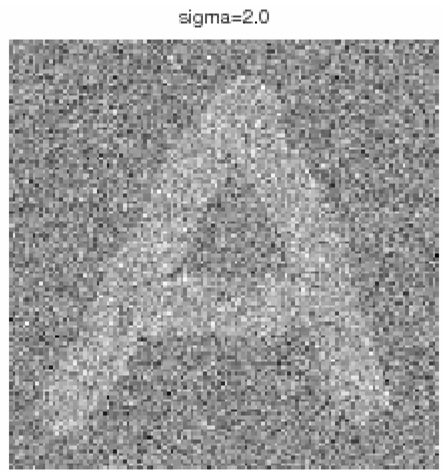
CS340 Machine learning
Gibbs sampling in Markov random fields

Image denoising

$\mathbf{x}$

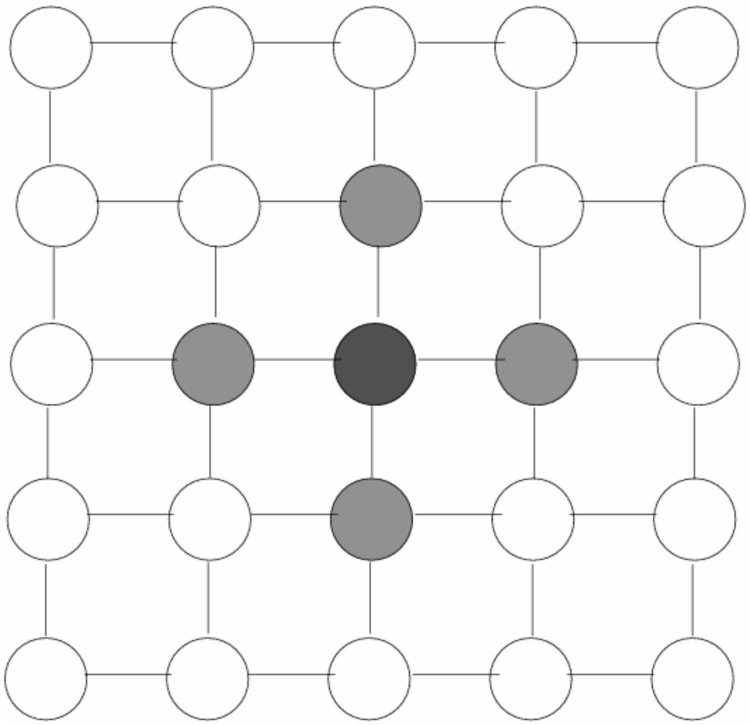


$\mathbf{y}$

$\hat{\mathbf{x}} = E[\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}]$

Ising model

- 2D Grid on {-1,+1} variables
- Neighboring variables are correlated



$$p(\mathbf{x}|\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} \prod_{<ij>} \psi_{ij}(x_i, x_j)$$

Ising model

$$\psi_{ij}(x_i, x_j) = \begin{pmatrix} e^{W} & e^{-W} \\ e^{-W} & e^{W} \end{pmatrix}$$

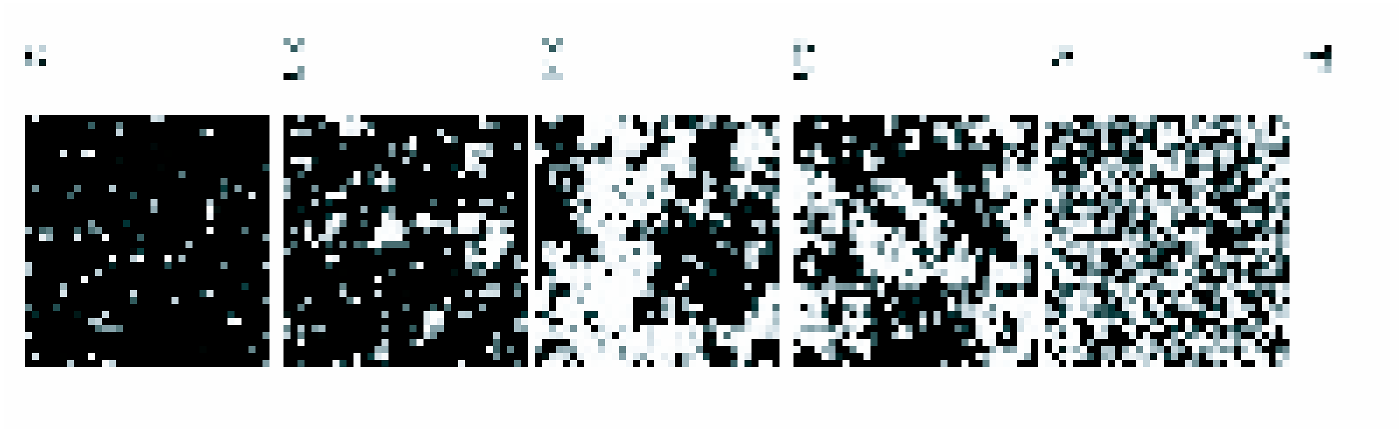
W > 0: ferro magnet
W < 0: anti ferro magnet (frustrated system)

$$\begin{aligned} p(\mathbf{x}|\boldsymbol{\theta}) &= \frac{1}{Z(\boldsymbol{\theta})} \exp[-\beta H(\mathbf{x}|\boldsymbol{\theta})] \\ H(\mathbf{x}) &= -\mathbf{x}^T \mathbf{W} \mathbf{x} = -\sum_{<ij>} W_{ij} x_i x_j \end{aligned}$$

β=1/T = inverse temperature

Samples from an Ising model

W=1



Cold T=2

Hot T=5

Mackay fig 31.2

Boltzmann distribution

- Prob distribution in terms of clique potentials

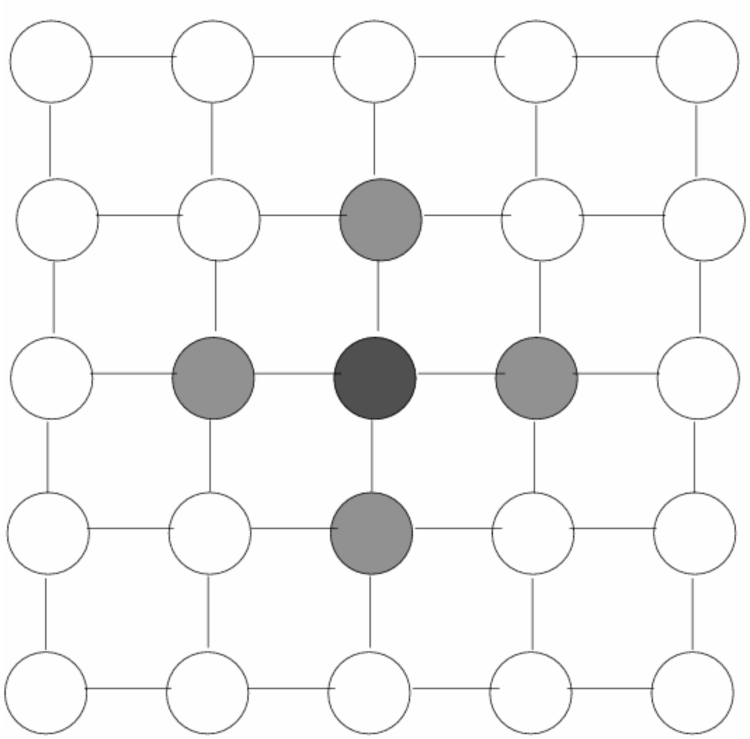
$$p(\mathbf{x}|\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} \prod_{c \in C} \psi_c(\mathbf{x}_c|\boldsymbol{\theta}_c)$$

- In terms of energy functions

$$\begin{aligned} \psi_c(\mathbf{x}_c) &= \exp[-H_c(\mathbf{x}_c)] \\ \log p(\mathbf{x}) &= -[\sum_{c \in \mathcal{C}} H_c(\mathbf{x}_c) + \log Z] \end{aligned}$$

Ising model

- 2D Grid on {-1,+1} variables
- Neighboring variables are correlated



$$H_{ij} = \begin{pmatrix} W_{ij} & -W_{ij} \\ -W_{ij} & W_{ij} \end{pmatrix}$$

W > 0: ferro magnet
W < 0: anti ferro magnet (frustrated)

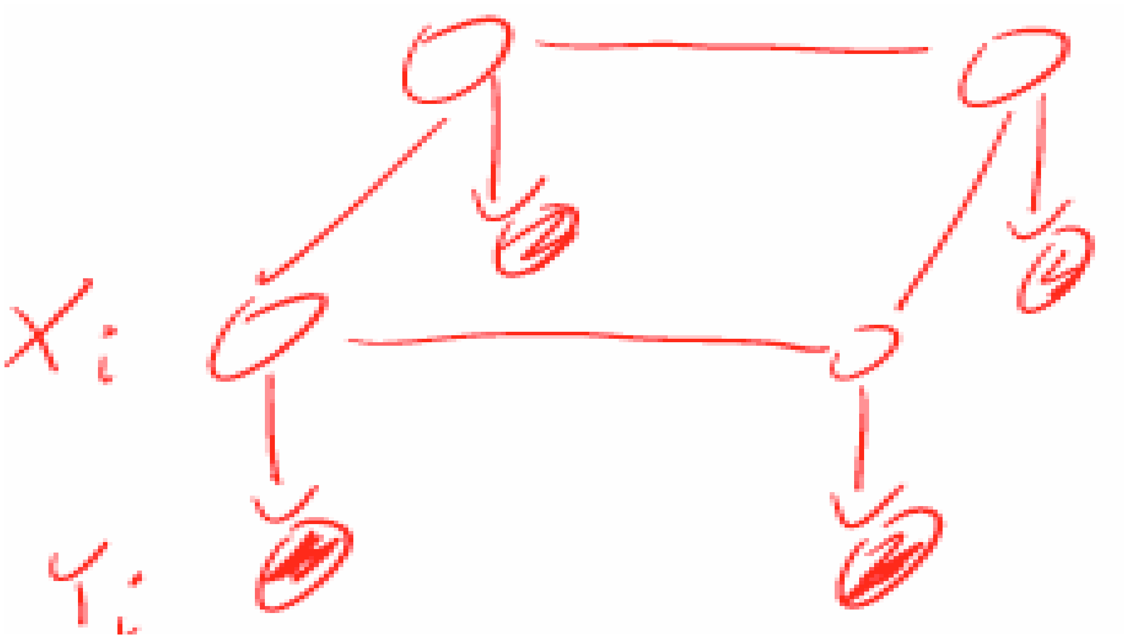
$$\begin{aligned} H(\mathbf{x}) &= -\mathbf{x}^T \mathbf{W} \mathbf{x} = -\sum_{<ij>} W_{ij} x_i x_j \\ p(\mathbf{x}|\boldsymbol{\theta}) &= \frac{1}{Z(\boldsymbol{\theta})} \exp[-\beta H(\mathbf{x}|\boldsymbol{\theta})] \end{aligned}$$

β=1/T = inverse temperature

Local evidence

$$p(x,y) = p(x)p(y|x) = \left[\frac{1}{Z}\prod_{<ij>}\psi_{ij}(x_i,x_j)\right]\left[\prod_i p(y_i|x_i)\right]$$

$$p(y_i|x_i) = \mathcal{N}(y_i|x_i,\sigma^2)$$

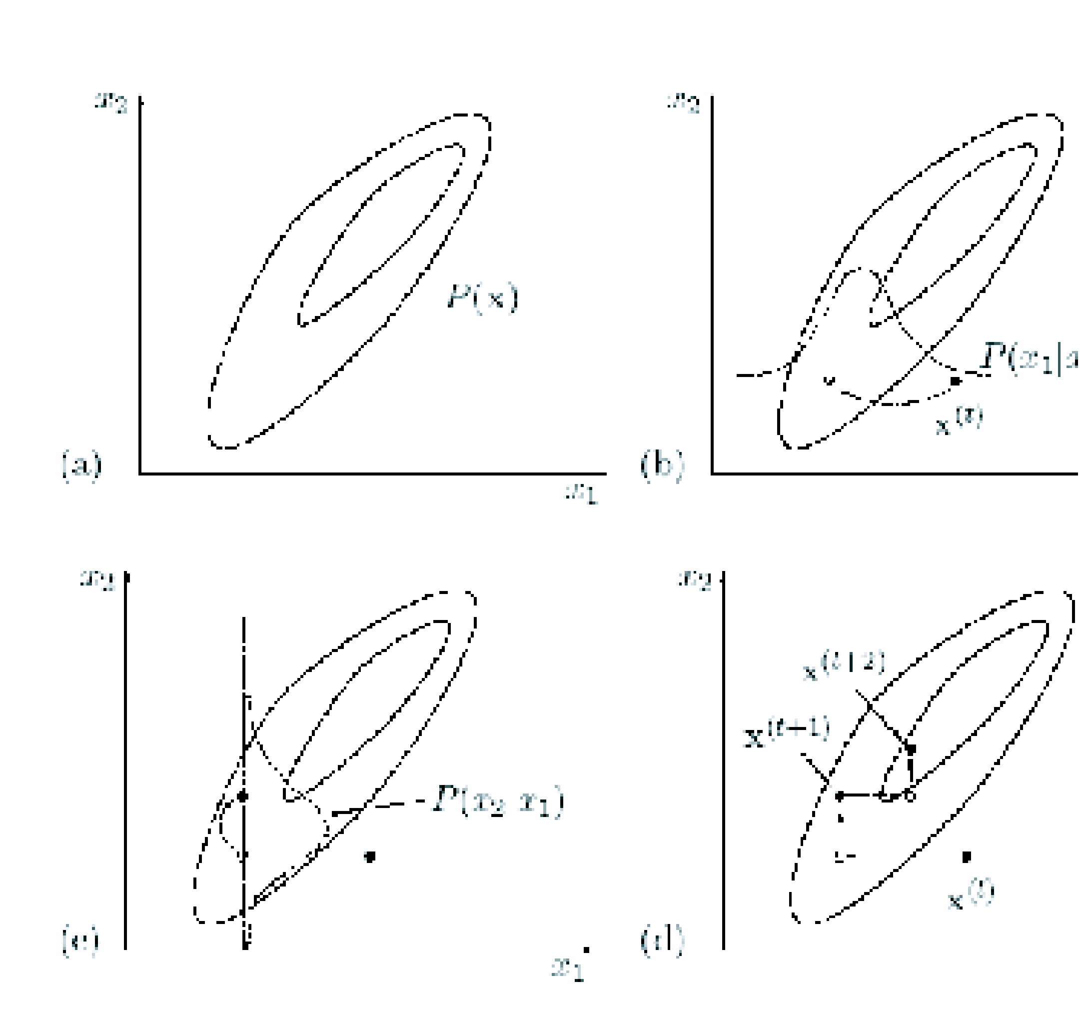


Gibbs sampling

- A way to draw samples from $p(x_{1:d}|y,\theta)$ one variable at a time, ie. $p(x_i|x_{-i})$

1. $x_1^{s+1} \sim p(x_1|x_2^s, \ldots, x_D^s)$
2. $x_2^{s+1} \sim p(x_2|x_1^{s+1}, x_3^s, \ldots, x_D^s)$
3. $x_i^{s+1} \sim p(x_i|x_{1:i-1}^{s+1}, x_{i+1:D}^s)$
4. $x_D^{s+1} \sim p(x_D|x_1^{s+1}, \ldots, x_{D-1}^{s+1})$

Gibbs sampling from a 2d Gaussian



Mackay 29.13

Gibbs sampling in an MRF

- Full conditional depends only on Markov blanket

$$
\begin{aligned}
p(X_i = \ell | x_{-i}) &= \frac{p(x_i = \ell, x_{-i})}{\sum_{\ell'} p(X_i = \ell', x_{-i})} \\
&= \frac{(1/Z)[\prod_{j \in N_i} \psi_{ij}(X_i = \ell, x_j)][\prod_{<jk>: j,k \notin F_i} \psi_{jk}(x_j, x_k)]}{(1/Z) \sum_{\ell'} [\prod_{j \in N_i} \psi_{ij}(X_i = \ell', x_j)][\prod_{<j,k>: j,k \notin F_i} \psi_{jk}(x_j, x_k)]} \\
&= \frac{\prod_{j \in N_i} \psi_{ij}(X_i = \ell, x_j)}{\sum_{\ell'} \prod_{j \in N_i} \psi_{ij}(X_i = \ell', x_j)}
\end{aligned}
$$

Gibbs sampling in an Ising model

- Let $\psi(x_i,x_j) = \exp(W\, x_i\, x_j)$, $x_i = +1,-1$.

$$
\begin{aligned}
p(X_i = +1 | x_{-i}) &= \frac{\prod_{j \in N_i} \psi_{ij}(X_i = +1, x_j)}{\prod_{j \in N_i} \psi_{ij}(X_i = +1, x_j) + \prod_{j \in N_i} \psi_{ij}(X_i = -1, x_j)} \\
&= \frac{\exp[J \sum_{j \in N_i} x_j]}{\exp[J \sum_{j \in N_i} x_j] + \exp[-J \sum_{j \in N_i} x_j]} \\
&= \frac{\exp[J w_i]}{\exp[J w_i] + \exp[-J w_i]} \\
&= \sigma(2 J w_i) \\
w_i &= \sum_{j \in N_i} x_j
\end{aligned}
$$

$$\sigma(u) = 1/(1+e^{-u})$$

Adding in local evidence

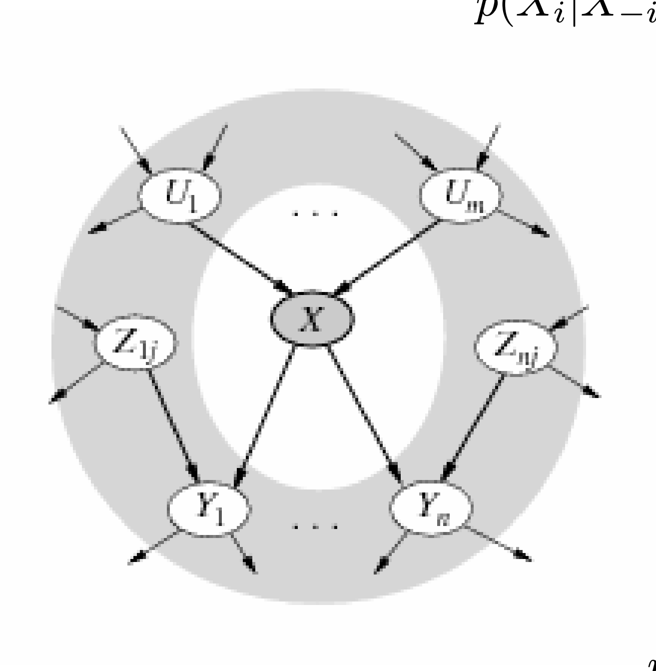
- Final form is

$$p(X_i = +1|x_{-i}, y) \quad = \quad \frac{\exp[Jw_i]\phi_i(+1, y_i)}{\exp[Jw_i]\phi_i(+1, y_i) + \exp[-Jw_i]\phi_i(-1, y_i)}$$

Run demo

Gibbs sampling for DAGs

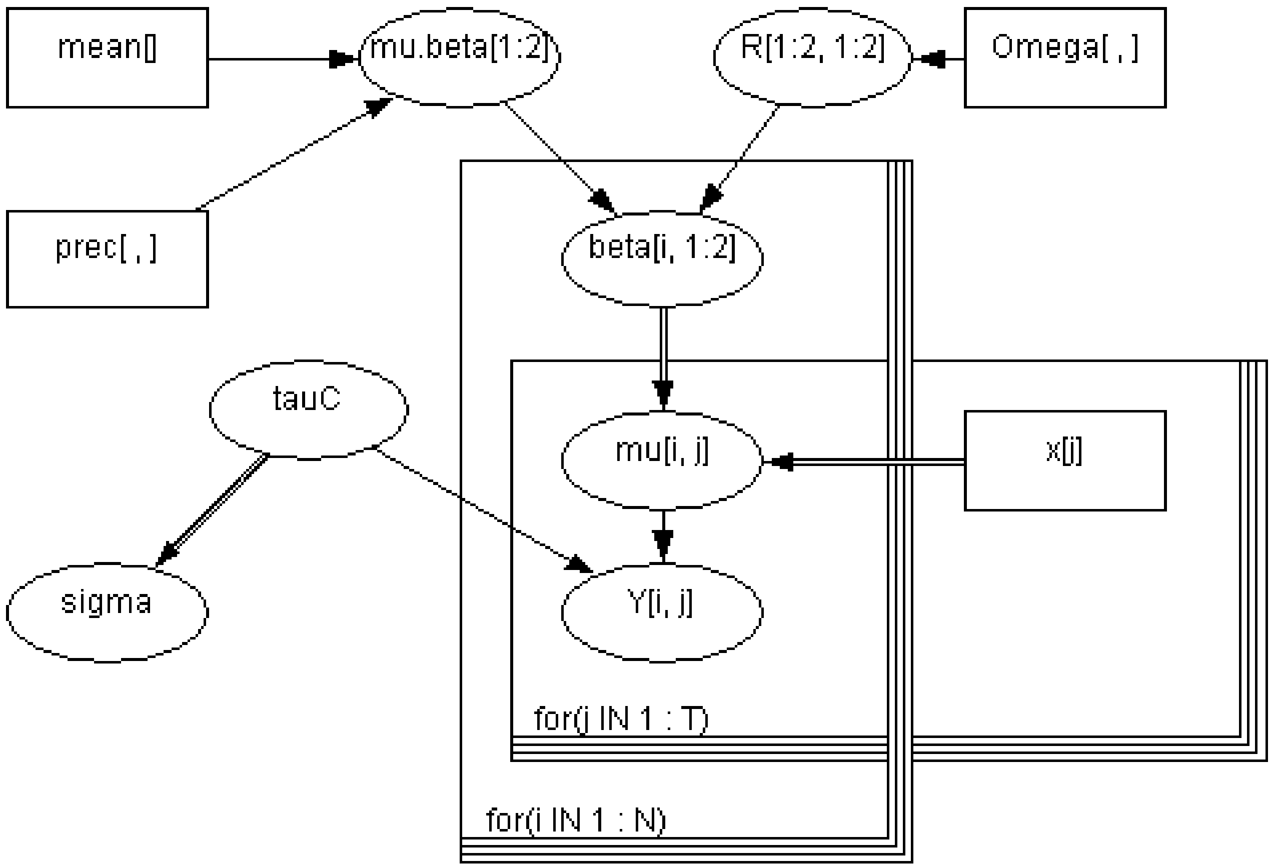
- The Markov blanket of a node is the set that renders it independent of the rest of the graph.
- This is the parents, children and co-parents.



$$
\begin{aligned}
p(X_i|X_{-i}) &= \frac{p(X_i, X_{-i})}{\sum_x p(X_i, X_{-i})} \\
&= \frac{p(X_i, U_{1:n}, Y_{1:m}, Z_{1:m}, R)}{\sum_x p(x, U_{1:n}, Y_{1:m}, Z_{1:m}, R)} \\
&= \frac{p(X_i|U_{1:n})[\prod_j p(Y_j|X_i, Z_j)]P(U_{1:n}, Z_{1:m}, R)}{\sum_x p(X_i = x|U_{1:n})[\prod_j p(Y_j|X_i = x, Z_j)]P(U_{1:n}, Z_{1:m}, R)} \\
&= \frac{p(X_i|U_{1:n})[\prod_j p(Y_j|X_i, Z_j)]}{\sum_x p(X_i = x|U_{1:n})[\prod_j p(Y_j|X_i = x, Z_j)]}
\end{aligned}
$$

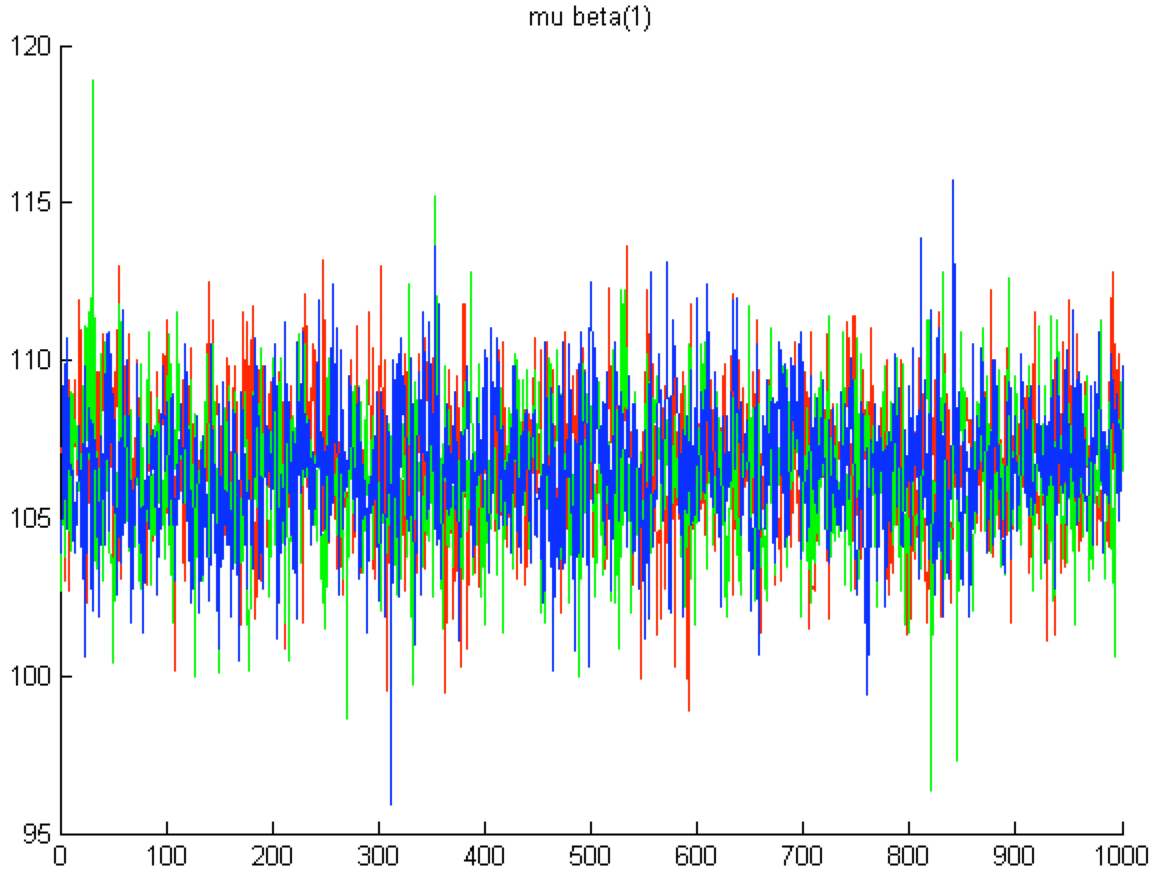
$$
p(X_i|X_{-i}) \propto p(X_i|Pa(X_i)) \prod_{Y_j \in ch(X_i)} p(Y_j|Pa(Y_j))
$$

Birats



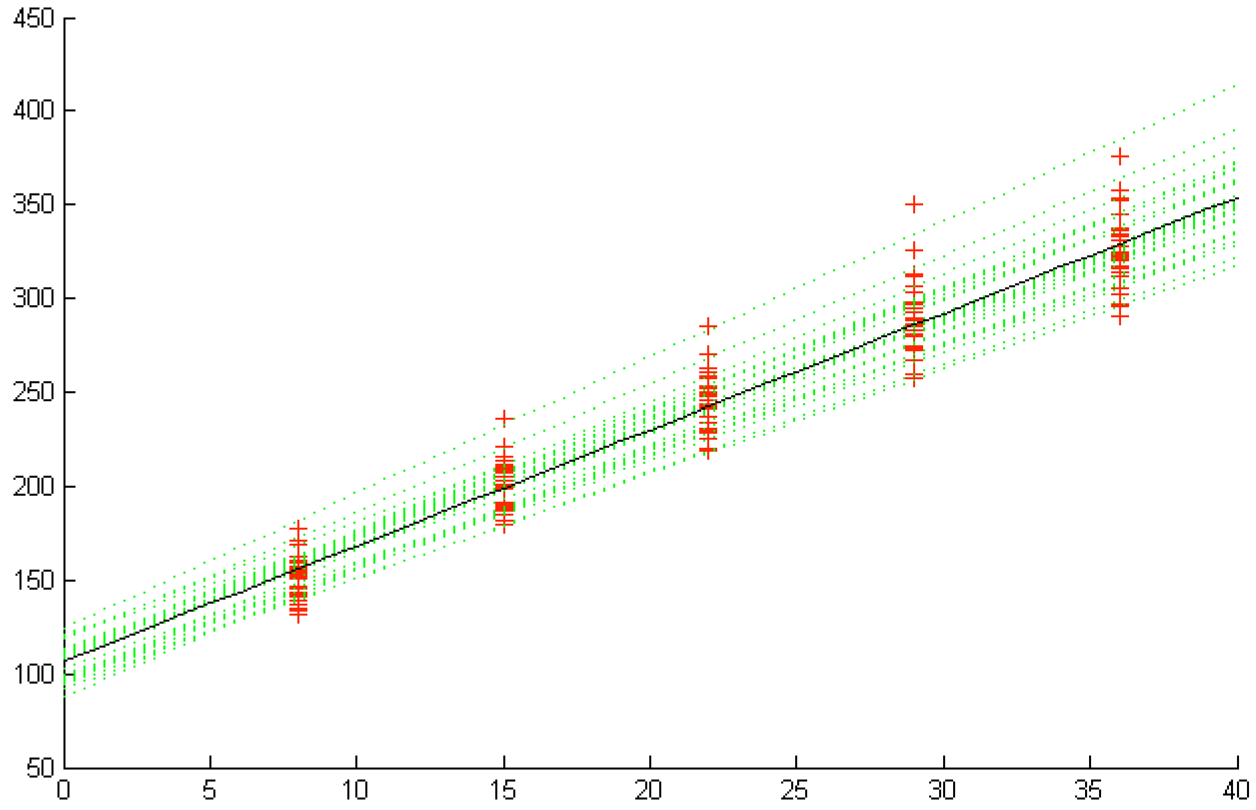
mean[]
mu.beta[1:2]
R[1:2, 1:2]
Omega[,]
prec[,]
beta[i, 1:2]
tauC
mu[i, j]
x[j]
sigma
Y[i, j]
for(j IN 1 : T)
for(i IN 1 : N)

Samples



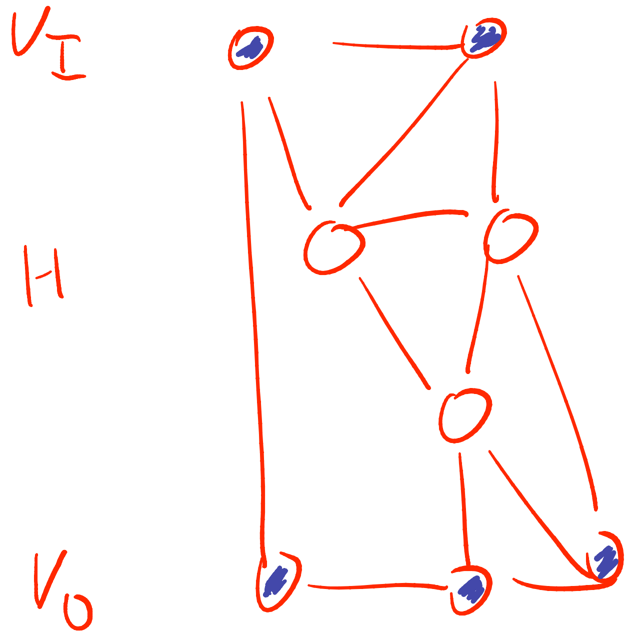
mu beta(1)
120
115
110
105
100
95
0
100
200
300
400
500
600
700
800
900
1000

Posterior predictive check

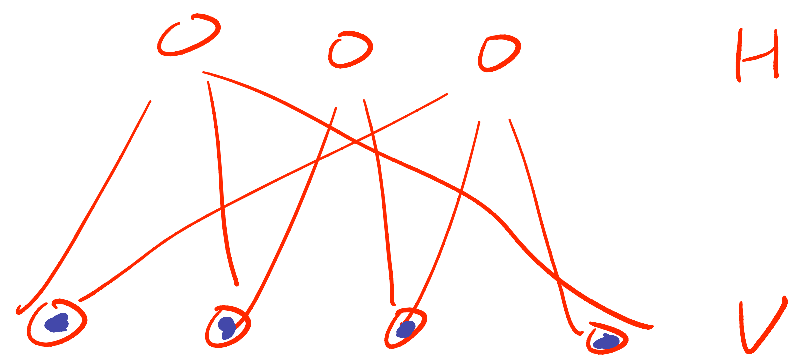


Boltzmann machines

Ising model where the graph structure is arbitrary, and the weights W are learned by maximum likelihood



Restricted Boltzmann machine



Hopfield network

Boltzmann machine with no hidden nodes (fully connected Ising model)