Bayesian Network

Some slides were adapted/taken from various sources, including Prof. Andrew Ng's Coursera Lectures, Stanford University, Prof. Kilian Q. Weinberger's lectures on Machine Learning, Cornell University, Prof. Sudeshna Sarkar's Lecture on Machine Learning, IIT Kharagpur, Prof. Bing Liu's lecture, University of Illinois at Chicago (UIC), CS231n: Convolutional Neural Networks for Visual Recognition lectures, Stanford University and many more. We thankfully acknowledge them. Students are requested to use this material for their study only and **NOT** to distribute it.

Problems with Naïve Bayes Classifier

- Unrealistic Assumption: Often X_i are not really conditionally independent

- Surprisingly, often the right classification, even when not the right probability

- Problems at hand: $P(X_1 \ldots X_n | Y)$

- Since calculating joint probability is intractable, we have used Naïve Bayes assumption which has been found much restrictive and sometime unrealistic.

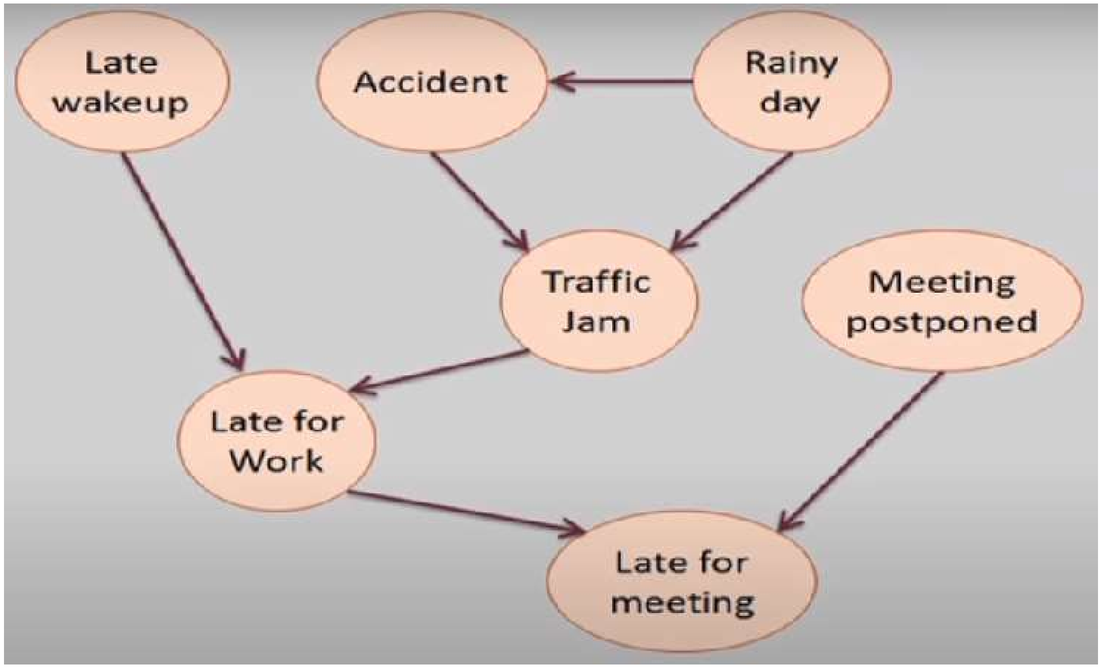
Bayesian Networks

- Problems at hand: $P(X_1 \dots X_n|Y)$
- Need not make full independent assumption or full dependent assumptions
- Rather, we denote the causal relationships and specific conditional independence of different attributes.
- In Belief networks, we also denotes causal relationships.
- Different approaches: Belief Network: Directed Graphical Model, Bayesian network: Un-directed Graphical model

Bayes Networks

- Directed type of graphical model which represents conditional independence relationships between different variables in the domain
- Helps to get tractable inference where all the variables are not connected (dependent) to each other
- Interested to represents the dependence (or independence) between different variables in the domain [$X_1 \ldots X_n$ Y]
- Nodes: represents the variables
- Arcs: represents the relation between the variables, lack of arcs represents the independence between the variables in certain ways
- Conditional independence relationships
- Causality: Some variables are the causes (effects) of the others
- If we represents the causality, we can get a more compact Bayesian network. However, Bayesian network without causality exists.

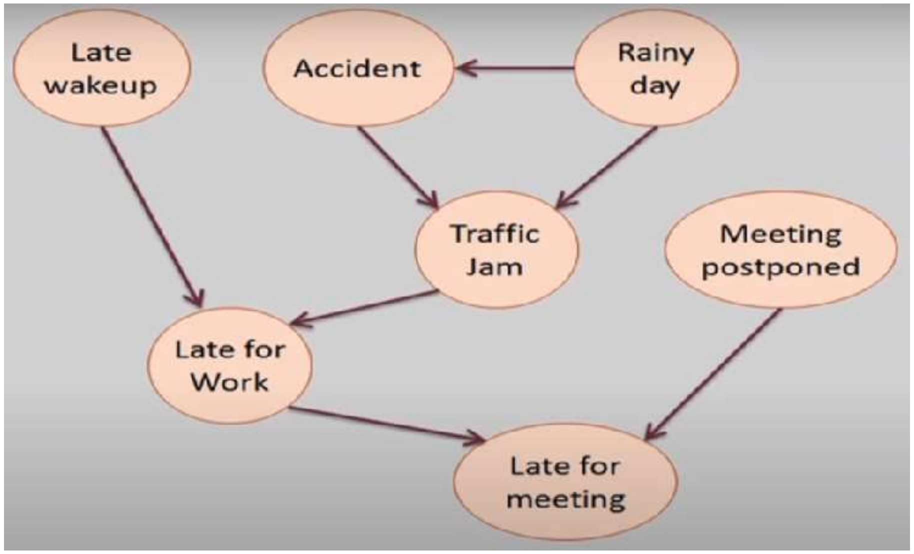
Bayes Networks: Example



- Arcs denotes the causal relationships between the variables (represented by nodes)
- Directed acyclic graph (DAG) represents the nodes and the relationships between them.
- Different conditional relationship can also be observed from this graph.

- "Late wake up" influences "Late for work". Again "Late for work" influences "Late for meeting". So, we can see "Late for meeting" may or may not be influenced by the node "Late wake up". Because, one may not be "Late for work" even if she/he woken up late. So, the node "Late wake up" and "Late for meeting" is not normally independent rather they are conditionally independent given if you know whether she/he is "Late for work".

Bayes Networks: Example



- Again, consider two nodes, "Late wakeup" and "Accident". These two nodes normally seems independent but given the node "Late for work" is true, they are may be related. Because, one didn't waken up late but still "Late for work", "Accident" may be the cause for the "Late for work".
- Similar for "Meeting postponed" and "Late for work". They are normally independent but may be related through the node "Late for meeting".

Bayesian Networks

- Bayesian network is a graphical representation which represents efficiently the joint probability distribution of the variables.

- Any probability or conditional probability of interest can be computed if you know the full joint probability distribution which is computationally expensive (intractable).

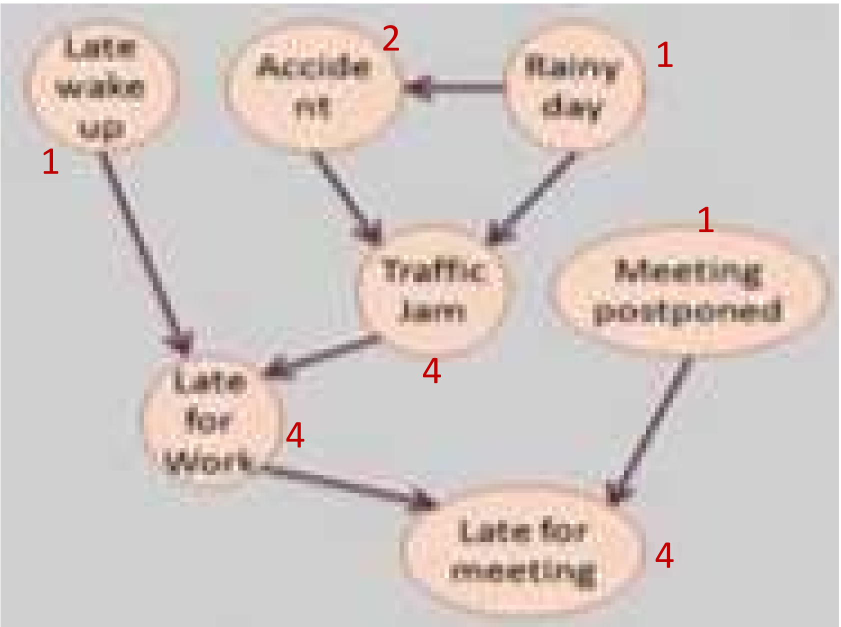
- Bayesian network (DAG) represents joint portability more compactly. Specially if the graph does not have too may edges.

Bayesian Networks

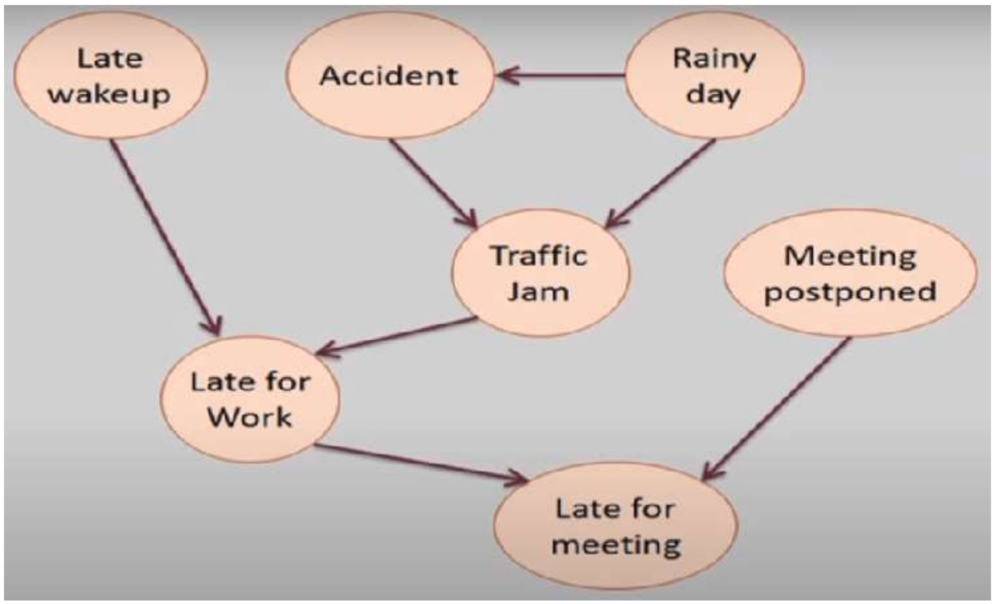
- In Bayesian network nodes are represented by X= $\{X_1 \ldots X_n\}$
- Arcs represents probabilistic dependence (or independence) among variables
- Absent of arcs denote independence or conditional independence.
- The network structure is a Directed Acyclic Graph (DAG).
- At each node, a local probability table is maintained which is called conditionally probability table.

Bayesian Networks

- For simplicity, let us assume, all the variables are Boolean. So for our example, there are 2^7 possible combinations and each having a probability.
- At the conditional probability, we have to keep the probability distribution of the node given the value of its parents.
- Now for the nodes having no parents can have one value. A nodes having a parent has $2^1=2$ values. A nodes having two parents has $2^2=4$ values and so on.
- So we have to store 17 probability values for the table whereas for representing fully connected Bayesian network, we have to store $2^7 - 1$ no. of probability values.



Bayesian Networks



Conditional probability table associated with each node specifies the conditional distribution for the variable given it immediate parents in the graph.

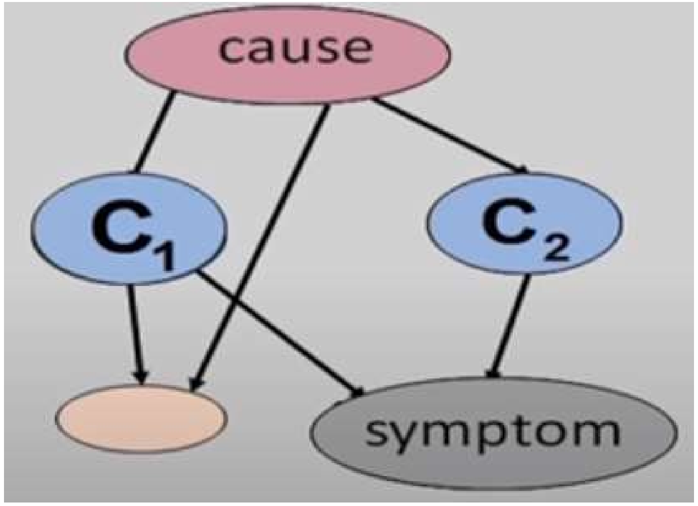
Each node is asserted to be conditionally independent of its non-descendants, given its immediate parents

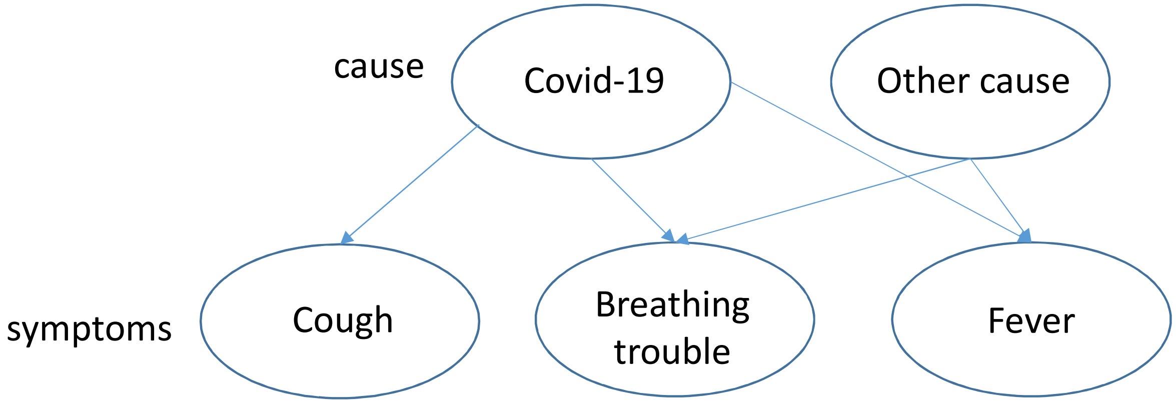
Inference

- Computes posterior probabilities given evidence about some nodes.
- Exploits the probabilistic conditional independence for efficient computations.
- Unfortunately, exact inference of probabilities in general for an arbitrary Bayesian Network is known to be NP-hard.
- In theory, approximate techniques (such as Monte Carlo methods) can also be NP-hard. Though in practice, many such methods were shown to be useful.
- Efficient algorithms that leverages the structures of the graph.

Applications of Bayesian Networks

- Diagnosis : P(cause|symptom)=?
- Prediction P(symptom|cause)=?
- Classification: P(class|data)
- Decision making (given a cost function)





Definition

- Structure of a graph: Conditional independence relations
- In general: $P(X_1, X_2, \ldots, X_3) = \prod P(X_i | parents(X_i))$

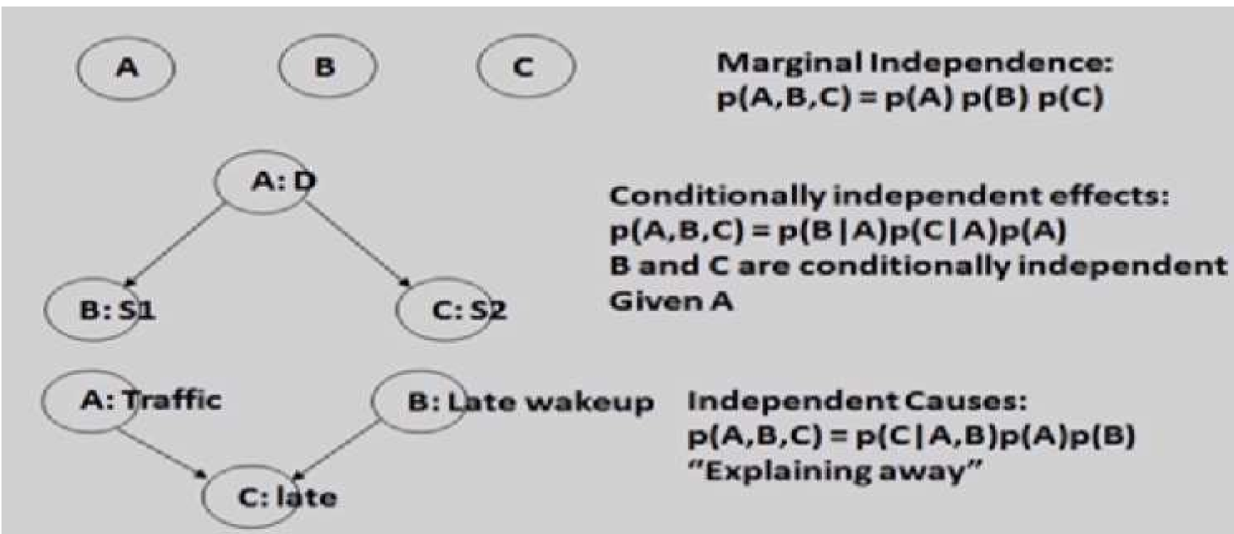
Full joint distribution

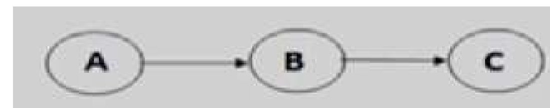
The graph structured approximations

- Requires that the graph is acyclic (no directed cycles)
- Two components to a Bayesian Network
  - The graph structures (conditional independence assumptions)
  - The numerical probabilities (for each variable given its parents)

Examples:

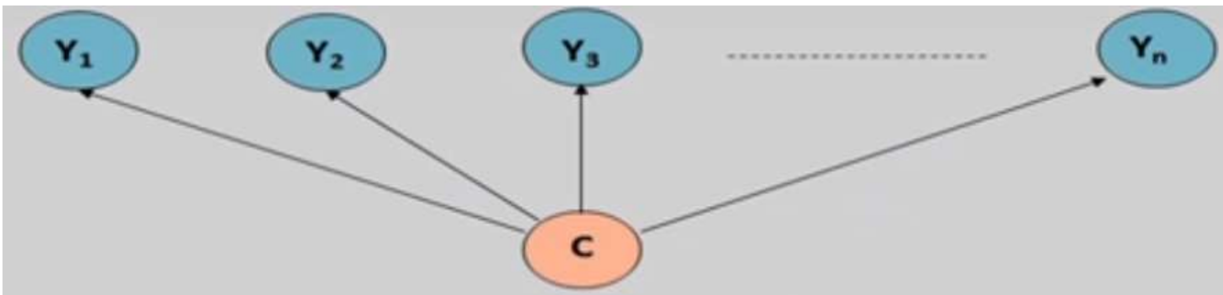
- $P(X_1, X_2, \ldots, X_3) = P(X_1)P(X_2|X_1)P(X_3|X_1X_2)\ldots P(X_n|X_1 \ldots X_{n-1})$
  $= \prod P(X_i|parents(X_i))$





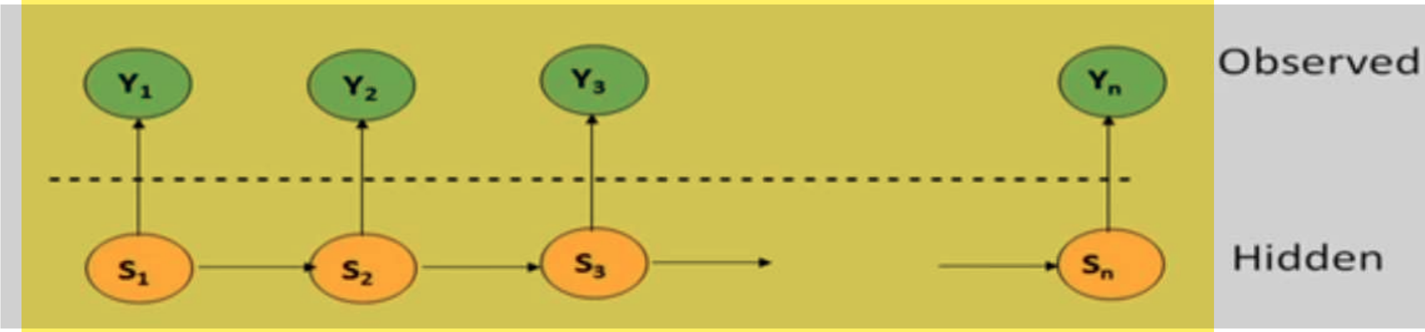
Markov dependence:
p(A,B,C) = p(C|B) p(B|A)p(A)

Examples of Different Bayesian Networks



Naïve Bayes Model

Hidden Markov Model (HMM)



Observations Y_t are conditionally independent of all other variables given S_t, so the observation at time t depends only on the current state S_t. ... The s_t's form a (first order) Markov chain, i.e., $p(s_t|s_{t-1},...,s_1) = p(st|st-1), t = 2,...,T$.

Assumptions:

- Hidden state sequence are markov
- Observations Y_t is conditionally independent of all other variables given S_t
- Widely used in sequence learning
- Inference is linear in n