Introduction to Probabilistic Graphical Models
Lecture 2

Directed Graphical Models, Conditional Independence

Télécom ParisTech,

Université Paris-Saclay, Paris, France

Instructor: Umut Şimşekli

Lecture Outline

- Graphical Models
  - Bayesian Networks

Disclaimer

- All the material that will be used within this course is adapted from the "Bayesian Statistics and Machine Learning" course that has been given by A. Taylan Cemgil at Boğaziçi University, Istanbul

- For more info, please see `http://www.cmpe.boun.edu.tr/~cemgil/`

Graphical Models

- formal languages for specification of probability models and associated inference algorithms
- historically, introduced in probabilistic expert systems (Pearl 1988) as a visual guide for representing expert knowledge
- today, a standard tool in machine learning, statistics and signal processing

Graphical Models

- provide graph based algorithms for derivations and computation
- pedagogical insight/motivation for model/algorithm construction
  - Statistics:
    "Kalman filter models and hidden Markov models (HMM) are equivalent upto parametrisation"
  - Signal processing:
    "Fast Fourier transform is an instance of sum-product algorithm on a factor graph"
  - Computer Science:
    "Backtracking in Prolog is equivalent to inference in Bayesian networks with deterministic tables"
- Automated tools for code generation start to emerge, making the design/implement/test cycle shorter

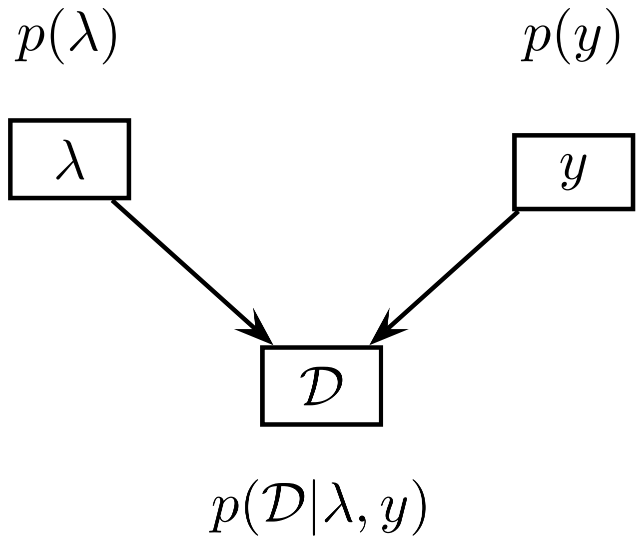
Important types of Graphical Models

- Useful for Model Construction
  - **Directed Acyclic Graphs (DAG), Bayesian Networks**
  - **Undirected Graphs, Markov Networks, Random Fields**
  - Influence diagrams
  - ...

- Useful for Inference
  - **Factor Graphs**
  - Junction/Clique graphs
  - Region graphs
  - ...

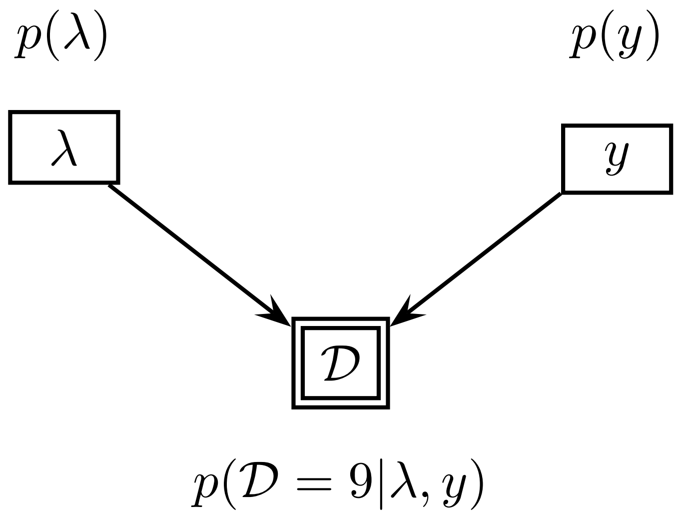
Directed Graphical models (DAG)

DAG Example: Two dice



$$p(\mathcal{D}, \lambda, y) \quad = \quad p(\mathcal{D}|\lambda, y)p(\lambda)p(y)$$

DAG with observations



$$\phi_{\mathcal{D}}(\lambda, y) \quad = \quad p(\mathcal{D} = 9|\lambda, y)p(\lambda)p(y)$$

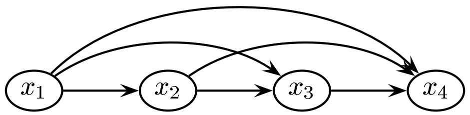
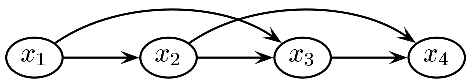
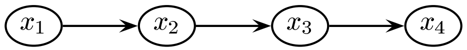
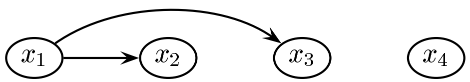
Directed Graphical models

- Each random variable is associated with a node in the graph,
- We draw an arrow from $A \to B$ if $p(B|\dots, A, \dots)$ ($A \in \text{parent}(B)$),
- The edges tell us *qualitatively* about the factorization of the joint probability
- For N random variables $x_1, \dots, x_N$, the distribution admits

$$p(x_1, \dots, x_N) = \prod_{i=1}^{N} p(x_i|\text{parent}(x_i))$$

- Describes in a compact way an algorithm to “generate” the data – “Generative models”

Examples

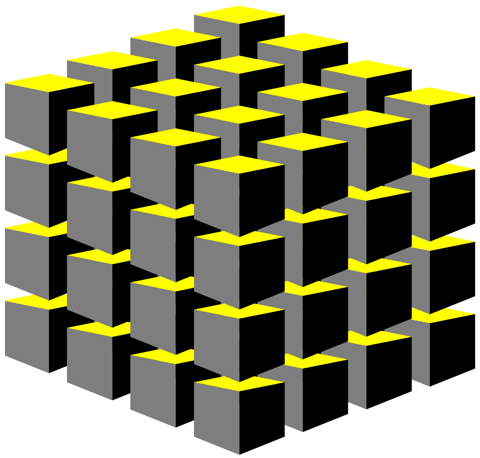
Model	Structure	factorization
Full		$p(x_1)p(x_2\|x_1)p(x_3\|x_1,x_2)p(x_4\|x_1,x_2,x_3)$
Markov(2)		$p(x_1)p(x_2\|x_1)p(x_3\|x_1,x_2)p(x_4\|x_2,x_3)$
Markov(1)		$p(x_1)p(x_2\|x_1)p(x_3\|x_2)p(x_4\|x_3)$
		$p(x_1)p(x_2\|x_1)p(x_3\|x_1)p(x_4)$
Factorized		$p(x_1)p(x_2)p(x_3)p(x_4)$

Removing edges eliminates a term from the conditional probability factors.

Probability Tables

- Assume all x_i are discrete with $|x_i| = k$. If N is large, a naive table representation is HUGE: k^N entries

Example: $p(x_1, x_2, x_3)$ with $|x_i| = 4$

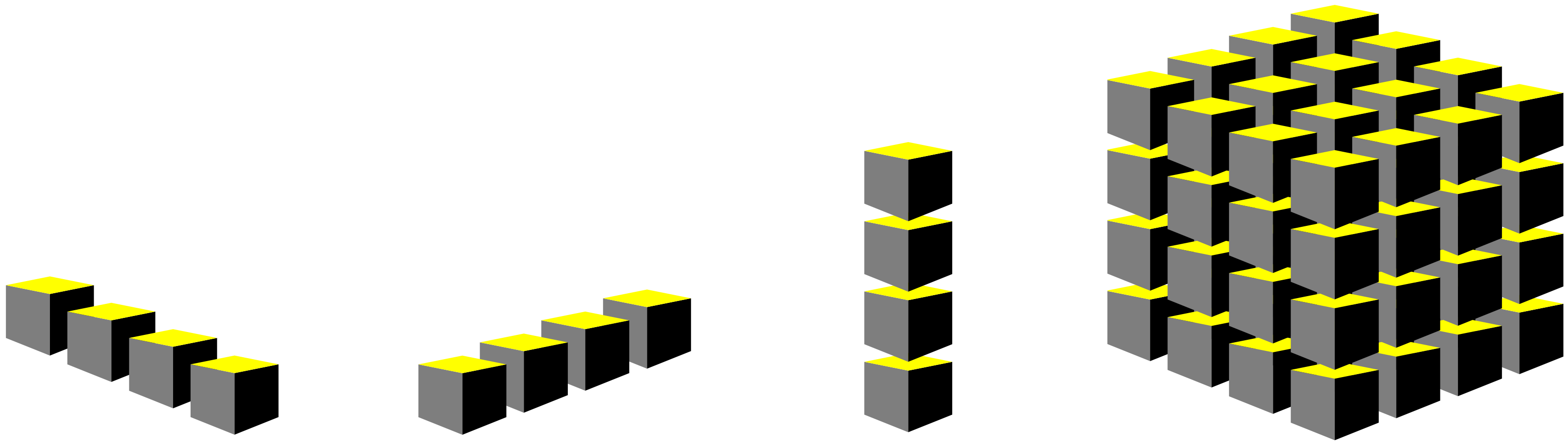


Each cell is a positive number s.t. $\sum_{x_1,x_2,x_3} p(x_1, x_2, x_3) = 1$

- We need efficient data structures to represent joint distributions $p(x_1, x_2, \ldots, x_N)$

Independence Assumption == Complete Factorization

- Assume $p(x_1, x_2, \ldots, x_N) = \prod_k p(x_k)$.

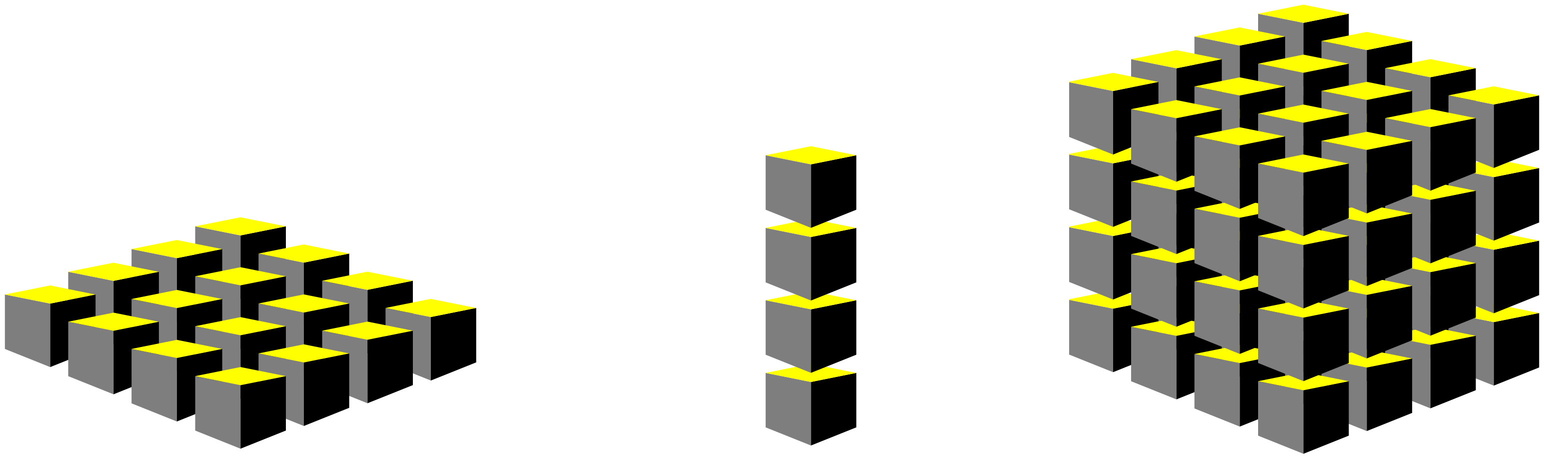


$$p(x_1) \quad \times \quad p(x_2) \quad \times \quad p(x_3) \quad = \quad p(x_1, x_2, x_3)$$

We need to store 4×3 numbers instead of 4^3 !

- However, complete independence is too restrictive and not very useful.

An alternative Factorization



$$p(x_1, x_2) \quad \times \quad p(x_3) \quad = \quad p(x_1, x_2, x_3)$$

We need to store $4^2 + 4$ numbers instead of 4^3.

- Still some variables are independent from rest. We will make conditional independence assumptions instead.

Conditional Independence

- Two disjoint sets of variables A and B are conditionally independent given a third disjoint set C if

$$p(A, B|C) \quad = \quad p(A|C)p(B|C)$$

- This is equivalent to

$$p(A|BC) \quad = \quad p(A|C)$$

- We denote this relationship with ($\perp\!\!\!\perp$)

$$A \perp\!\!\!\perp B|C$$

Conditional Independence

- Conditional Independence is a key concept in probabilistic models
- Conceptual and Computational simplifications
  - Understanding key factors in a domain
  - Reducing computational burden for inference

Conditional Independence Properties

- Directed Graphical Models
  - d-separation
- Markov Random Fields (MRF's : Undirected Graphical Models)
  - Path Blocking
- Testing for conditional independence in MRF is simpler

Keywords Summary

Bayes Theorem

Likelihood

Prior

Posterior

Evidence, Marginal Likelihood

Bayesian Network, Directed Graphical Model, parents

Undirected Graphical Model, Markov Blanket

Factor Graph, Factor node, variable node, edges, bipartite graph

Probability Tables, Conditional Probability

Probability Distributions, Density function, stable computation

Factorisation, Proportional-to notation

Marginalisation, Integrating-out, Clamping, Conditioning, Max-marginal

Clique, Clique potential, local compatibility functions

Model topology, Model Selection