Bayesian Networks

Leonid Zhukov

School of Applied Mathematics and Information Science
National Research University Higher School of Economics

28.11.2013

Basic Probability

- Joint probability $P(A, B)$
- Marginalization (sum rule): $P(A) = \sum_B P(A, B)$
- Conditioning (product rule): $P(A, B) = P(B|A)P(A)$
- Bayes Theorem

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

Independence

- Independence:

$$P(A, B) = P(A)P(B)$$

- Conditional independence:

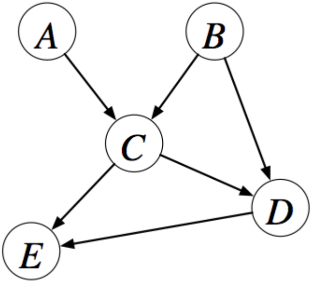
$$\begin{aligned} P(A|B, C) &= P(A|C) \\ P(B|A, C) &= P(B|C) \\ P(A, B|C) &= P(A|B, C)P(B|C) = P(A|C)P(B|C) \end{aligned}$$

Notation

$$A \perp\!\!\!\perp B \mid C$$

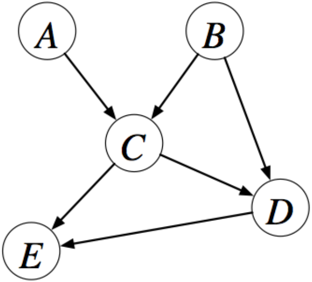
Basian Networks

- Directed graphical models
- Basian Networks are directed acyclic graphs, DAG
- Nodes - random variables
- Edges - dependencies between random variables (direct influence)



Basian Networks

- Each node associated with conditional distribution P(X|parents)
- Parents - nodes, sending arrows to X
- Root nodes associated with priors P(X)

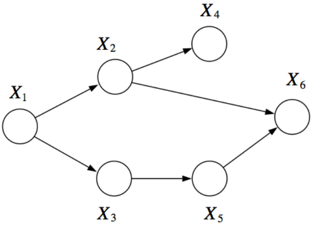


Joint distribution

- $P(x_1, x_2) = P(x_2|x_1)P(x_1)$
- $P(x_1, x_2, x_3) = P(x_3|x_1, x_2)P(x_2|x_1)P(x_1)$
- $P(x_1, ...x_k) = P(x_k|x_1, ...x_{k-1})..P(x_2|x_1)P(x_1)$
- Fully connected, link between every pair of nodes
- Missing edges means conditional independence
- Factorization properties of joint distribution

$$P(x_1..x_k) = \prod_{k=1}^{K} P(x_k|parents(x_k))$$

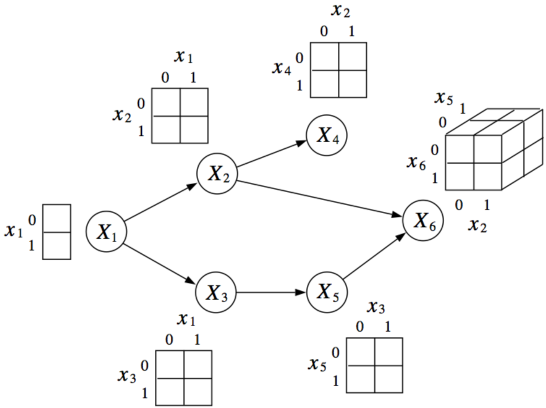
Directed graphical model



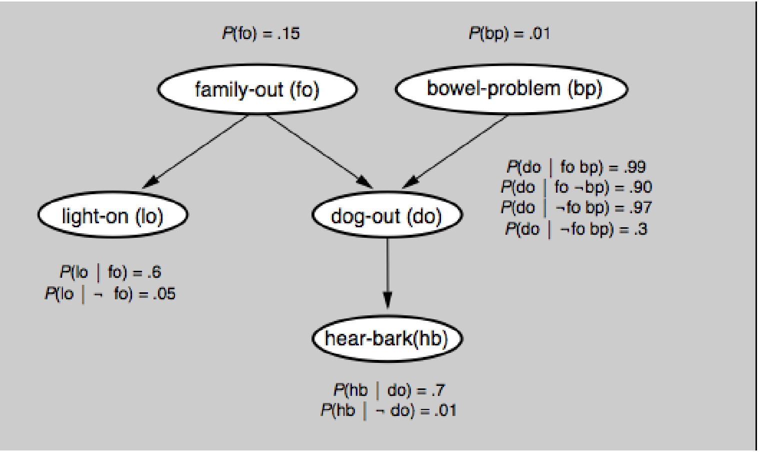
Factorization:

$$P(x_1, x_2, x_3, x_4, x_5, x_6) = P(x_1)P(x_2|x_1)P(x_3|x_1)P(x_4|x_2)P(x_5|x_3)P(x_6|x_2, x_5)$$

Directed graphical model

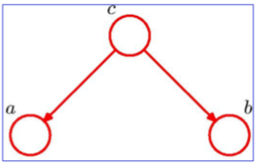


Example

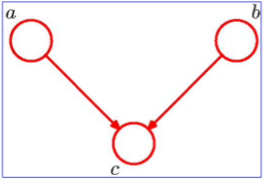


P(fo) = .15
P(bp) = .01
family-out (fo)
bowel-problem (bp)
light-on (lo)
dog-out (do)
hear-bark(hb)
P(lo | fo) = .6
P(lo | ¬ fo) = .05
P(do | fo bp) = .99
P(do | fo ¬bp) = .90
P(do | ¬fo bp) = .97
P(do | ¬fo bp) = .3
P(hb | do) = .7
P(hb | ¬ do) = .01

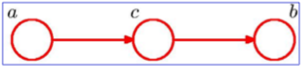
3 canonical graphs



$p(a,b,c)=p(a|c)p(b|c)p(c)$



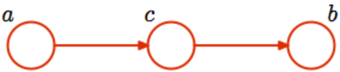
$p(a,b,c)=p(a)p(b)p(c|a,b)$



$p(a,b,c)=p(a)p(c|a)p(b|c)$

19

Head-to-tail node

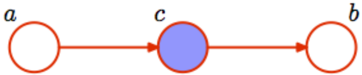


Independence:

$$p(a,b) = \sum_c p(a,b,c) = \sum_c p(a)p(c|a)p(b|c) =$$
$$= p(a)\sum_c p(b|c)p(c|a) = p(a)p(b|a) \neq p(a)p(b)$$

not independent! $a \not\perp\!\!\!\perp b$

Head-to-tail node

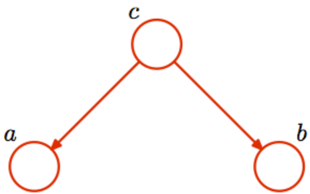


Conditional independence:

$$p(a, b|c) = \frac{p(a, b, c)}{p(c)} = \frac{p(a)p(c|a)p(b|c)}{p(c)} = p(a|c)p(b|c)$$

Conditionally independend: $a \perp\!\!\!\perp b \mid c$

Tail-to-tail node

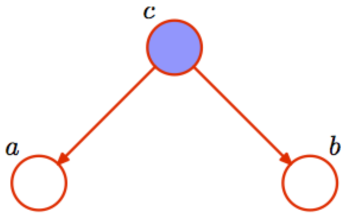


Independence:

$$p(a,b) = \sum_c p(a,b,c) = \sum_c p(a|c)p(b|c)p(c) \neq p(a)p(b)$$

not independent! $a \not\perp\!\!\!\perp b$

Tail-to-tail node

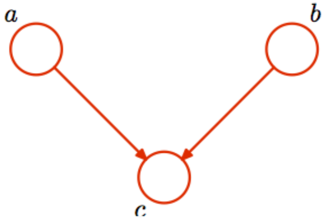


Conditional independence:

$$p(a, b|c) = \frac{p(a, b, c)}{p(c)} = \frac{p(a|c)p(b|c)p(c)}{p(c)} = p(a|c)p(b|c)$$

Conditionally independend: $a \perp\!\!\!\perp b \mid c$

Head-to-head node

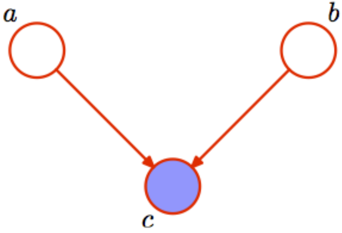


Independence:

$$p(a,b) = \sum_c p(a,b,c) = \sum_c p(c|a,b)p(a)p(b) =$$
$$= p(a)p(b)\sum_c p(c|a,b) = p(a)p(b)$$

independent! $a \perp\!\!\!\perp b$

Head-to-head node

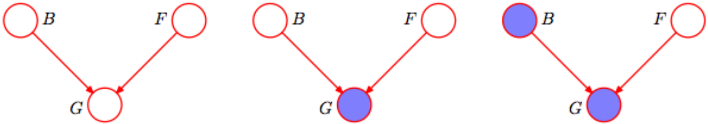


Conditional independence:

$$p(a, b|c) = \frac{p(a, b, c)}{p(c)} = \frac{p(c|a, b)p(a)p(b)}{p(c)} \neq p(a|c)p(b|c)$$

Conditionally not independend: $a \not\perp\!\!\!\perp b \mid c$

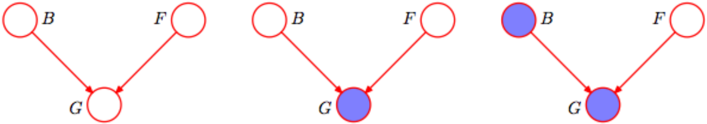
Conditional independence



- Binary variables, fuel system on a car
- B - battery, B=1 charged, B = 0 dead
- F - fuel tank, F = 1 full F = 0 empty
- G - electric gauge, G=1 shows full tank, G =0 shows empty tank
- Prior distribution P(B = 1) = 0.9, P(F = 1) = 0.9
- Gauge is unreliable! Conditional on gauge read "full"

$$P(G = 1|B = 1, F = 1) = 0.8$$
$$P(G = 1|B = 1, F = 0) = 0.2$$
$$P(G = 1|B = 0, F = 1) = 0.2$$
$$P(G = 1|B = 0, F = 0) = 0.1$$

Conditional independence

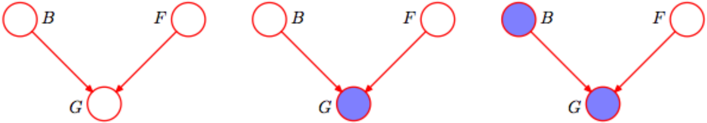


- Observe fuel gauge, reads "empty G = 0. Is tank really empty? $P(F = 0|G = 0)$- ?
- Bayes

$$P(F = 0|G = 0) = \frac{P(G = 0|F = 0)P(F = 0)}{P(G = 0)}$$

- $P(G = 0|F = 0) = P(G = 0|B = 0, F = 0)P(B = 0) + P(G = 0|B = 1, F = 0)P(B = 1) = 0.9 * 0.1 + 0.8 * 0.9 = 0.81$
- $P(G = 0) = \sum_{B \in 0,1} \sum_{F \in 0,1} P(G = 0|B, F)P(B)P(F) = 0.315$
- $P(F = 0|G = 0) = 0.257 > P(F = 0) = 0.1$

Conditional independence



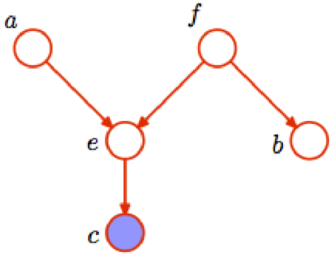
- We also check (observe) battery, it is dead, B = 0
- Bayes

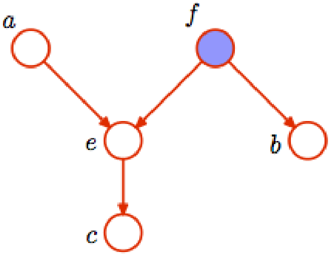
$$P(F=0|G=0,B=0) = \frac{P(G=0|B=0,F=0)P(F=0)P(B=0)}{\sum_{F\in 0,1} P(G=0|B=0,F)P(F)P(B=0)}$$

- $P(F=0|G=0,B=0) = 0.111$
- State of fuel tank and battery became dependent through the gauge observation

Blocked path

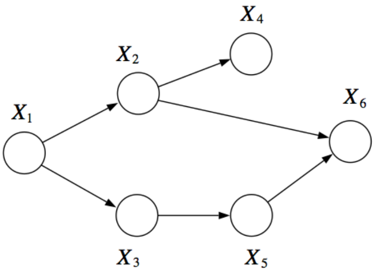
- An observed TT or HT node, or
- A HH node which is not observed, nor any of its descendants is observed





D-separatioin

- A set of nodes A is said to be d-separated from a set of Types of Graphical Models п»ёnodes B by a set of nodes C if every path from A to B is blocked when C is in the conditioning set.



Factorization

- Theorem: Factorization -> CI
  If a probability distribution factorizes according to a directed acyclic graph, and if A, B and C are disjoint subsets of nodes such that A is d-separated from B by C in the graph, then the distribution satisfies $A \perp\!\!\!\perp B \mid C$.
- Theorem: CI -> Factorization
  If a probability distribution satisfies the conditional independence statements implied by d-separation over a particular directed graph, then it also factorizes according to the graph.

Factorization

Has local, wants global

- CI statements are usually what is known by the expert
- The expert needs the model p(x) in order to compute things
- The CI -> Factorization part gives p(x) from what is known (CI statements)

Learning networks

- Network - joint probability distribution of random variables
- Structure can be learned, often set up "by hand"using expert knowldege
- Probabilites can be estimated from data using MAP/MLE

Probabilistic inference

- Evaluate probability of some set of variables given the values (observations) of another set.
- Exact inference of arbitrary Baysian network NP-hard
- Exact solutions for polytrees (no undirected cycles, exactly one undirected path between any two nodes)
- Approximate, numerical solutions, structure dependent

References

- Bayesian Networks without Tears. Eugene Charniak, AI magazine, vol 12, No4 , 1991 pp 50-63
- A Tutorial on Learning With Bayesian Networks, David Heckerman, Technical Report MSR-TR-95-06, 2006.
- Pattern Recognition and Machine Learning, Chapter 6, Christopher Bishop, Springer 2006