Introduction to Probabilistic Graphical Models
Lecture 3

Undirected Graphs, Factor Graphs, Inference

Télécom ParisTech,

Université Paris-Saclay, Paris, France

Instructor: Umut Şimşekli

2016-2017

Updated Grading

- New Grading Schema!
  - % 50 Homeworks (7+7+12+12+12)
  - % 50 Final
  - % 10 Attendance (bonus)
  - % No Quizzes

Lecture Outline

- Graphical Models
  - Bayesian Networks
  - Undirected Graphical models, Markov Random Fields
  - Factor graphs
- Learning and Inference

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/`

d-Separation

- Three disjoint sets of variables A, B and C

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

- A path from A to B is blocked by C if

  a the arrows on the path meet either head-to-tail or tail-to-tail at the node, and the node is in the set C, or

  b the arrows meet head-to-head at the node, and neither the node, nor any of its descendants, is in the set C.

Markov Blanket For DAG's

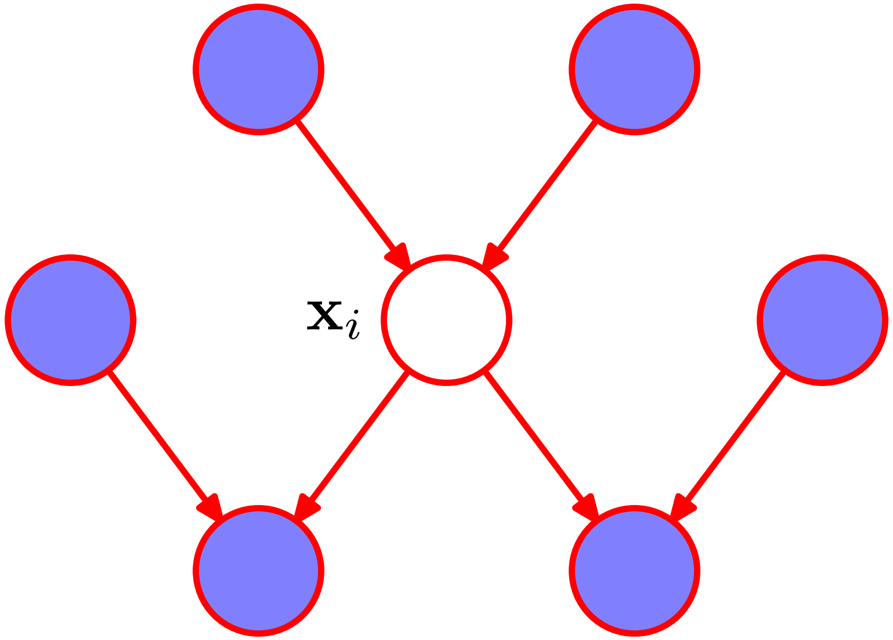


Figure 8.26 from Bishop PRML

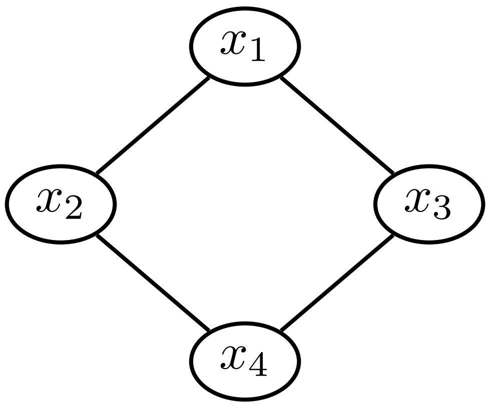
Undirected Graphical Models

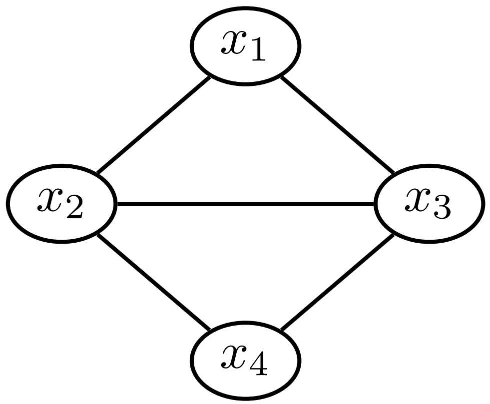
- Define a distribution by non-negative *local compatibility functions* $\phi(x_\alpha)$

$$p(\mathbf{x}) = \frac{1}{Z} \prod_\alpha \phi(x_\alpha)$$

where α runs over **cliques** : fully connected subsets

- Examples

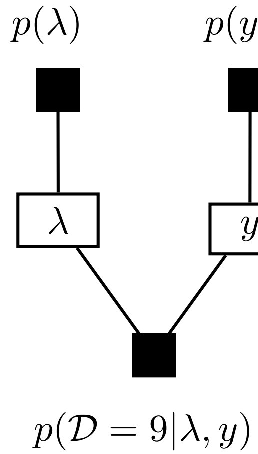




$p(\mathbf{x}) = \frac{1}{Z}\phi(x_1, x_2)\phi(x_1, x_3)\phi(x_2, x_4)\phi(x_3, x_4)$ $\qquad$ $p(\mathbf{x}) = \frac{1}{Z}\phi(x_1, x_2, x_3)\phi(x_2, x_3, x_4)$

Factor graphs

- A bipartite graph. A powerful graphical representation of the inference problem
  - **Factor nodes**: Black squares. Factor potentials (local functions) defining the posterior.
  - **Variable nodes**: White Nodes. Define collections of random variables
  - **Edges**: denote membership. A variable node is connected to a factor node if a member variable is an argument of the local function.

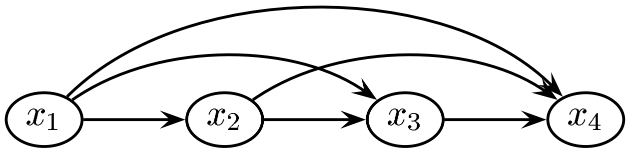


$$\phi_{\mathcal{D}}(\lambda, y) \quad = \quad p(\mathcal{D} = 9|\lambda, y)p(\lambda)p(y) = \phi_1(\lambda, y)\phi_2(\lambda)\phi_3(y)$$

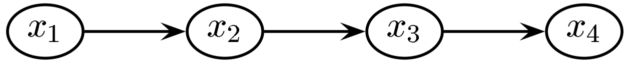
Exercise

- For the following Graphical models, write down the factors of the joint distribution and plot an equivalent factor graph and an undirected graph.

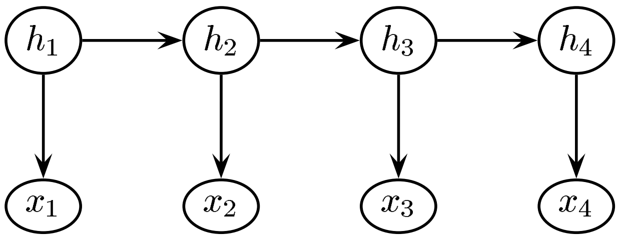
Full



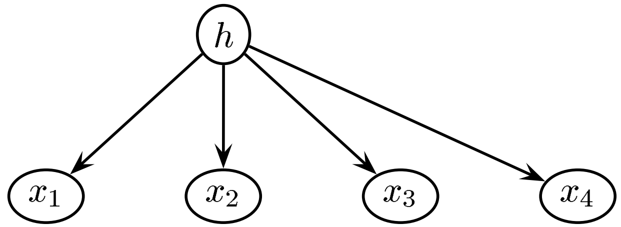
Markov(1)



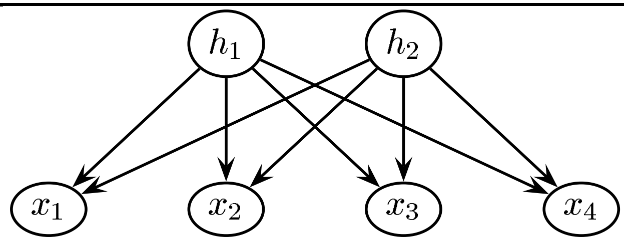
HMM



MIX

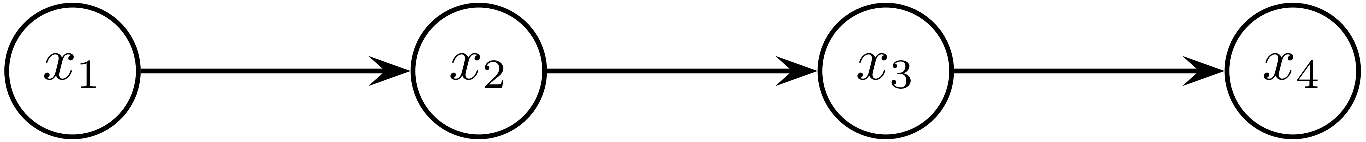


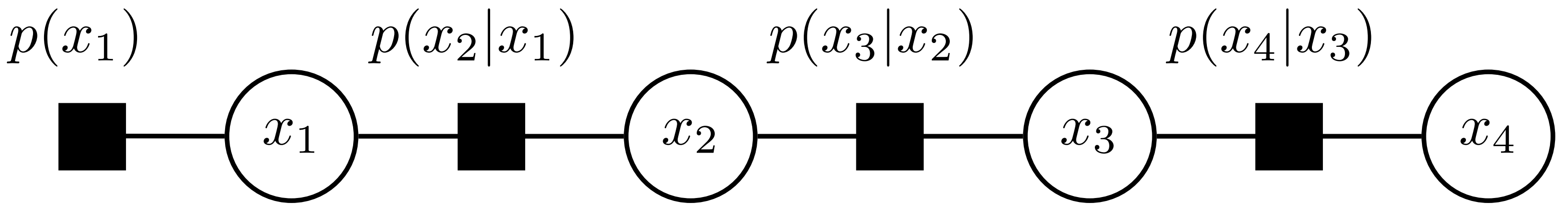
IFA

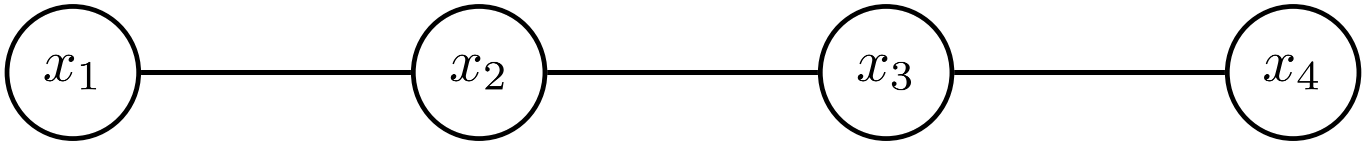


Factorized

Answer (Markov(1))

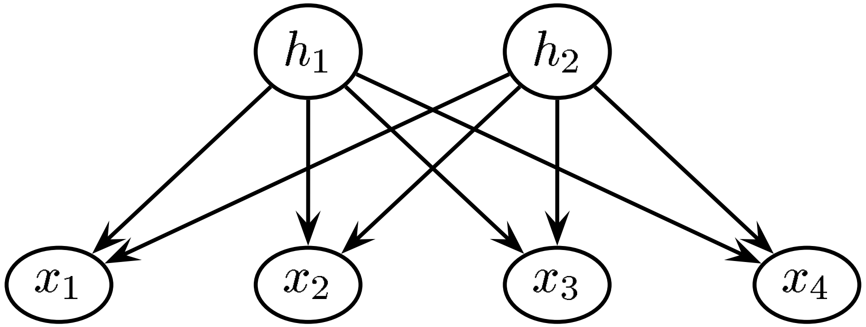




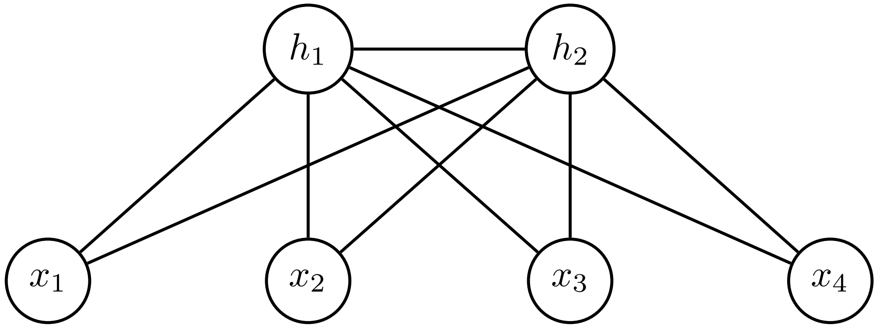


$$\underbrace{p(x_1)p(x_2|x_1)}_{\phi(x_1,x_2)}\underbrace{p(x_3|x_2)}_{\phi(x_2,x_3)}\underbrace{p(x_4|x_3)}_{\phi(x_3,x_4)}$$

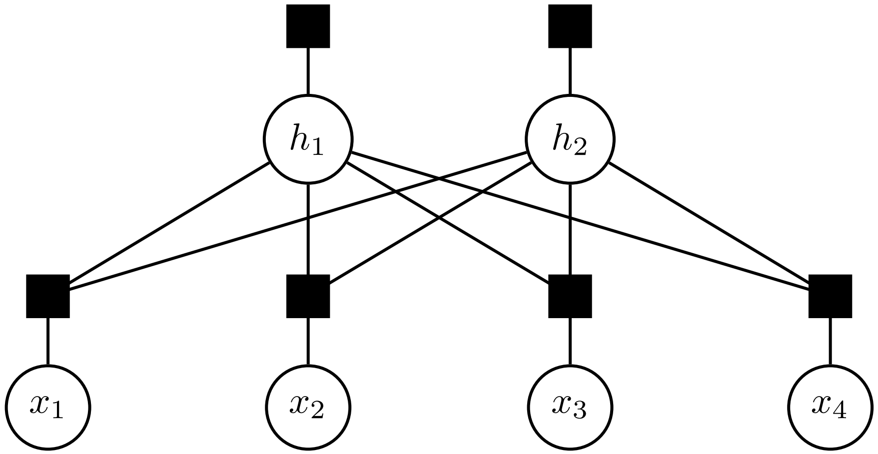
Answer (IFA – Factorial)



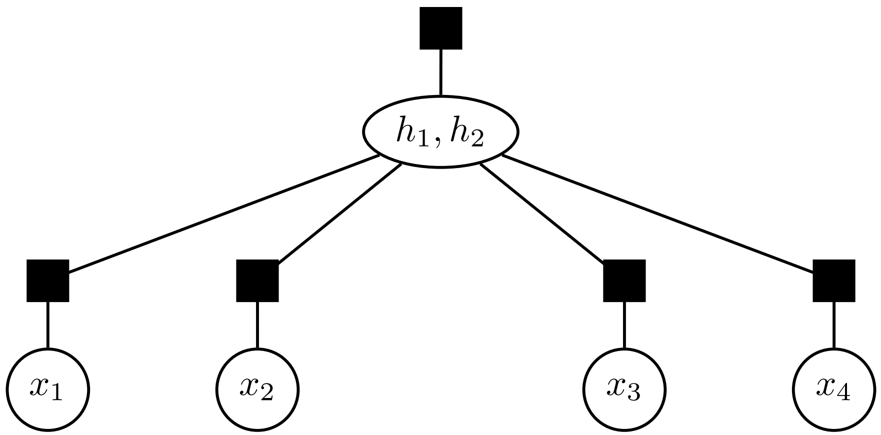
$$p(h_1)p(h_2)\prod_{i=1}^{4} p(x_i|h_1,h_2)$$



Answer (IFA – Factorial)



- We can also cluster nodes together



Inference and Learning

- Data set

$$\mathcal{D} = \{x_1, \ldots x_N\}$$

- Model with parameter λ

$$p(\mathcal{D}|\lambda)$$

- Maximum Likelihood (ML)

$$\lambda^{\mathsf{ML}} = \arg\max_{\lambda} \log p(\mathcal{D}|\lambda)$$

- Predictive distribution

$$p(x_{N+1}|\mathcal{D}) \approx p(x_{N+1}|\lambda^{\mathsf{ML}})$$

Regularisation

- Prior

$$p(\lambda)$$

- Maximum a-posteriori (MAP) : Regularised Maximum Likelihood

$$\lambda^{\text{MAP}} = \arg\max_{\lambda} \log p(\mathcal{D}|\lambda)p(\lambda)$$

- Predictive distribution

$$p(x_{N+1}|\mathcal{D}) \approx p(x_{N+1}|\lambda^{\text{MAP}})$$

Keywords Summary

Undirected Graphical Model, Markov Blanket

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

Probability Tables, Conditional Probability

Clique, Clique potential

Maximum Likelihood (ML), Maximum a-posteriori (MAP)