Graphical Models

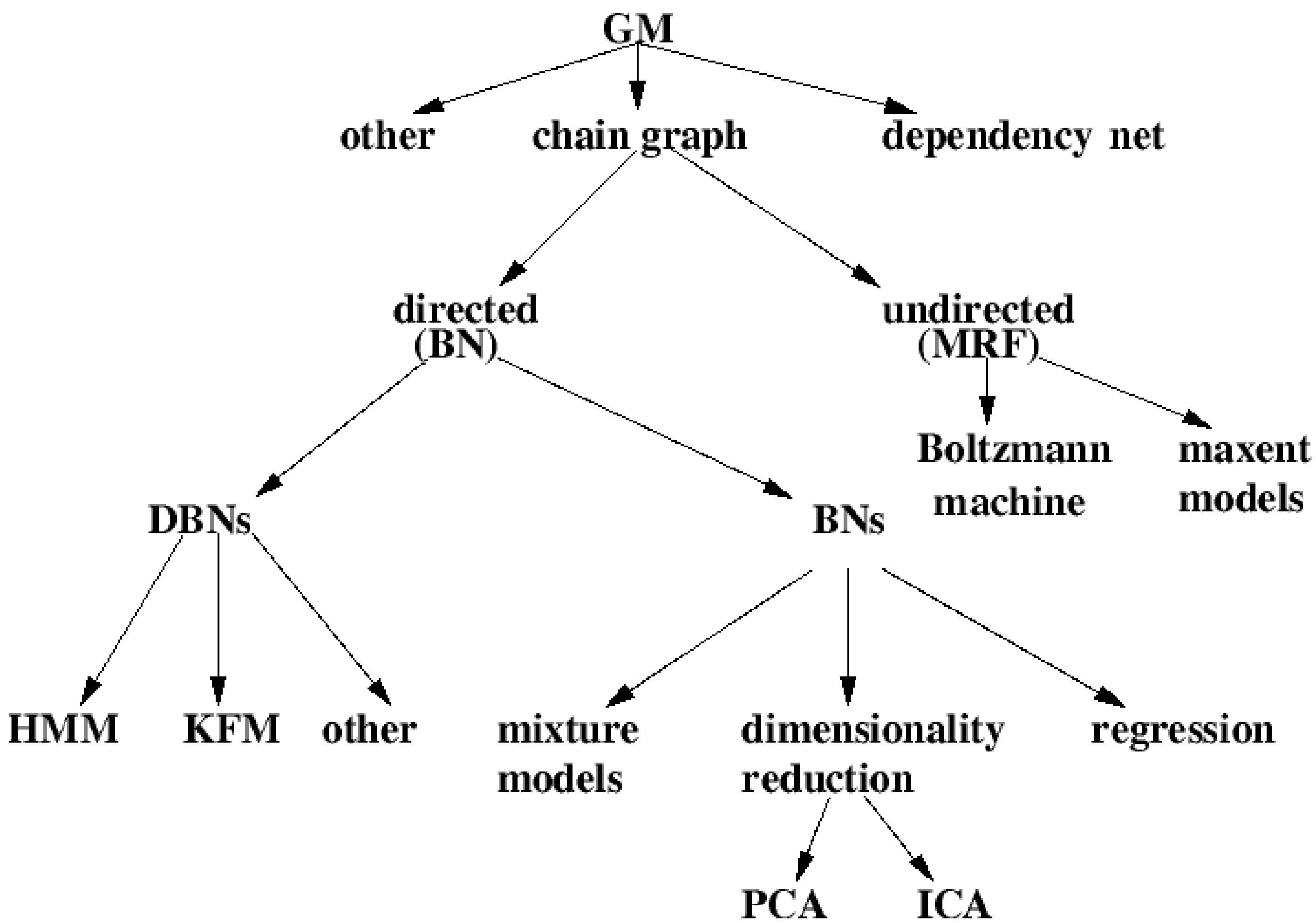
Franz Pernkopf

Laboratory of Signal Processing and Speech Communication
Graz University of Technology
Inffeldgasse 16c , 8010 Graz, Austria
Tel: +43 316 873 4436
E-Mail: pernkopf@tugraz.at

Course Topics

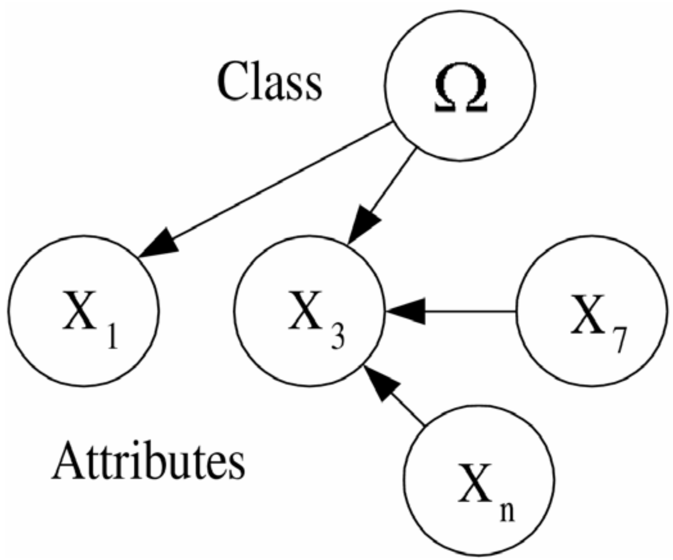
Topics	Speaker	Date
Introduction to GMs	Franz Pernkopf	11-04-05
Parameter and Structure Learning	Stefan Tertinek	
Exact Inference	Rene Hirschmanner	
Generalized Belief Propagation	available	
Loopy Belief Propagation	available	
Variational Inference	available	
Sampling	available	
Linear Gaussian Models / HMM	available (Tuan ?)	
Particle Filter	Dimitri Shutin	
Factor Graphs	Thomas Blocher	
Bayesian Network Classifiers	Cornelia Falch, Markus Noisternig	
GM Tools with application to String Edit Distance	Stefan Petrik	

Taxonomy of GM



- A Bayesian Network $B = \langle G, \Theta \rangle$ consists of

1. a network structure G
   - Nodes: Random variables $\Rightarrow$ $U = \{\Omega, X_1, \ldots, X_n\}$
   - Arcs (edges): Dependency between random variables
   - directed acyclic graph (DAG)
2. Symbol Θ represents a set of parameters which quantify the network
   - a set of (conditional) probability distributions $\Rightarrow$ $P = \{P(U_i \mid \mathrm{pa}(U_i))\}$



Directed GM (Bayesian Network)

- Conditional independence: Each variable is independent of its non-descendants in the network given its parents.

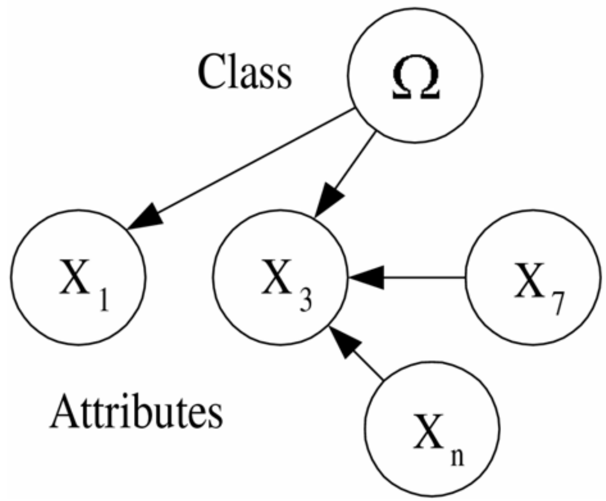
$$U \perp \mathrm{nd}(U) \mid \mathrm{pa}(U)$$

$$X_1 \perp X_3 \mid \Omega : \qquad P(X_1 = x_1 \mid \Omega = \omega, X_3 = x_3) = P(X_1 = x_1 \mid \Omega = \omega)$$

- Joint probability distribution:

$$P(U) = P(U_1,\ldots,U_n) = \prod_{i=1}^{n+1} P(U_i \mid \mathrm{pa}(U_i)) = P(\Omega, X_1,\ldots, X_n) =$$

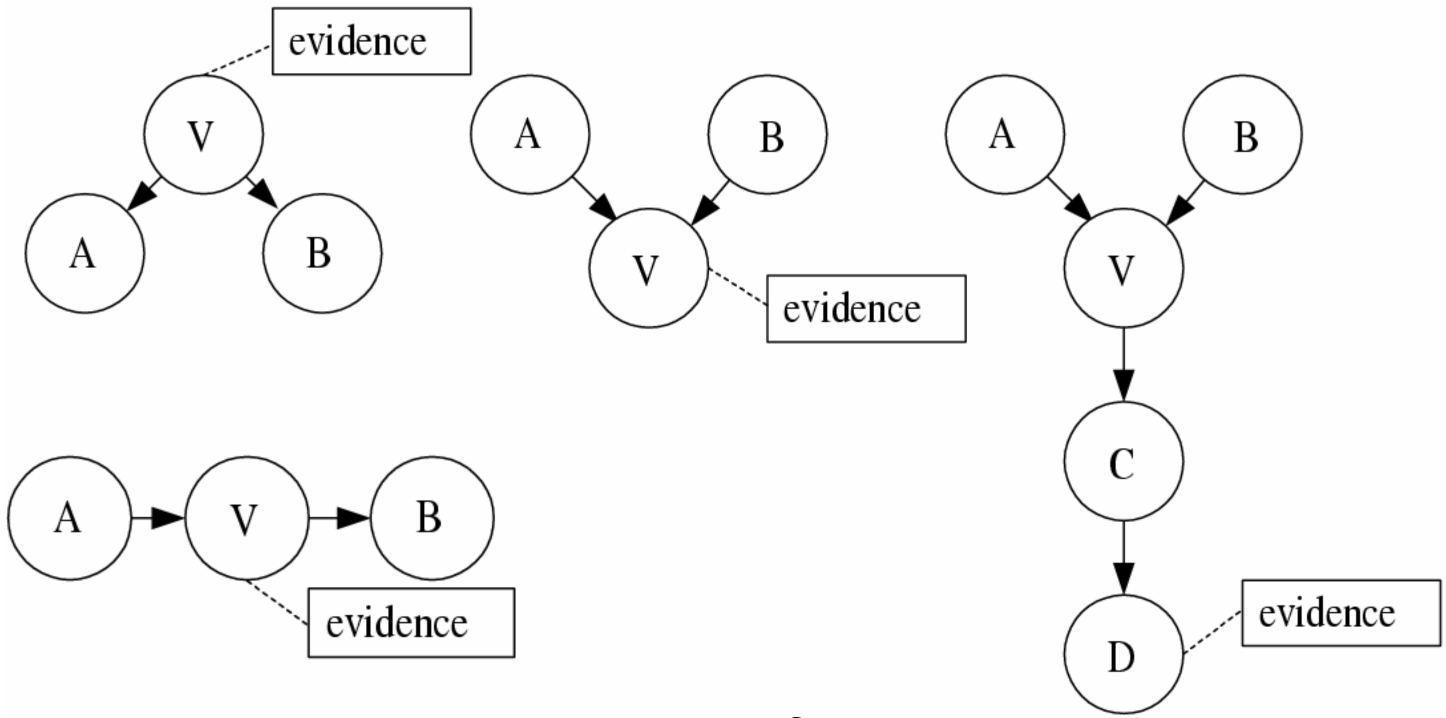
$$P(\Omega)P(X_7)P(X_n)P(X_1 \mid \Omega)P(X_3 \mid \Omega, X_7, X_n)$$



- A Bayesian network encodes the joint probability distribution over a set of variables.

Directed GM (Bayesian Network)

- Dependence relationships in DAGs are modeled by a separability criterion called d-separation.
- **d-separation:** Two variables *A* and *B* are d-separated if for all paths between *A* and *B* there is an intermediate variable *V* such that either
  - The connection is serial or diverging and the state of *V* is known.
  - The connection is converging and neither *V* nor any of *V*s descendants have received evidence.



Inference

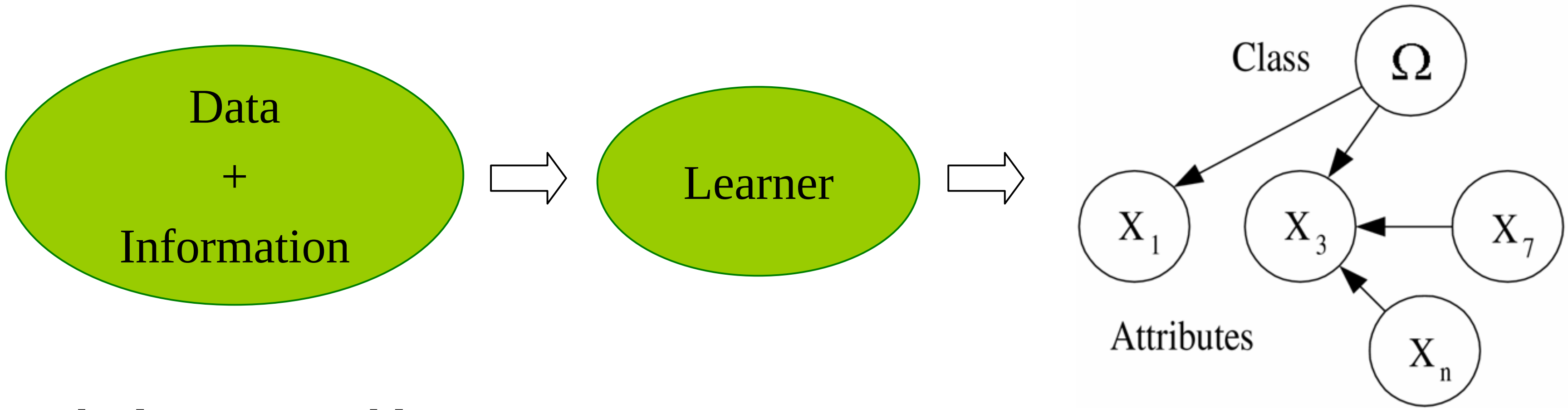
- Probabilistic inference means computing

$$P(U_Q \mid U_E = u_E)$$

- "Evidence " means observing the values of certain nodes
- Absorbing evidence divides the units of the network into two groups:
  - **Evidence variables** U_E: Set of evidence variables (value is known)
  - **Query variables** U_Q: Set of query variables

- Exact Inference
  - Junction Tree Algorithm (Message Passing)
- Approximate Inference
  - Variational Inference
  - Loopy Belief Propagation
  - Stochastic Sampling Methods (MCMC)

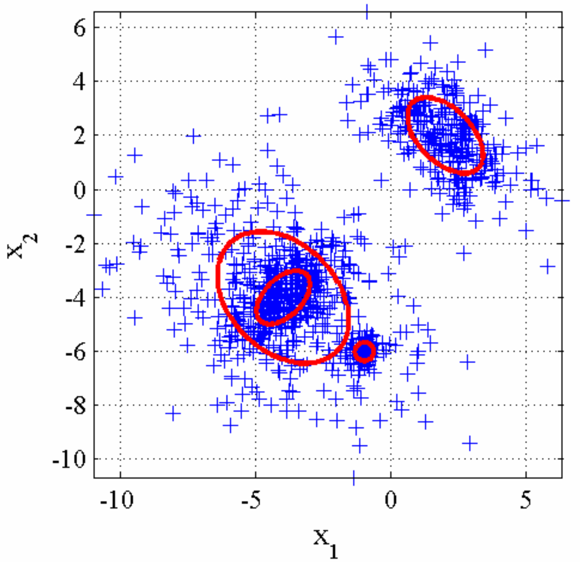
Learning Bayesian Networks



- The learning problem:

	Known Structure	**Unknown Structure**
Complete Data	Parameter estimation: Maximum likelihood (ML), Maximum a-posterior (MAP), …	Search model space (HCS, GA,…) Criterion: Classification Rate (CR), Conditional Mutual Information (CMI),…
Incomplete Data	Parameter estimation: Inference (EM),…	Structural EM

- A Gaussian Mixture Model (GMM) is a distribution composed of weighted Gaussian distributions.



- Problem of EM for learning GMM:
  - Solution depends on initial parameters
  - EM algorithm assumes that the number of Gaussian components is known
- Combination of EM and GA (GA-EM)
  - "Solves" the initialization issue
  - Employ the EM algorithm within the GA framework to speed up the optimization
  - Minimum Description Length (MDL) for determination of the number of components