Machine Learning

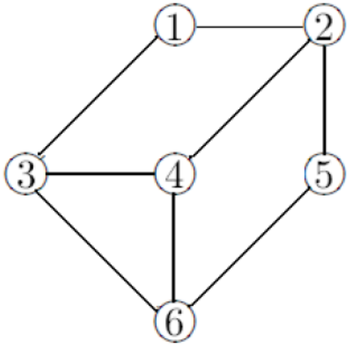
CSE 6363 (Fall 2016)

Lecture 21 Markov Random Fields

Heng Huang, Ph.D.

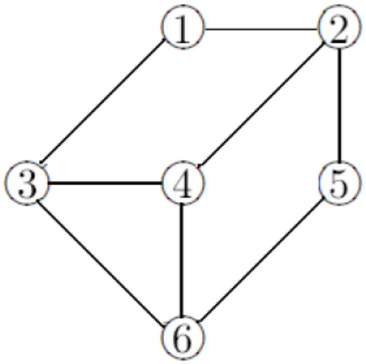
Department of Computer Science and Engineering

Basic Graph Concepts



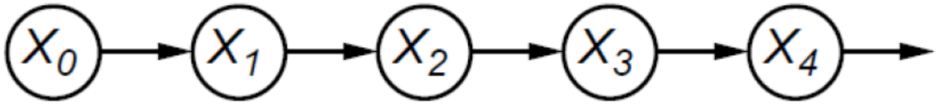
- $\mathcal{V} = \{1, 2, 3, 4, 5, 6\}$
- $\mathcal{E} = \{(1,2), (1,3), (2,4), (2,5), (3,4), (3,6), (4,6), (5,6)\}$
- $\mathcal{N}(4) = \{2, 3, 6\}$
- Examples of cliques: $\{(1), (3,4,6), (2,5)\}$
- Set of all cliques: $\mathcal{V} \cup \mathcal{E} \cup \{3, 4, 6\}$

Separation



- Let A, B, C be three disjoint subsets of $\mathcal{V}$
- C separates A from B if any path from a node in A to a node in B contains some node in C
- Example: $C = \{1, 4, 6\}$ separates $A = \{3\}$ from $B = \{2, 5\}$

Markov Chains



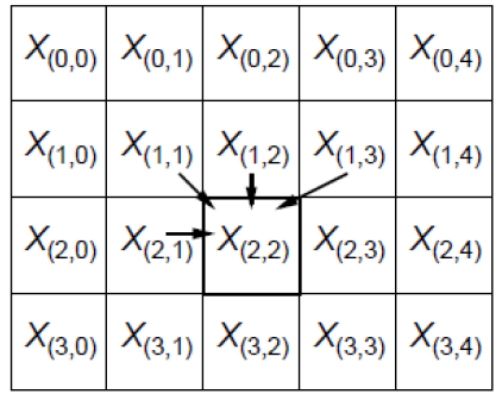
- Graphical model: Associate each node of a graph with a random variable (or a collection thereof)
- Homogeneous 1-D Markov chain:

$$p(x_n|x_i, i < n) = p(x_n|x_{n-1})$$

- Probability of a sequence given by:

$$p(x) = p(x_0) \prod_{n=1}^{N} p(x_n|x_{n-1})$$

2D Markov Chains



- Advantages:
  - Simple expressions for probability
  - Simple parameter estimation
- Disadvantages:
  - No natural ordering of image pixels
  - Anisotropic model behavior

Random Fields on Graph

- Consider a collection of random variables $\mathbf{x} = (x_1, x_2, \ldots, x_N)$ with associated joint probability distribution $p(\mathbf{x})$
- Let A, B, C be three disjoint subsets of $\mathcal{V}$. Let $\mathbf{x}_A$ denote the collection of random variables in A.
  - Conditional independence: $A \perp\!\!\!\perp B \mid C$
  - $A \perp\!\!\!\perp B \mid C \Leftrightarrow p(\mathbf{x}_A, \mathbf{x}_B | \mathbf{x}_C) = p(\mathbf{x}_A | \mathbf{x}_C) p(\mathbf{x}_B | \mathbf{x}_C)$
- Markov random field: undirected graphical model in which each node corresponds to a random variable or a collection of random variables, and the edges identify conditional dependencies.

Markov Properties

Pairwise Markovianity:

- $(n_i, n_j) \notin \mathcal{E} \Rightarrow x_i$ and x_j are independent when conditioned on all other variables

$$p(x_i, x_j | \mathbf{x}_{\backslash\{i,j\}}) = p(x_i | \mathbf{x}_{\backslash\{i,j\}}) p(x_j | \mathbf{x}_{\backslash\{i,j\}})$$

Local Markovianity:

- Given its neighborhood, a variable is independent on the rest of the variables

$$p(x_i | \mathbf{x}_{\mathcal{V}\backslash\{i\}}) = p(x_i | \mathbf{x}_{\mathcal{N}(i)})$$

Global Markovianity:

- Let A, B, C be three disjoint subsets of $\mathcal{V}$. If

  C separates A from $B \Rightarrow p(\mathbf{x}_A, \mathbf{x}_B | \mathbf{x}_C) = p(\mathbf{x}_A | \mathbf{x}_C) p(\mathbf{x}_B | \mathbf{x}_C)$,

  then $p(\cdot)$ is global Markov w.r.t. $\mathcal{G}$.

Hammersley-Clifford Theorem

Consider a random field $\mathbf{x}$ on a graph $\mathcal{G}$, such that $p(\mathbf{x}) > 0$. Let $\mathcal{C}$ denote the set of all maximal cliques of the graph.

- If the field has the local Markov property, then $p(\mathbf{x})$ can be written as a Gibbs distribution:

$$p(\mathbf{x}) = \frac{1}{Z} \exp\left\{ - \sum_{C \in \mathcal{C}} V_C(\mathbf{x}_C) \right\},$$

  where Z, the normalizing constant, is called the partition function; $V_C(\mathbf{x}_C)$ are the clique potentials

- If $p(\mathbf{x})$ can be written in Gibbs form for the cliques of some graph, then it has the global Markov property.

Fundamental consequence: every Markov random field can be specified via clique potentials.

Markov Random Field

- Posterior probability of the labelling y given observation x is :

$$P(y|x) = \frac{1}{Z(x)} \prod_c \varphi_c(\boldsymbol{y}_C; x)$$

cliques

Potential functions

where $Z(x) = \sum \prod_c \varphi_c(\boldsymbol{y}_C; x)$ is called the *partition function*.

- Since we define potential function is strictly positive, we can express them as exponentials:

$$\mathrm{P}(\mathbf{y} \mid \mathbf{x}) = \frac{1}{Z(\mathbf{x})} \exp\{-E(\mathbf{y}; \mathbf{x})\}$$

The *energy function* $E(\mathbf{y}; \mathbf{x})$ usually has some structured form:

$$E(\mathbf{y}; \mathbf{x}) = \sum_c \psi_c(\mathbf{y}_c; \mathbf{x})$$

MAP Inference

- Posterior probability of the labelling y given observation x is :

Bayes Rule

$$\underbrace{\mathrm{P}(\mathbf{y} \mid \mathbf{x})}_{\text{posterior}} = \frac{\overbrace{\mathrm{P}(\mathbf{x} \mid \mathbf{y})}^{\text{likelihood}} \cdot \overbrace{\mathrm{P}(\mathbf{y})}^{\text{prior}}}{\mathrm{P}(\mathbf{x})}$$

Maximum a Posteriori (MAP) inference: $\mathbf{y}^\star = \operatorname{argmax}_{\mathbf{y}} \mathrm{P}(\mathbf{y} \mid \mathbf{x})$.

- The most possible labeling is to minimize the Energy

$$\mathbf{y}^\star = \operatorname*{argmax}_{\mathbf{y}} \mathrm{P}(\mathbf{y} \mid \mathbf{x}) = \operatorname*{argmin}_{\mathbf{y}} E(\mathbf{y}; \mathbf{x}) .$$

Pairwise MRF

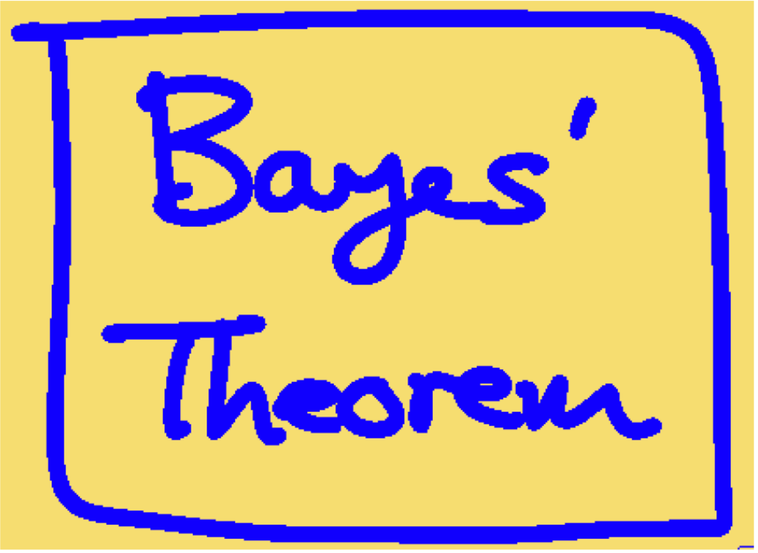
- Most common energy function for image labeling

$$E(\mathbf{y};\mathbf{x}) = \sum_{c} \psi_c(\mathbf{y}_c;\mathbf{x})$$

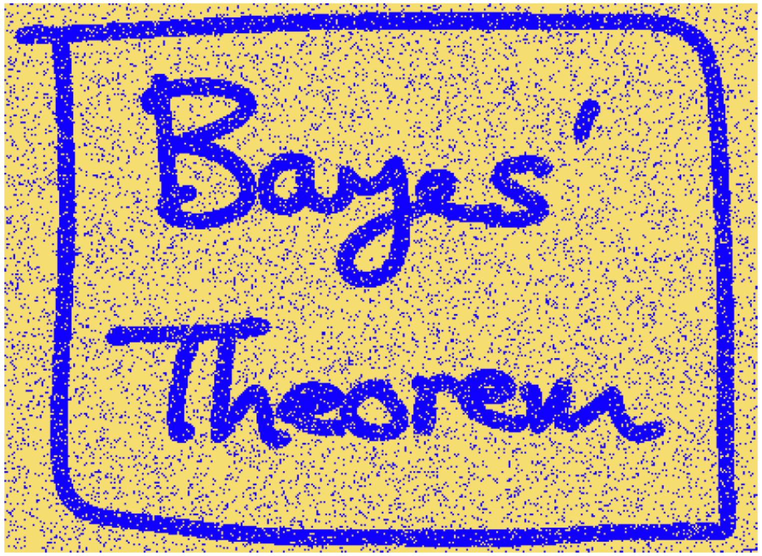
$$= \sum_{i \in \mathcal{V}} \underbrace{\psi_i^U(y_i;\mathbf{x})}_{\text{Unary}} + \sum_{ij \in \mathcal{E}} \underbrace{\psi_{ij}^P(y_i, y_j)}_{\text{Pairwise}}$$

Example MRF Model: Image Denoising

- How can we retrieve the original image given the noisy one?



Original image Y

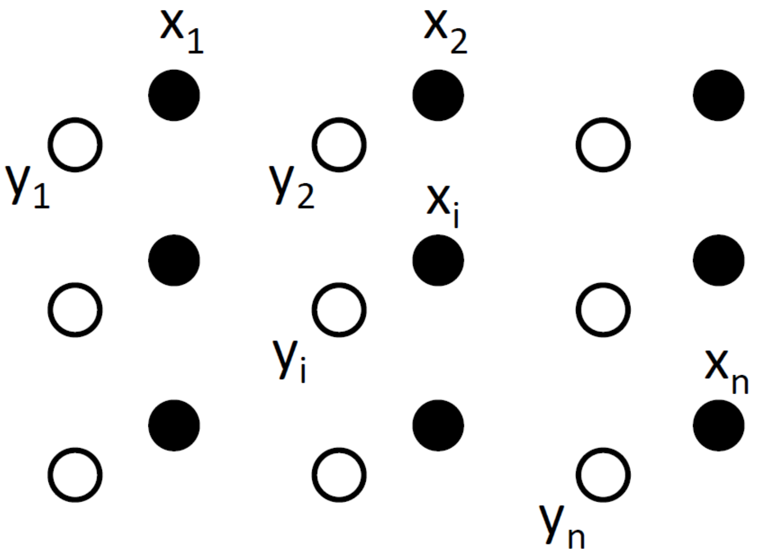


Noisy image X(**Input**)

MRF Formulation

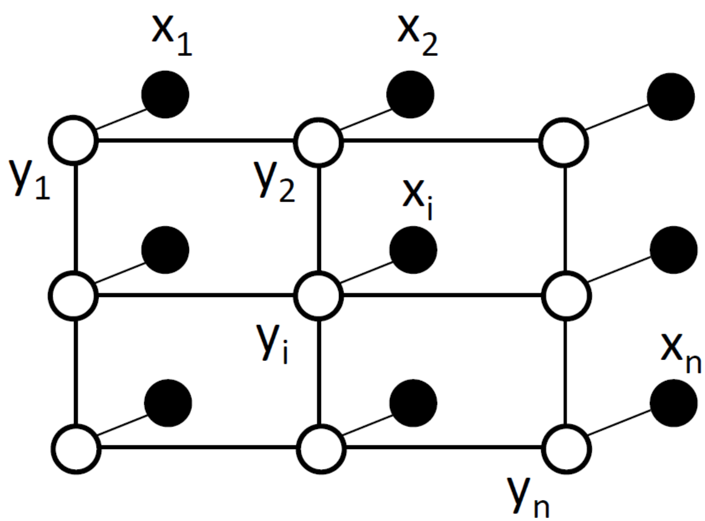
- <u>Nodes</u>
  - For each pixel i,
    - y_i : latent variable (value in original image)
    - x_i : observed variable (value in noisy image)

    Simple setting : $x_i, y_i \in \{-1,1\}$



MRF Formulation

- Edges
  - x_i,y_i of each pixel i correlated
  - neighboring pixels, similar value(smoothness)



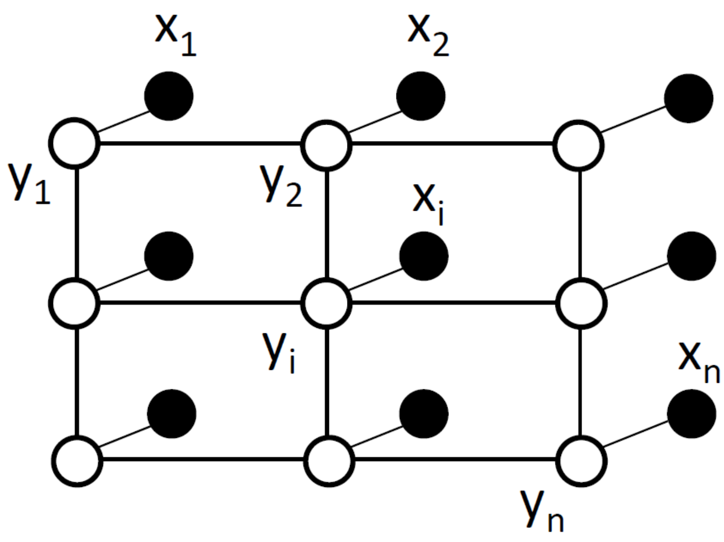
$\phi(x_i, y_i) = -\beta x_i y_i$

$\psi(y_i, y_j) = -\alpha y_i y_j$

MRF Formulation

Energy function

$$E(y; x) = \sum_{ij} \psi(y_i, y_j) + \sum_i \phi(x_i, y_i)$$
$$= -\alpha \sum_{ij} y_i y_j - \beta \sum_i x_i y_i$$



Optimization

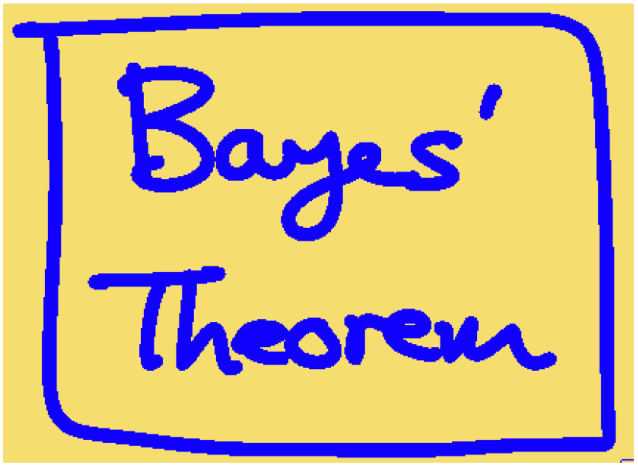
Energy function

$$E(y; x) = \sum_{ij} \psi(y_i, y_j) + \sum_i \phi(x_i, y_i)$$
$$= -\alpha \sum_{ij} y_i y_j - \beta \sum_i x_i y_i$$

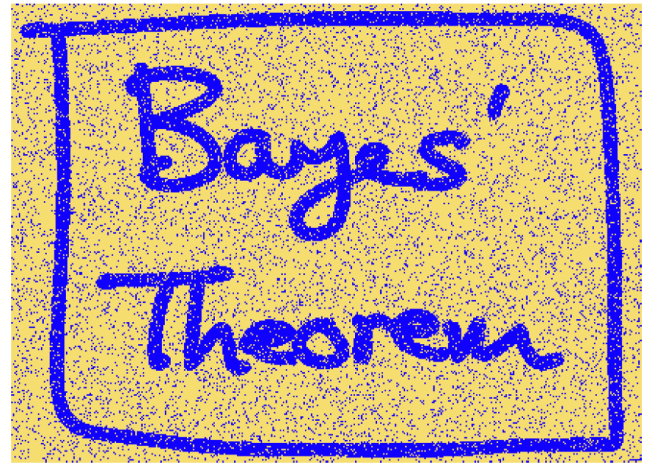
Iterated Conditional Modes (ICM)

- Initialize $y_i = x_i$ for all i
- Take a y_i, fix others, flip y_i if $-y_i$ make energy lower
- Repeat until converge

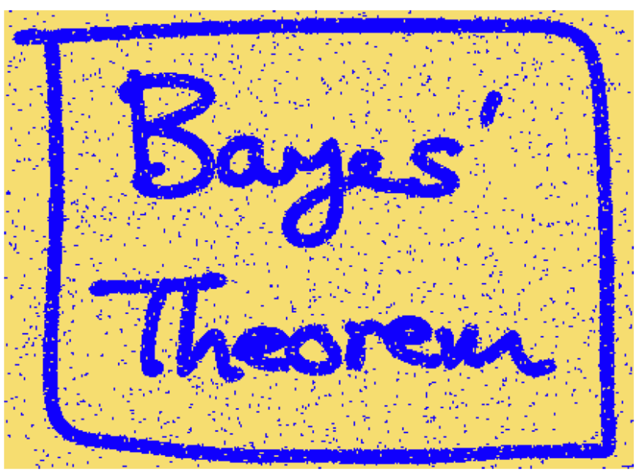
Results



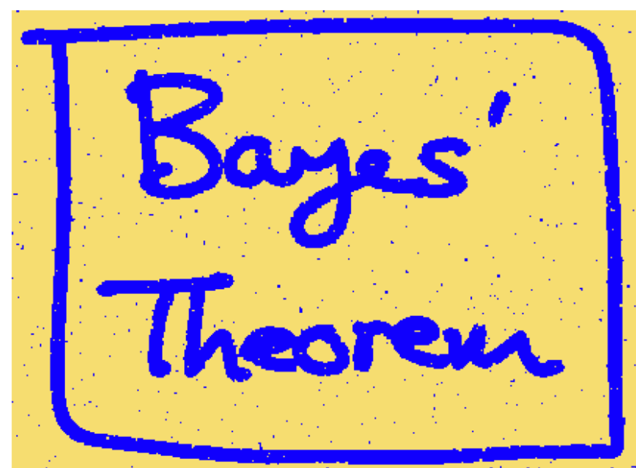
original image



noise image(10%)



Restored by ICM(4%)



Restored by Graph Cut(<1%)