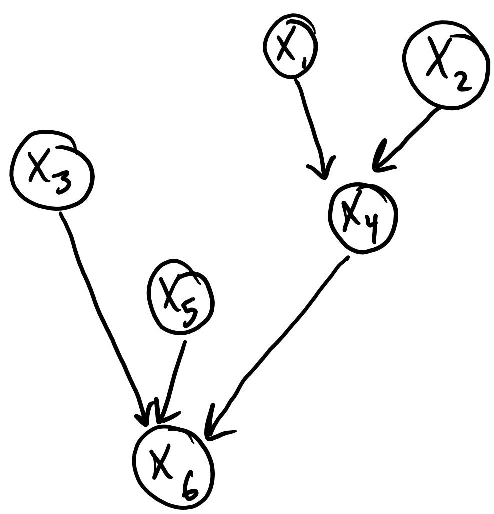
Graphical Models

Tuesday, June 6, 2017 9:59 AM

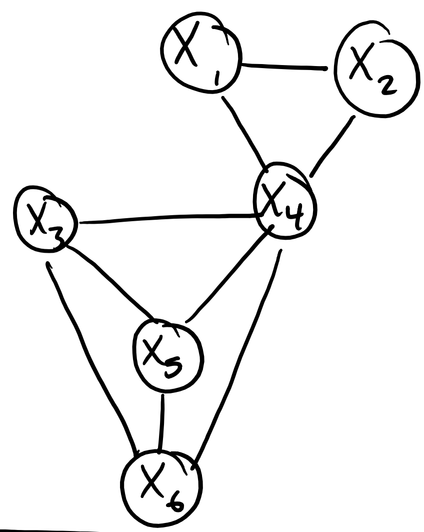
Koller & Friedman "Probabilistic Graphical Models"

Wainwright & Jordan "Graphical Models, Exponential Families, and Variational Inference"

Idea: Use graphs to describe "qualitatively" dependencies between random variables.



Directed



Undirected

How does info about parent variables propogate to children variables?

How do variables depend on one another?

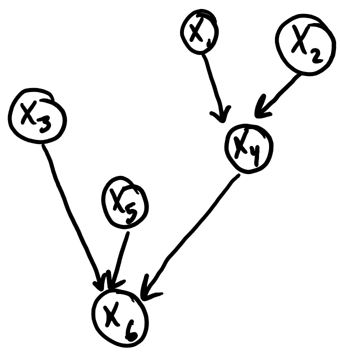
Why are we doing this?

Representation: understand models & roles of variables

Inference: Algorithms for computing conditional distributions given evidence

Learning: Learn the graphical structure, joint distrⁿ

<u>Directed</u>:

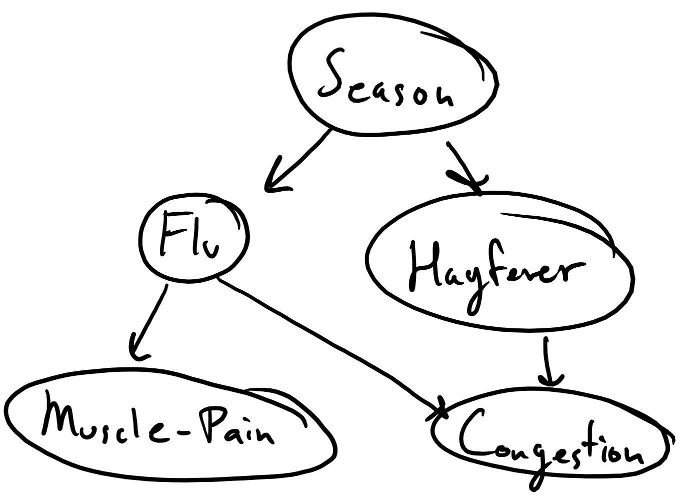


$X_1 \sim f_{X_1}$ $\qquad$ $X_4 \mid X_1, X_2 \sim f_{X_4 \mid X_1, X_2}$

$X_2 \sim f_{X_2}$ $\qquad$ $X_6 \mid X_3, X_4, X_5$

$X_3 \sim f_{X_3}$ $\qquad$ $\sim f_{X_6 \mid X_3, X_4, X_5}$

$X_5 \sim f_{X_5}$



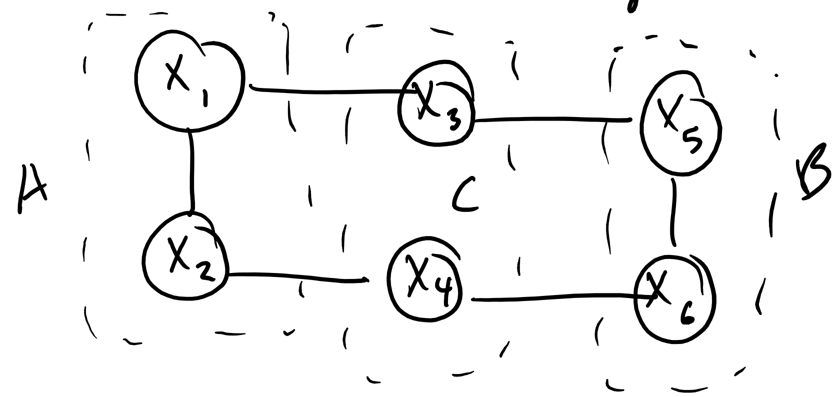
Undirected Graphical Models

Marginal Correlation Graphs: No edge if $X_j \perp\!\!\!\perp X_k$

Conditional Independence Graphs:

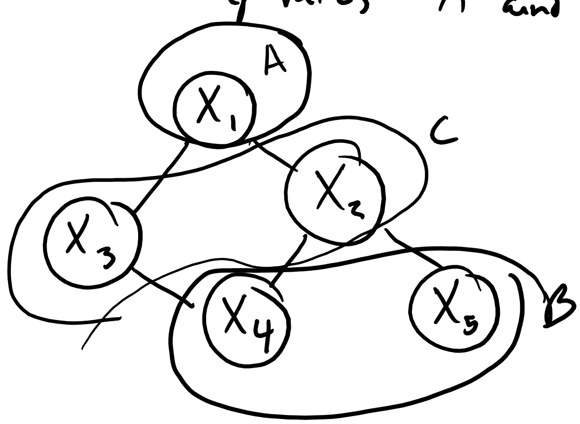
No edge if $X_j \perp\!\!\!\perp X_k \mid \text{rest}$ (pairwise Markov property)

<u>def</u> A group of variables C <u>separates</u> A, B if every path from A to B passes through C.



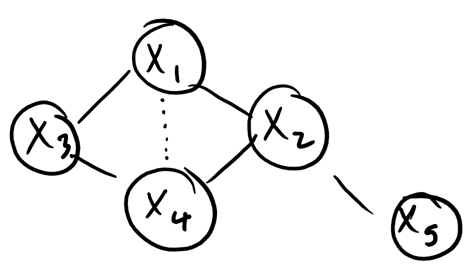
Local Markov Property: X_i conditional on it neighbors is independent of rest of RVs.

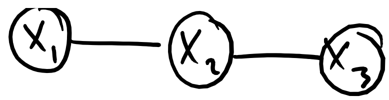
Global " " : For every disjoint A, B, C s.t. C separates A and B $X_A \perp\!\!\!\perp X_B \mid X_C$



Global $\Rightarrow$ Local $\Rightarrow$ Pairwise

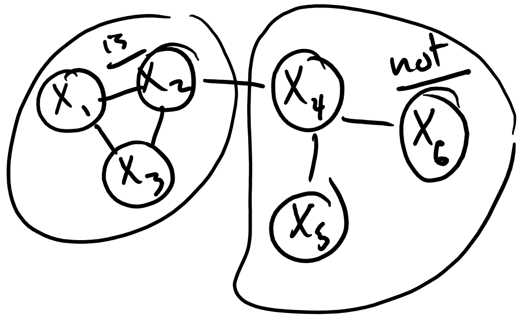
$X_1, X_4 \mid X_3, X_2, X_5$



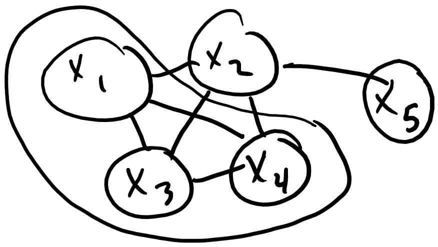


$$f_{X_1 X_2 X_3} = f_{X_1} \cdot f_{X_2} \cdot f_{X_3}$$

def A clique is a complete subgraph



A <u>maximal</u> clique is not contained w/in a larger clique.

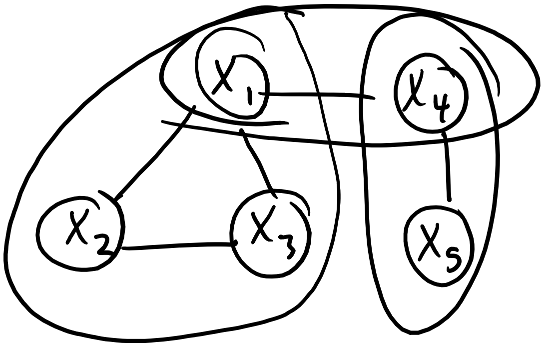


def The density factorizes wrt. a set of cliques $\mathcal{C}$ if $f_X = \frac{1}{Z} \prod_{c \in \mathcal{C}} \psi_c(x_c)$

thm For positive dist$^{\underline{n}}$s

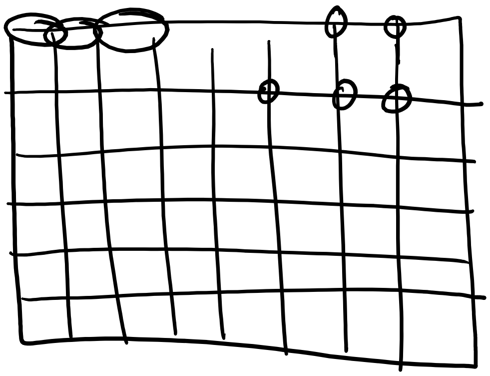
global Markov $\Leftrightarrow$ local Markov

$\Leftrightarrow$ factorizes wrt. maximal cliques



$$f_X(x) = \frac{1}{Z} \exp\left\{ \sum_{c \in \mathcal{C}} \log \psi_c(x_c) \right\}$$

Gibbs <u>representation</u>



grid graph

→ Cliques are edges

$$f_X(x) = \frac{1}{Z} \exp\left\{ \sum_{e \in E} \log \psi_e(x_e) \right\}$$

Gaussian Graphical Model

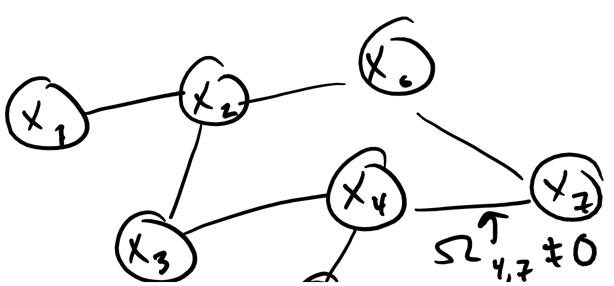
$$f_X(x) = \frac{1}{(2\pi)^{T/2}\sqrt{\det(\Sigma)}} e^{-\frac{1}{2}x^T \Sigma^{-1} x} \qquad (\Sigma^{-1} \text{ exists})$$

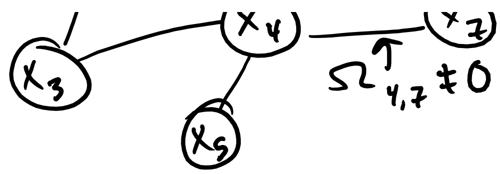
$\Sigma^{-1} =: \Omega$ (Precision matrix)

$$e^{-\frac{1}{2}x^T \Omega x} = e^{-\frac{1}{2}\sum_{i,j} \Omega_{ij} x_i x_j}$$

$$= \prod_{i,j : \Omega_{ij} \neq 0} e^{-\frac{1}{2}\Omega_{ij} x_i x_j}$$

Graph w/ edge set $\{(i,j) : \Omega_{ij} \neq 0\}$
factorizes for the GGM w/ precision Ω





Suppose we have copies (iid)

X_{ij} - i^{th} replicate of j^{th} variable

Estimate $\hat{\Sigma} = \frac{1}{n} \sum_i x_i x_i^T$

(1) $\hat{\Omega} = \hat{\Sigma}^{-1}$ is natural

(2) threshold components of $\hat{\Omega}$

Graphical Lasso

$$-\ln(L) \propto -\ln|\Omega| + x^T \Omega x$$

$$R_n = \frac{1}{n} \sum_i \left(-\ln|\Omega| + x_i^T \Omega x_i\right)$$

$$\min_{\Omega} \quad R_n + \lambda |||\Omega|||_1$$

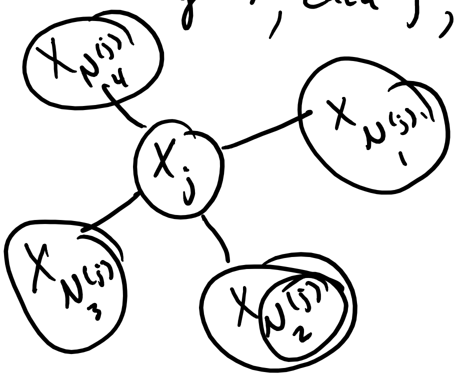
$$\text{w/} \quad |||\Omega|||_1 = \sum_{j \neq k} |\Omega_{jk}|$$

▷ in general is called semi-definite prog

Alternative is through the Lasso

(1) Regress, each j, X_j onto rest X_{-j}

(1) Regress, each j, X_j onto rest X_{-j}



$$\min_{\beta} \frac{1}{n}\sum_{i=1}^{n} \left(x_{ij} - \beta^T x_{i,-j}\right)^2 + \lambda \|\beta\|_1$$

$\beta_{N^{(j)}_k} \neq 0$ "learns the neighborhood"

(2) Combine neighborhoods using "and" or "or"

Binary $X_j \in \{-1, 1\}$ then implies Gibbs rep, for edge factorized structure.

$$f_X(x) = \frac{1}{Z} \exp\left\{-\left(\sum_{i,j} \omega_{ij} x_i x_j + \sum_i a_i x_i\right)\right\}$$

is called the Ising model