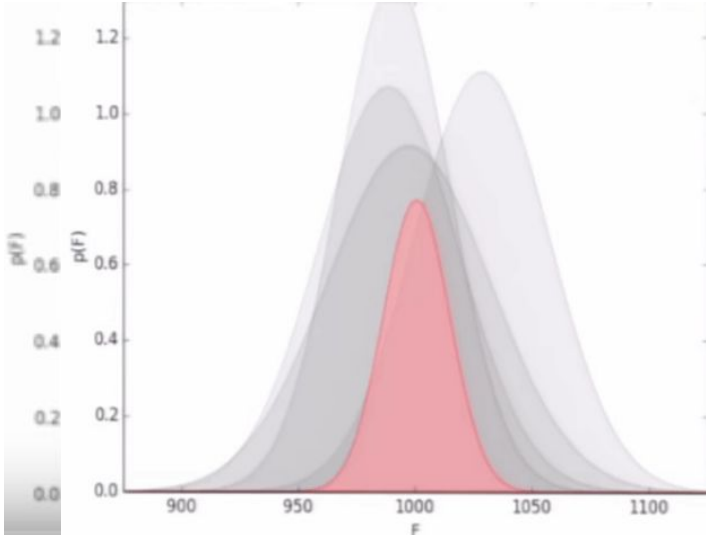
Frequentist Approach - Maximum Likelihood

- Fundamentally related to the frequency of related events

$$\mathcal{L}(D \mid F_{\text{true}}) = \prod_{i=1} P(D_i \mid F_{\text{true}})$$

- Models being <u>fixed</u> and data <u>vary</u> around them



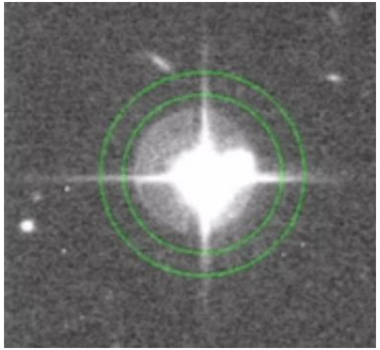


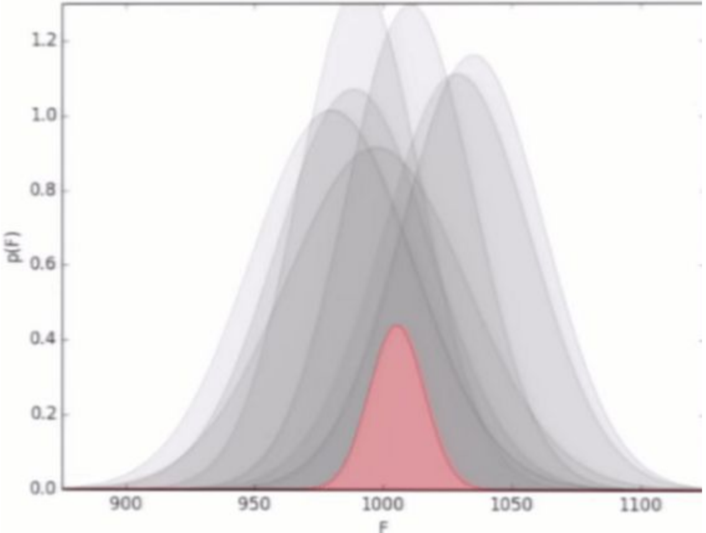
Frequentist Approach - Maximum Likelihood

- Fundamentally related to the frequency of related events

$$\mathcal{L}(D \mid F_{\text{true}}) = \prod_{i=1} P(D_i \mid F_{\text{true}})$$

- Models being <u>fixed</u> and data <u>vary</u> around them



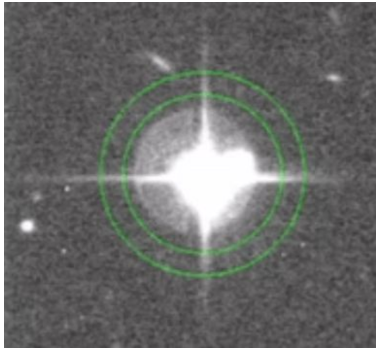


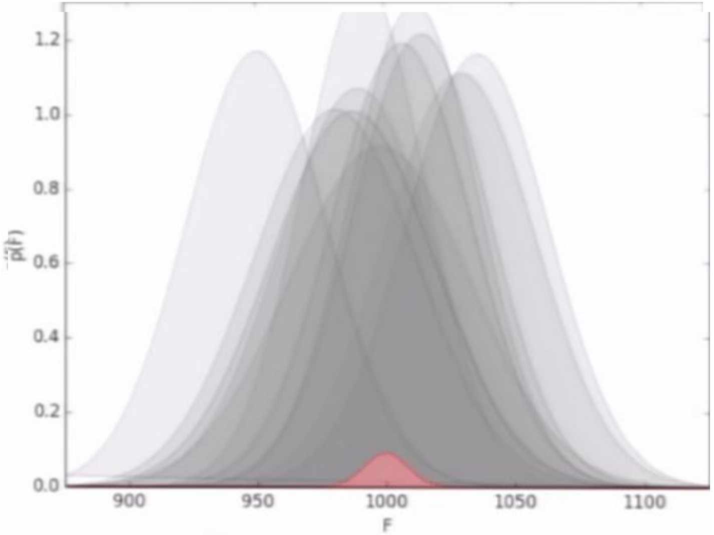
Frequentist Approach - Maximum Likelihood

- Fundamentally related to the frequency of related events

$$\mathcal{L}(D \mid F_{\text{true}}) = \prod_{i=1} P(D_i \mid F_{\text{true}})$$

- Models being <u>fixed</u> and data <u>vary</u> around them





Bayesian Approach - Posterior Probability

- Fundamentally related to our own certainty or uncertainty of events
- Observed data being fixed and models vary around those
- Computation of Knowledge F, given the data, encoded as probability.

$P(F_{true} | \text{Data})$

- To Compute the above probability, we use Bayes' Theorem as follows:

$P(F_{true} | \text{Data}) = P(D \mid F_{true})\, P(F_{true}) / P(D)$

- Setting a prior is a challenge - Emperical prior or Non-emperical prior.

Bayesian Approach - Maximum A Aposteriori Hypothesis (MAP)

- The goal of Bayesian learning is finding the <u>most probable hypothesis</u>, given training data:

  $h_{MAP} = \arg_{h \epsilon H} \max P(h|D) = \arg_{h \epsilon H} \max P(D|h)P(h) / P(D) = \arg_{h \epsilon H} \max P(D|h)P(h)$

  $h_{ML} = \arg_{h \epsilon H} \max P(h|D)$; h_{ML} is the most likely hypothesis when a non-emperical prior of P(h) is considered

Frequentist versus Bayesian - Practical Example

- Frequentist approach is the traditional approach in clinical trial design.
- Long process and takes years to reach completion
- Bayesian Approach shows promising results and adaptivity in Clinical trials.
- More convenient when dealing with new information as it can be incorporated either before the study or during the study
- Covid-19 vaccine developed by Oxford researchers to be undergoing mass production by Serum Institute of India and AstraZeneca

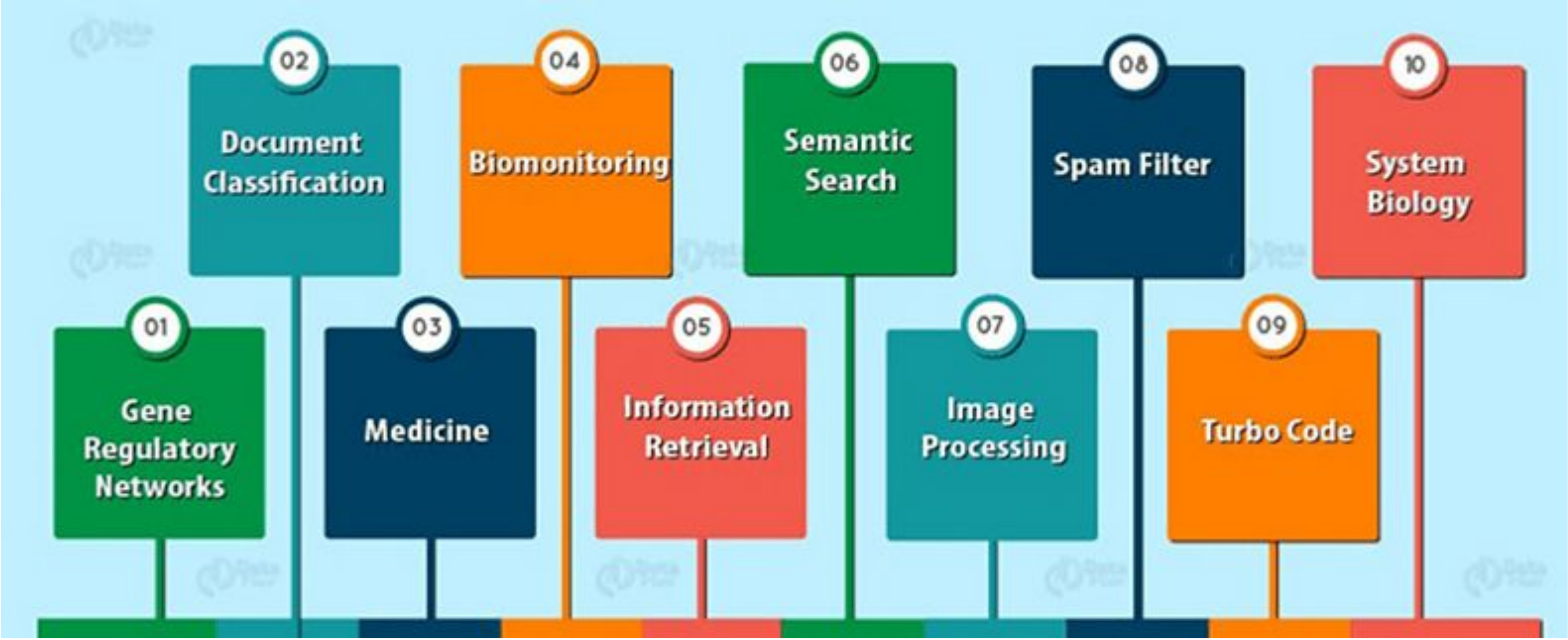
Bayesian Approach makes difference in Complex Models

- Handling of nuisance parameters
- Interpretation of uncertainty
- Incorporation of prior information
- Comparison and Evaluation of Models

Bayesian Networks (BN)

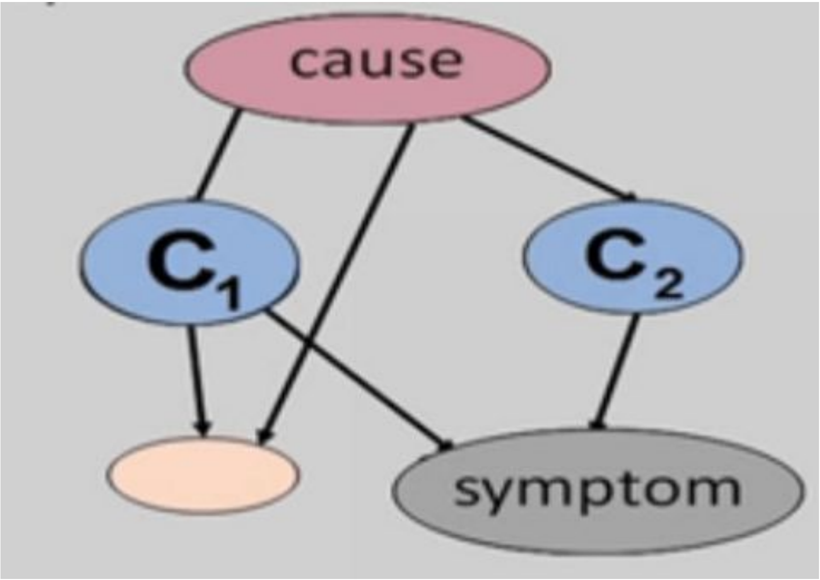
- A Bayesian network represents a joint distribution using a graph. Specifically, it is a directed acyclic graph in which each edge is a conditional dependency, and each node is a distinctive random variable.
- It has many other names: belief network, decision network, causal network, Bayes(ian) model or probabilistic directed acyclic graphical model, etc.
- It relies on the important assumption: each variable is conditionally independent of its non-descendants, given its parents.
- For a Bayesian Network, with a maximum of k parents for any node, we need only O(n * 2^k) probabilities.
- For n = 30 binary variables, k = 4 maximum parents for nodes, Unconstrained Joint Distribution: needs 2^30 (about 1 million) probabilities Bayesian Network needs only 480 probabilities

Fields where BN can be applied



Marrying BN with Machine Learning

- Diagnosis: P(cause|symptoms)
- Prediction: P(symptoms|cause)
- Classification: P(class|data)
- Decision-Making: Given a cost fn



Software Packages / Tools for bayesian Networks

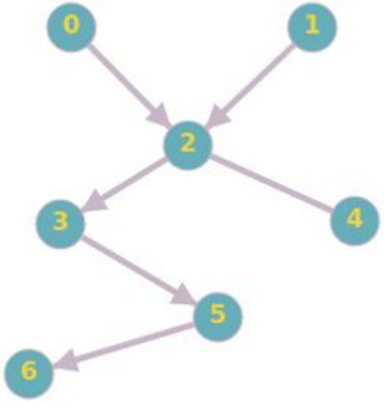
- Bayesian Network tools in Java - Java
- JavaBayes - Java
- jBNC - Java
- Stan: R, Python, Matlab, Julia , Stata
- Bayesian Network Inference with Java Objects - Java
- Structural Modeling, Inference, and Learning Engine - C++
- BayesPy - Python

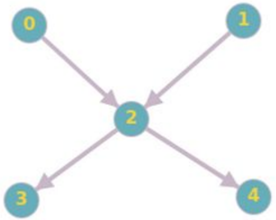
d-separation: Determining Independent Variables

- Construct the ancestral graph of all variables mentioned in the probability expression.
- Moralize the ancestral graph by "marrying" the parents. For each pair of variables with a common child, draw an undirected edge (line) between them. (If a variable has more than two parents, draw lines between every pair of parents.)
- Disorient the graph by replacing the directed edges (arrows) with undirected edges (lines).
- Delete the givens and their edges.If the independence question had any given variables, erase those variables from the graph and erase all of their connections, too.
- If the variables are disconnected in this graph, they are guaranteed to be independent. If the variables are connected in this graph, they are not guaranteed to be independent.
- If one or both of the variables are missing (because they were givens, and were therefore deleted), they are independent.

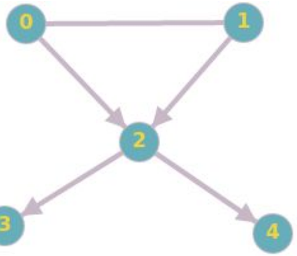
d-separation - Example

Are 3 & 4 conditionally independent given 2?

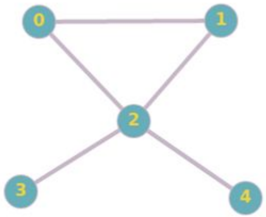




Ancestral Graph



Moralize



Disorient

Delete Givens

Case Study: A BN for medical diagnosis of the respiratory system.

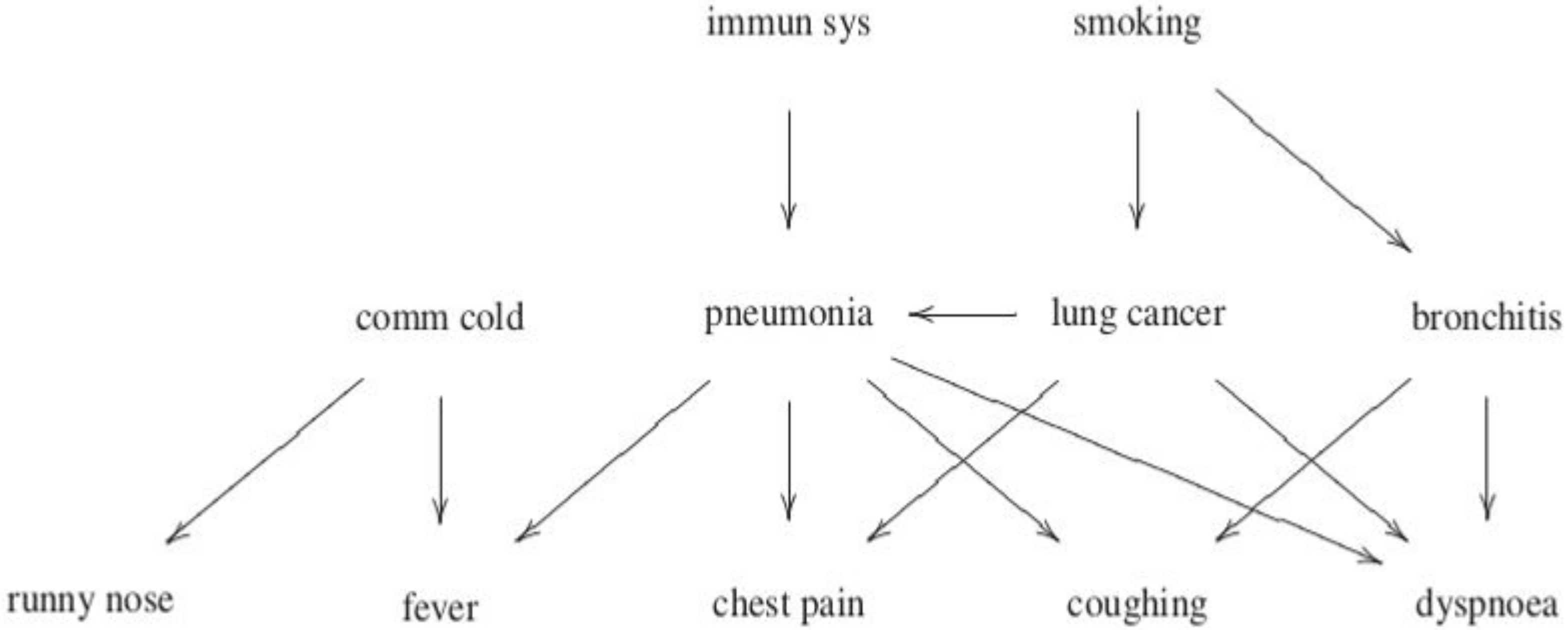
The symptom dyspnoea (shortness of breath) may be due to the diseases pneumonia, lung cancer, and/or bronchitis. Patients with pneumonia, and/or bronchitis often have a very nasty wet coughing. Pneumonia, and/or lung cancer are often accompanied by a heavy chest pain. Pneumonia is often causing a severe fever, but this may also be caused by a common cold. However, a common cold is often recognized by a runny nose. Sometimes, wet coughing, chest pain, and/or dyspnoea occurs unexplained, or are due to another cause, without any of these diseases being present. Sometimes diseases co-occur. A weakened immune-system (for instance, homeless people, or HIV infected) increases the probability of getting an pneumonia. Also, lung cancer increases this probability. Smoking is a serious risk factor for bronchitis and for lung cancer.

Modeling - Qualitative

- Identifying the variables.
- All variables are assumed to be binary.
- Model allows diseases to co-exist.
- DAG is constructed using causal modeling.
- General causal modeling assumption: Risk factors cause disease and disease cause symptoms.
- Risk factors are weakened immune-system (for pneumonia), smoking (for bronchitis and for lung cancer), and lung cancer (also for pneumonia).
- At the first level, the diseases and risk factors are modeled.
- Common cold has no risk factor.
- At the second level, the symptoms are modeled.

Modeling - Qualitative

DAG of the variables from the qualitative description



Modeling - Quantitative

- Determining the conditional probability tables(CPT).
- The numbers entered are arbitrary guesses.
- In determining the conditional probabilities, we used some modeling assumptions such as that the probability of a symptom in the presence of an additional causing diseases is at least as high as the probability of that symptom in the absence of that disease.
- The left column represents the probability values in the true state, P(variable name) ≡ P(variablename = true)
- The other columns indicate the joint states of the parent variables.

Modeling - Quantitative

P(immun syst)
0.05

P(smoking)
0.3

P(common cold)
0.35

P(lung cancer	smoking)
0.1	true
0.01	false

P(bronchitis	smoking)
0.3	true
0.01	false

P(runny nose	common cold)
0.9	true
0.01	false

P(pneumonia	immun syst,	lung cancer)
0.3	true	true
0.3	true	false
0.05	false	true
0.001	false	false

P(fever	pneumonia,	common cold)
0.9	true	true
0.9	true	false
0.2	false	true
0.01	false	false

P(cough	pneumonia,	bronchitis)
0.9	true	true
0.9	true	false
0.9	false	true
0.1	false	false

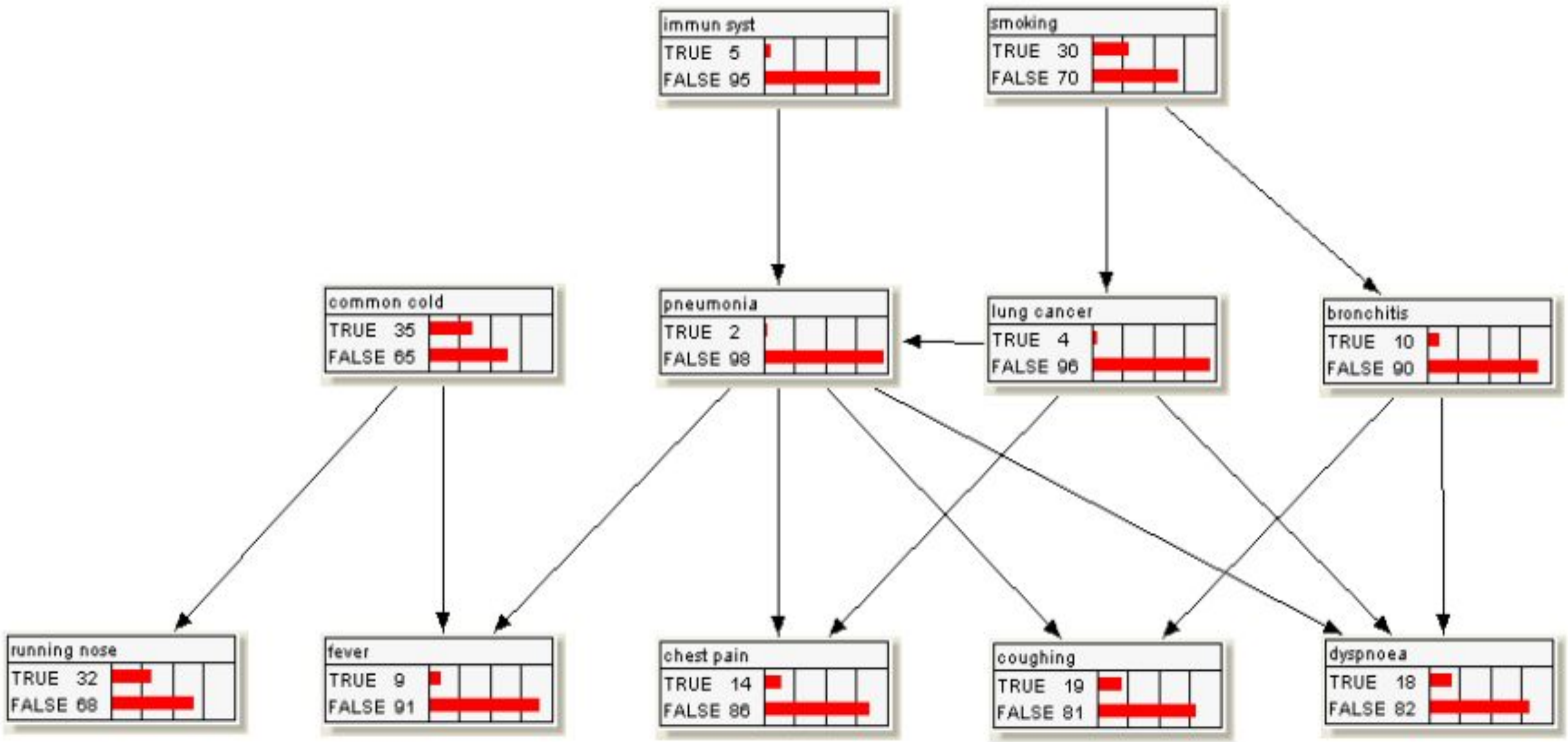
P(chest pain	pneumonia,	bronchitis)
0.9	true	true
0.9	true	false
0.9	false	true
0.1	false	false

P(dyspnoea	bronchitis,	lung cancer,	pneumonia)
0.8	true	true	true
0.8	true	true	false
0.8	true	false	true
0.8	true	false	false
0.5	false	true	true
0.5	false	true	false
0.5	false	false	true
0.1	false	false	false

Reasoning

- The BN is realized using BayesBuilder and is visualized for better interpretation.
- The program uses the junction tree inference algorithm to compute the marginal node probabilities and displays them on screen.
- The marginal node probabilities are the probability distributions of each of the variables in the absence of any additional evidence.
- In the program, evidence can be entered by clicking on a state of the variable. This procedure is sometimes called ‘clamping’.
- The node probabilities will then be conditioned on the clamped evidence. With this, we can easily explore how the models reasons.

Reasoning



immun syst
TRUE 5
FALSE 95
smoking
TRUE 30
FALSE 70
common cold
TRUE 35
FALSE 65
pneumonia
TRUE 2
FALSE 98
lung cancer
TRUE 4
FALSE 96
bronchitis
TRUE 10
FALSE 90
running nose
TRUE 32
FALSE 68
fever
TRUE 9
FALSE 91
chest pain
TRUE 14
FALSE 86
coughing
TRUE 19
FALSE 81
dyspnoea
TRUE 18
FALSE 82

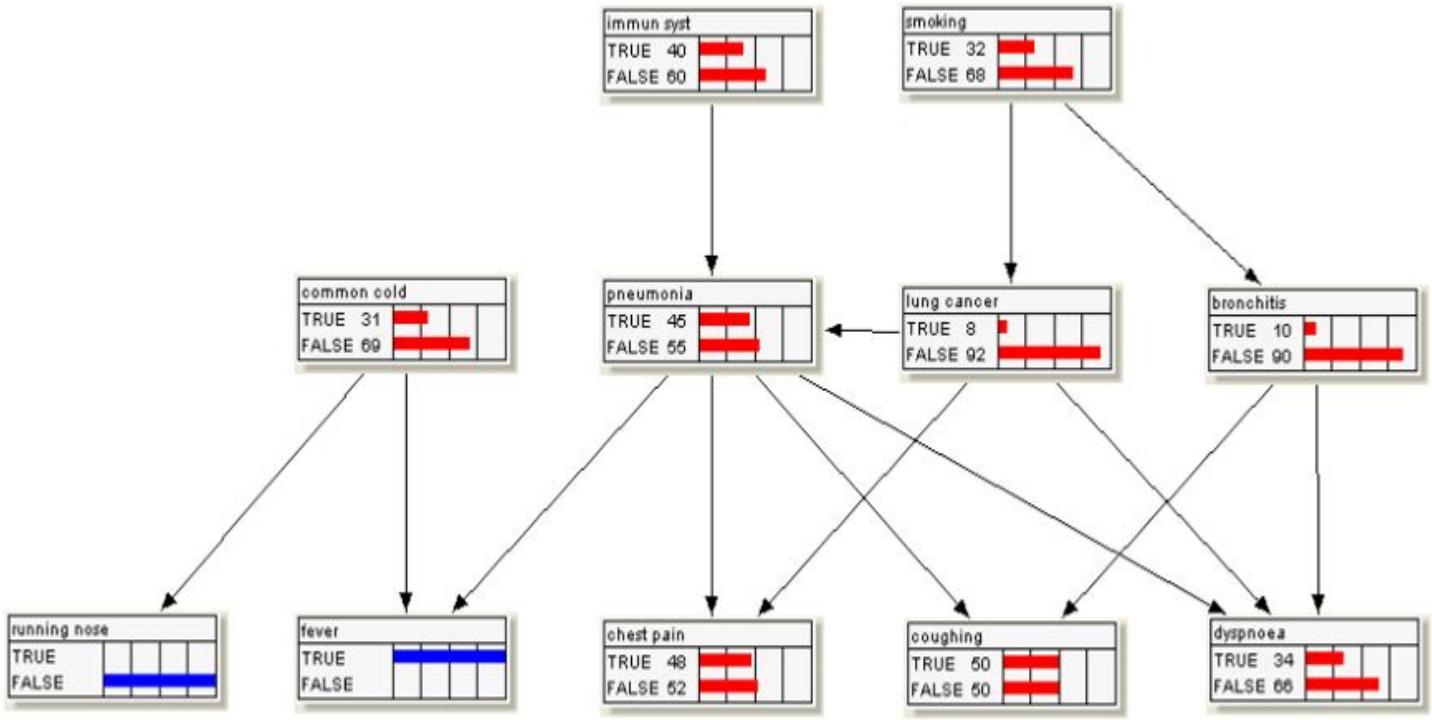
Knowledge Representation

- The BN will serve as a rich knowledge-base.
- Some hypothetical medical guidelines are given below.
- In case of high fever in absence of a runny nose, pneumonia should be considered.

- Inference: We clamp fever = true and runny nose = false and look at the conditional probabilities of the four diseases. We see that in particular the probability of pneumonia is increased from about 2% to 45%.

- Comment: There are two causes in the model for fever, namely has parents pneumonia and common cold. However, the absence of acommon cold makes common cold less likely. This makes the other explaining cause pneumonia more likely

Knowledge Representation



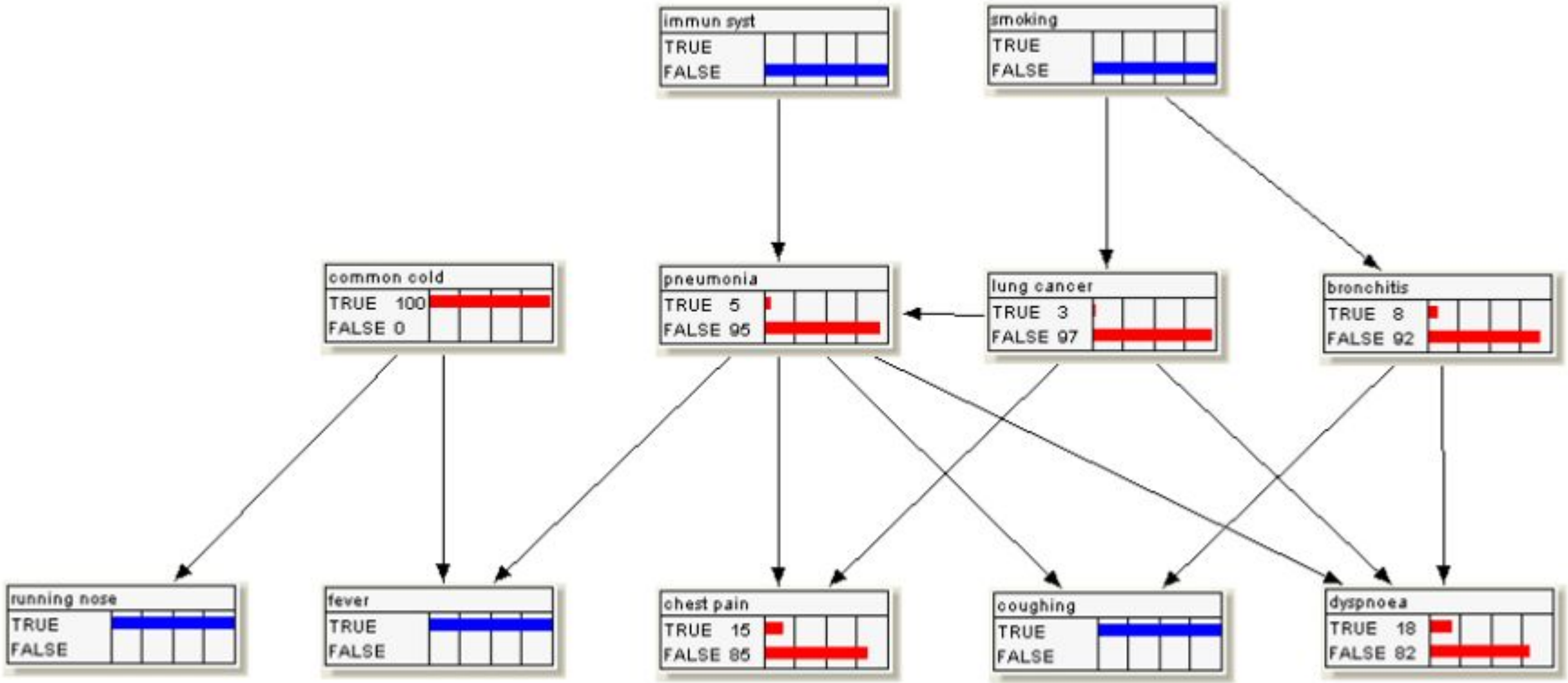
Diagnostic Reasoning

- The idea is to enter the patient observations, i.e. symptoms and risk factors into the system. Then diagnosis (i.e. finding the cause(s) of the symptoms) is done by Bayesian inference.

- <u>Hypothetical Case:</u> Mr. Appelflap calls. He lives with his wife and two children in a nice little house in the suburb. You know him well and you have good reasons assume that he has no risk of a weakened immune system. Mr. Appelflap complains about high fever and a nasty wet cough (although he is a non-smoker). In addition, he sounds rather nasal. What is the diagnosis?

- <u>Inference</u>: We clamp the risk factors immun sys = false, smoking = false and the symptoms fever = true, runny nose = true. We find all disease probabilities very small, except common cold, which is almost certainly true.

Diagnostic Reasoning



Advantages of Bayesian Networks

- Can be used for statistically-based learning.
- Can in some cases outperform other learning methods.
- Prior knowledge can be (incrementally) combined with newer knowledge to better approximate perfect knowledge.
- Can make probabilistic predictions.

Disadvantages of Bayesian Networks

- The Computationally expensive. Eg: Approximate structure learning is too NP-Complete
- Forces random variables to be in a cause-effect relationship. As a result, it does not depicts variables which are correlated. BN does not provides any guarantee of depicting the cause and effect relationship.
- BN is a DAG that said. If the data was generated from a model where there at least 3 variables correlated to each other (cyclic relationship) then Bayesian networks (BNs) will not be able to model this relationship.
- One of the most important issues with BNs is that some of the sophisticated scoring functions require reliable priors in order to find a structure closer to the original model

Concluding Remarks

- Bayesian Networks provide an efficient way to model the causality relationships and make predictions based on the same.
- They are robust and ideal for modeling real-life complex systems.
- A bit on the philosophical side: Our thought process is heavily influenced on the initial belief and new information gathered. Accordingly, the beliefs get updated.

References

Book Chapters:

Wiegerinck, Wim, Willem Burgers, and Bert Kappen. "Bayesian networks, introduction and practical applications." *Handbook on Neural Information Processing*. Springer, Berlin, Heidelberg, 2013. 401-431.

Books:

1. Koller, Daphne, and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.
2. Husmeier, Dirk, Richard Dybowski, and Stephen Roberts, eds. *Probabilistic modeling in bioinformatics and medical informatics*. Springer Science & Business Media, 2006.
3. Pourret, Olivier, Patrick Naïm, and Bruce Marcot, eds. *Bayesian networks: a practical guide to applications*. John Wiley & Sons, 2008.

Web Resoures :

1. https://hunch.net/ 2. https://towardsdatascience.com/ 3. https://data-flair.training/

Thank You