Probabilistic Graphical Models (PGM)

PA154 Jazykové modelování (8.1)

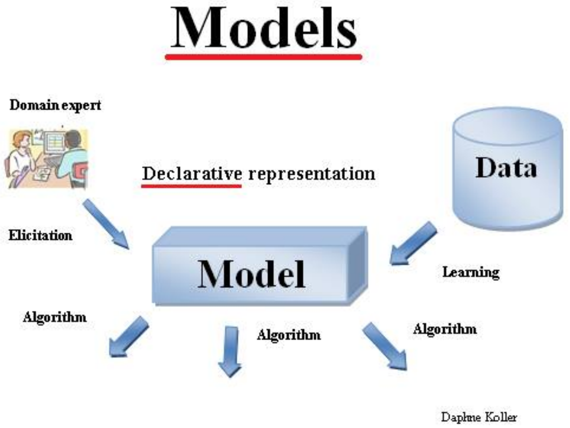
Pavel Rychlý

pary@fi.muni.cz

April 20, 2021

Source: Probabilistic Graphical Models
Daphne Koller
http://www.coursera.org/learn/probabilistic-graphical-models

Models

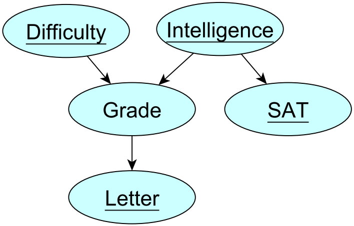


Models
Domain expert
Declarative representation
Data
Elicitation
Model
Learning
Algorithm
Algorithm
Algorithm
Daphne Koller

Graphical models

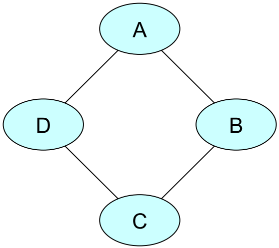
Bayesian networks

$X_1, \ldots, X_n$ - nodes
directed graph



Markov networks

undirected graph

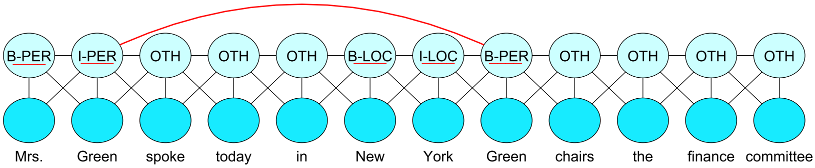


Daphne Koller

Textual Information Extraction

Mrs. Green spoke today in New York. Green chairs the finance committee.

Person — Location — Person — Organization



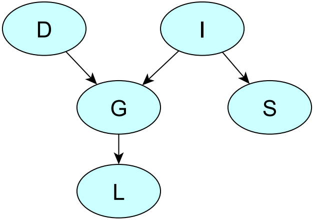
Daphne Koller

Graphical models

Bayesian networks

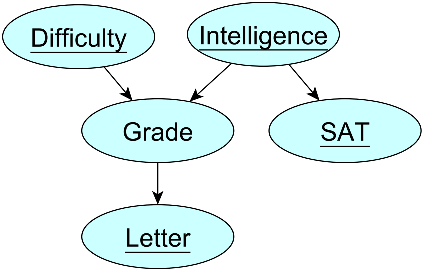
- **G**rade
- Course **D**ifficulty
- Student **I**ntelligence
- Student **S**AT
- Reference **L**etter

P(G,D,I,S,L)



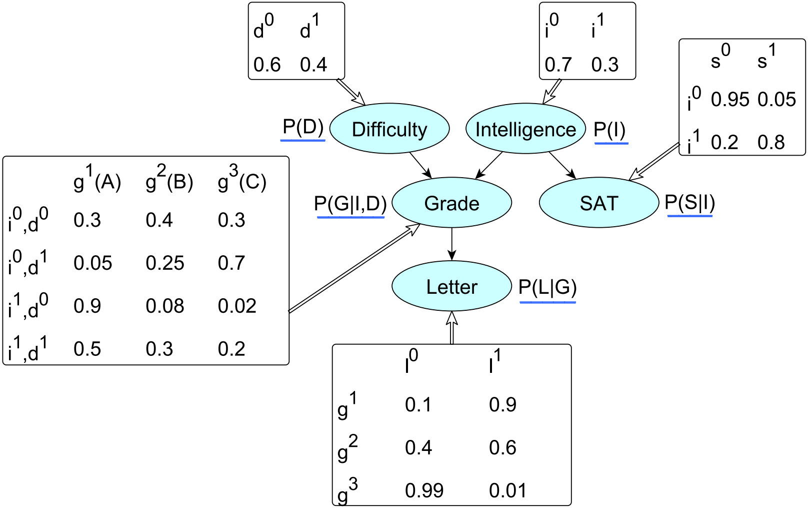
Daphne Koller

Graphical models



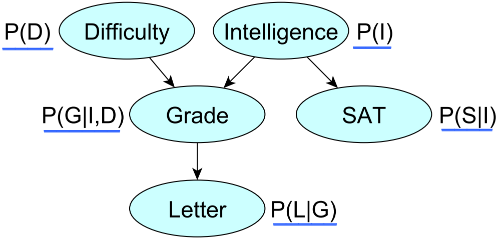
Daphne Koller

Graphical models



Daphne Koller

Chain Rule for Bayesian Networks

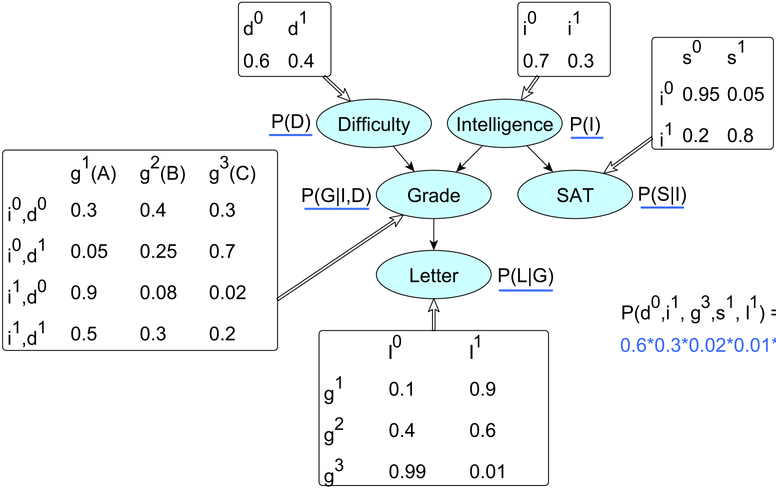


P(D,I,G,S,L) $= P(D)P(I)P(G|I,D)P(S|I)P(L|G)$

Distribution defined as a product of factors!

Daphne Koller

Chain Rule for Bayesian Networks



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	g^1(A)	g^2(B)	g^3(C)
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

$P(d^0, i^1, g^3, s^1, l^1) =$

$0.6*0.3*0.02*0.01*0.8$

Daphne Koller

Bayesian network

- A Bayesian network is:
  - A directed acyclic graph (DAG) G whose nodes represent random variables $X_1, \ldots, X_n$
  - For each node X_i a CPD $P(X_i | Par_G(X_i))$
- The BN represents a joint distribution via the chain rule for Bayesian networks

$$P(X_1, \ldots, X_n) = \prod_i P(X_i | Par_G(X_i))$$

Daphne Koller

BN Is a Legal Distribution: $P \geq 0$

- P is a product of CPDs
- CPDs are non-negative

Daphne Koller

BN Is a Legal Distribution: $\sum P = 1$

$$
\begin{aligned}
\sum_{D,I,G,S,L} P(D,I,G,S,L) &= \sum_{D,I,G,S,L} P(D)P(I)P(G|I,D)P(S|I)P(L|G) \\
&= \sum_{D,I,G,S} P(D)P(I)P(G|I,D)P(S|I)\sout{\sum_L P(L|G)} \\
&= \sum_{D,I,G,S} P(D)P(I)P(G|I,D)P(S|I) \\
&= \sum_{D,I,G} P(D)P(I)P(G|I,D)\sout{\sum_S P(S|I)} \\
&= \sum_{D,I} P(D)P(I)\sout{\sum_G P(G|I,D)}
\end{aligned}
$$

Daphne Koller

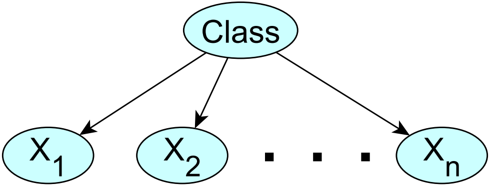
P Factorizes over G

- Let G be a graph over $X_1, \ldots, X_n$.
- P factorizes over G if

$$P(X_1, \ldots, X_n) = \prod_i P(X_i | Par_G(X_i))$$

Daphne Koller

Naïve Bayes Model

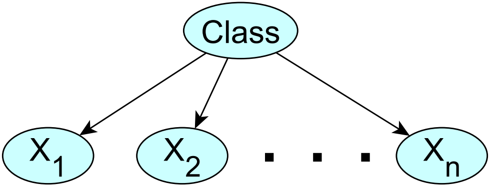


$$P(C, X_1, ..., X_n) = P(C) \prod_{i=1}^{n} P(X_i | C)$$

features X_i, X_j independent

Daphne Koller

Naïve Bayes Classifier

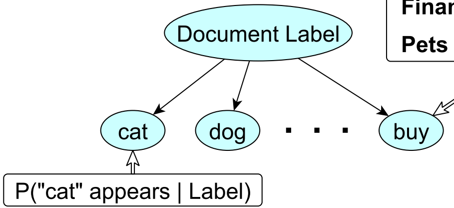


$$\frac{P(C=c^1|x_1,\ldots,x_n)}{P(C=c^2|x_1,\ldots,x_n)} = \frac{P(C=c^1)}{P(C=c^2)} \prod_{i=1}^{n} \frac{P(x_i|C=c^1)}{P(x_i|C=c^2)}$$

Daphne Koller

Bernoulli Naïve Bayes for Text

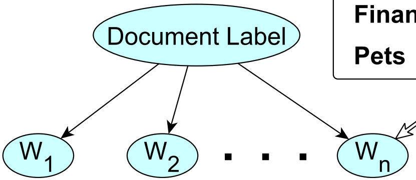
	cats	dogs	buy	sell	...
Financial	0.001	0.001	0.2	0.3	...
Pets	0.3	0.4	0.02	0.0001	...



$$\frac{P(C=c^1|x_1,\dots,x_n)}{P(C=c^2|x_1,\dots,x_n)} = \frac{P(C=c^1)}{P(C=c^2)} \prod_{i=1}^{n} \frac{P(x_i|C=c^1)}{P(x_i|C=c^2)}$$

Daphne Koller

Multinomial Naïve Bayes for Text



	cats	dogs	buy	sell	...
Financial	0.001	0.001	0.02	0.02	...
Pets	0.02	0.03	0.003	0.001	...

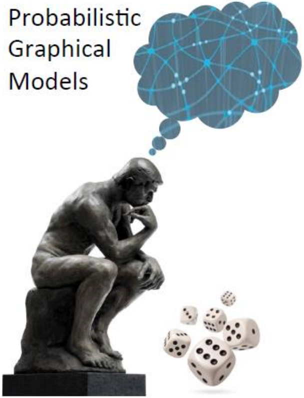
$$\frac{P(C=c^1|x_1,\ldots,x_n)}{P(C=c^2|x_1,\ldots,x_n)} = \frac{P(C=c^1)}{P(C=c^2)} \prod_{i=1}^{n} \frac{P(x_i|C=c^1)}{P(x_i|C=c^2)}$$

Daphne Koller

Summary

- Simple approach for classification
  - Computationally efficient
  - Easy to construct
- Surprisingly effective in domains with many weakly relevant features
- Strong independence assumptions reduce performance when many features are strongly correlated

Daphne Koller



Representation Bayesian Networks

Application: Diagnosis

Daphne Koller

Medical Diagnosis: Pathfinder (1992)

- Help pathologist diagnose lymph node pathologies (60 different diseases)
- Pathfinder I: Rule-based system
- Pathfinder II used naïve Bayes and got superior performance

Daphne Koller

Medical Diagnosis: Pathfinder (1992)

- Pathfinder III: Naïve Bayes with better knowledge engineering
- No incorrect zero probabilities
- Better calibration of conditional probabilities
  - $P(finding|disease_1)$ to $P(finding|disease_2)$
  - Not $P(finding_1|disease)$ to $P(finding_2|disease)$

Heckerman et al.

Daphne Koller

Medical Diagnosis: Pathfinder (1992)

- Pathfinder IV: Full Bayesian network
  - Removed incorrect independencies;
  - Additional parents led to more accurate estimation of probabilities
- BN model agreed with expert panel in 50/53 cases, vs 47/53 for naïve Bayes model
- Accuracy as high as expert that designed the model

Heckerman et al.

Daphne Koller