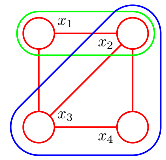
Semantics of undirected graphical models

- Intuitively, for any two nodes, x_i and x_j, not connected by a link, $x_i \perp x_j \mid \mathbf{x}/\{i,j\}$.
- So, $p(\ldots,x_i,\ldots,x_j,\ldots) = p\left(x_i \mid \mathbf{x}/\{i,j\}\right) p\left(x_j \mid \mathbf{x}/\{i,j\}\right) p\left(\mathbf{x}/\{i,j\}\right)$
- This suggests that an appropriate factorization should not have factors with these two nodes together.
- Direct links imply that we have a relation, and so we cannot put directly linked nodes into the same factor.
- A group of nodes that are all connected cannot be factored by the above rule.

Semantics of undirected graphical models

- So, we add nodes into factors, provided that they are all connected.
- This leads to describing the semantics in terms of maximal cliques.
  - A clique is fully connected subset of nodes from the graph
  - A maximal clique is a clique where no node in the graph can be added to it without it ceasing to be a clique.



All parwise linked nodes are cliques. For example $\{x_1,x_2\}$ is a clique (green). However, it is not a maximal clique. $\{x_2,x_3,x_4\}$ is a maximal clique (blue). If we add another node (only x_1 is left) we no longer have a clique.

Semantics of undirected graphical models (2)

Let C index maximal cliques. Then

$$p(\mathrm{x}) = \frac{1}{Z}\prod_c \psi_C(\mathrm{x}_C)$$

where Z $= \sum_x \prod_c \psi_C(\mathrm{x}_C)$ (or $\int_x \prod_c \psi_C(\mathrm{x}_C)$) is the partition function, and $\psi_C(\mathrm{x}_C)$ are the clique potentials.

If x_i and x_j do not share an edge, then they do not share cliques.

So $p(\mathrm{x}) = \frac{1}{Z}\prod_{c(i)} \psi_C(\mathrm{x}_C) \prod_{c(j)} \psi_C(\mathrm{x}_C) \prod_{c \notin c(i) \cup c(j)} \psi_C(\mathrm{x}_C)$

Semantics of undirected graphical models (3)

We will assume that all $\psi_C(\mathrm{x}_C) > 0$.

In general, we leave the semantics of $\psi_C(\mathrm{x}_C)$ open, but for undirected graphs that come from directed graphs where each node has one parent, the semantics follows that for the directed graphs.

Since $\psi_C(\mathrm{x}_C) > 0$ we will often write $\psi_C(\mathrm{x}_C) = \exp\{-E(\mathrm{x}_C)\}$ where E() is the energy function.

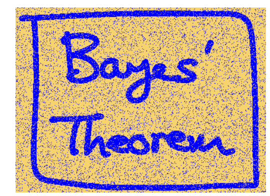
Semantics of undirected graphical models (3)

Writing $\psi_C(\mathrm{x}_C) = \exp\{-E(\mathrm{x}_C)\}$ means that

$$
\begin{aligned}
p(x) &= \frac{1}{Z}\prod_c \psi_x(\mathrm{x}_C) \\
&= \frac{1}{Z}\prod_c \exp\{-E(\mathrm{x}_C)\} \\
&= \frac{1}{Z}\exp\left\{\sum_c -E(\mathrm{x}_C)\right\} \\
&= \frac{1}{Z}\exp\{-E(x)\} \qquad \text{Where } E(x) = \sum_c E(\mathrm{x}_C)
\end{aligned}
$$

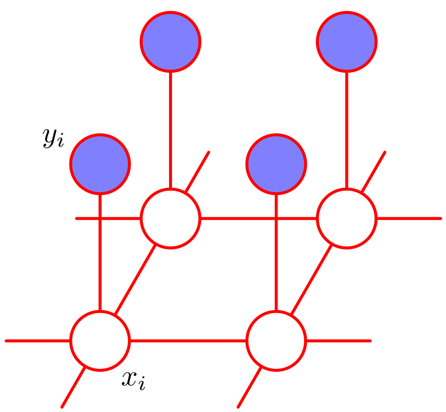
Example of a Markov random field

- Consider a binary image (pixels are either black or white).
- Pixels are represented by {-1,1}.
- Suppose the image have is an underlying accurate image where some of the bits have been flipped by a noise process.



Example of a Markov random field (2)

- Undirected graphical model.



Example of a Markov random field (2)

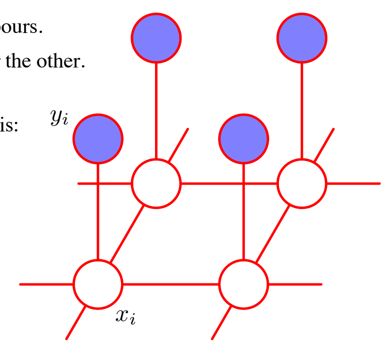
- For low energy (high probability)

$x_i = y_i$ most of the time (set by noise level)
$x_i = x_j$ most of the time if i and j are neighbours.
x_i could be biased to have one value or the other.

A simple energy function for the entire grid is:

$$E(\mathbf{x},\mathbf{y}) = h\sum x_i - \beta\sum_{i,j} x_i x_j - \eta\sum_i x_i y_i$$

Because values are 1 and -1, being the same makes the sums bigger, different smaller.



Example of a Markov random field (3)

$x_i = y_i$ most of the time (set by noise level)

$x_i = x_j$ most of the time if i and j are neighbours.

x_i could be biased to have one value or the other.

For each $\{x_i, y_i\}$ maximum clique, $\mathrm{E}(x_i, y_i) = -\eta \bullet x_i \bullet y_i \quad (\eta > 0)$
(high probablity corresponds to low energy)

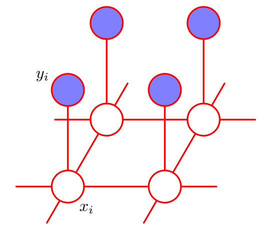
For unique $\{x_i, x_{j \in neighbor(i)}\}$ max clique, $\mathrm{E}(x_i, x_j) = -\beta \bullet x_i \bullet x_j \quad (\beta > 0)$

For a subset of the above cliques, one for each i, add in a term $h \bullet x_i$.

Example of a Markov random field (4)

- Notice in the previous analysis we assigned arguably symmetric cliques different potentials
  - Left boundary x_i might get different potentials than right boundary x_i.
  - Some x_{ij} get a factor for the bias, other do not.
- Notice that exact assignment to clique potentials may not matter
- We can jump readily quickly to the overall picture, hence:

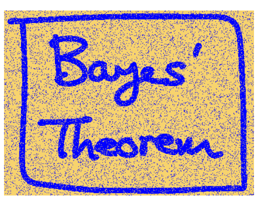
$$E(\mathbf{x}, \mathbf{y}) = h \sum x_i - \beta \sum_{i,j} x_i x_j - \eta \sum_i x_i y_y$$



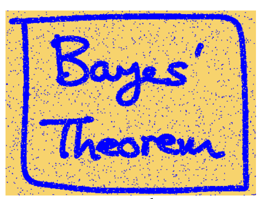
Example of a Markov random field (3)

- Finding a low energy (high probability) state using ICM (iterated conditional modes).
  - Initialize x_i to y_i.
  - For each i, change x_i if energy decreases.
  - Repeat until energy no longer can be decreased.
- Converges to a local minimum because we only decrease.

original



with noise



result

From directed to undirected

- Easy case (all nodes have at most one parent).
- Example:

- Becomes:

From directed to undirected

- Convert:

$$p(\mathrm{x}) = p(x_1)\,p(x_2|x_1)\,p(x_3|x_2)\;\cdots\;p(x_{N-1}|x_{N-2})\,p(x_N|x_{N-1})$$

- To:

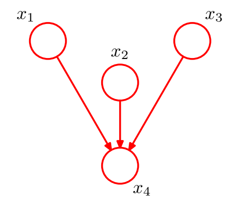
$$p(\mathrm{x}) = \Psi(x_1,x_2)\,\Psi(x_2,x_3)\;\ldots\;\Psi(x_{N-2},x_{N-1})\,\Psi(x_{N-1},x_N)$$

- Inspection suggests:

$$\Psi(x_1,x_2) = p(x_1)\,p(x_2|x_1)$$
$$\Psi(x_2,x_3) = p(x_3|x_2)$$
$$\cdots$$
$$\Psi(x_{N-2},x_{N-1}) = p(x_{N-1}|x_{N-2})$$
$$\Psi(x_{N-1},x_N) = p(x_N|x_{N-1})$$

From directed to undirected

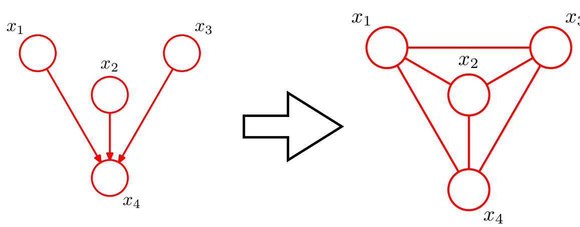
- Harder case (some nodes have multiple parents).
- Example:



- Because this implies conditioning on three variables, the potentials for the clique are a function of four variables.
- These nodes need to be part of a clique (but they are not).

From directed to undirected

- Solution is to marry the parents.
- This makes the graph "moral".
- Note that moralization looses conditional independence information.



From directed to undirected

- Complete algorithm
  - Make the graph moral.
  - Initialize each maximal clique potential to one.
  - Multiply each factor in p() into an appropriate clique potential.
  - Note that Z=1