Aprendizaje automatizado

Campos aleatorios de Markov

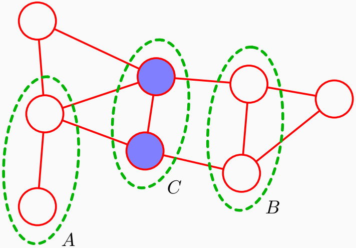
Gibran Fuentes-Pineda

Abril 2023

Campos aleatorios de Markov (MRFs)

- Factorización de la distribución conjunta e independencias condicionales se representan con grafos no dirigidos

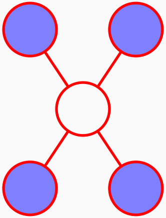
$$A \perp\!\!\!\perp B \mid C$$



Tomada de PRML (Bishop 2009)

Cobija de Markov para MRFs

- Cualquier nodo es condicionalmente independiente de cualquier otro nodo en el grafo dado únicamente sus vecinos



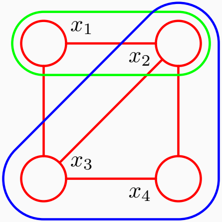
Tomada de PRML (Bishop 2009)

- Distribución conjunta se descompone de acuerdo a funciones sobre los "cliques" del grafo

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

Cliques

- Un **clique** c es un subconjunto de nodos completamente conectados



Tomada de PRML (Bishop 2009)

Probabilidad conjunta en MRFs

- $P(\mathsf{x}) > 0$ satisface las propiedades de independencia condicional de un grafo no dirigido $\mathcal{G}$ si y sólo si puede representarse como un producto de factores, uno por clique máximo, es decir

$$P(\mathsf{x}) = \frac{1}{Z} \prod_{\mathcal{C}} \psi_{\mathcal{C}}(\mathsf{x}_{\mathcal{C}})$$

  donde $\mathcal{C}$ es el conjunto de todos los cliques máximos de $\mathcal{G}$, $\psi_{\mathcal{C}}(\mathsf{x}_{\mathcal{C}})$ se conocen como funciones potencial y Z es la función de partición dada por

$$Z \triangleq \sum_{\mathsf{x}} \prod_{\mathcal{C}} \psi_{\mathcal{C}}(\mathsf{x}_{\mathcal{C}})$$

Funciones potencial

- Como $P(\mathsf{x}) > 0$, las funciones potencial se pueden expresar como exponenciales

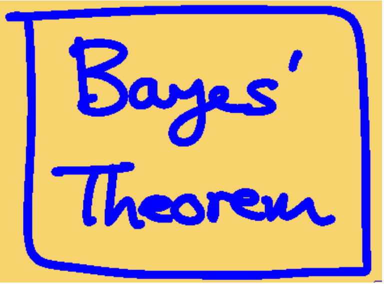
$$\psi_{\mathcal{C}}(\mathsf{x}_{\mathcal{C}}) = exp(-E(\mathsf{x}_{\mathcal{C}}))$$

  donde $E(\mathsf{x}_{\mathcal{C}})$ es la función de energía y la representación exponencial se conoce como distribución de Boltzmann

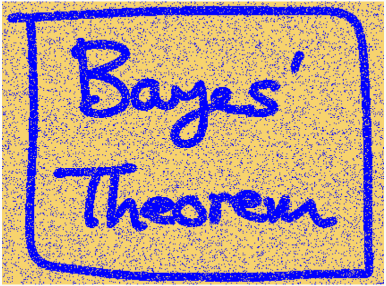
- La distribución conjunta está dada por el producto de funciones potencial, por lo que la energía total se obtiene sumando las energías de cada clique máximo

MRF para quitar ruido en una imagen binaria (1)

Original ($x_i \in \{-1, +1\}$)



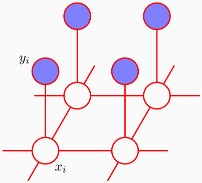
Ruidosa ($y_i \in \{-1, +1\}$)



Imágenes tomadas de Bishop, *Pattern Recognition and Machine Learning*, 2006.

MRF para quitar ruido en una imagen binaria (2)

- Presuposiciones
  - Ruido se genera cambiando el signo de la imagen original
  - Signo de pixel x_i en imagen original se correlaciona con el de sus vecinos x_j y con el del pixel y_i de imagen ruidosa
- MRF
  - Nodo corresponden a pixeles en imagen original y ruidosa
  - Cliques máximos entre cada par de pixeles vecinos ($\{x_i, x_j\}$) y entre cada pixel de imagen original y ruidosa ($\{x_i, y_i\}$)



Imágenes tomadas de Bishop, *Pattern Recognition and Machine Learning*, 2006.

MRF para quitar ruido en una imagen binaria (3)

- Funciones de energía
  - Para $\{x_i, x_j\}$: $-\beta x_i x_j$
  - Para $\{x_i, y_i\}$: $-\eta x_i y_i$
  - Preferencia a un signo: hx_i
- Energía total

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

- Distribución conjunta

$$P(\mathrm{x}, \mathrm{y}) = \frac{1}{Z} e^{-E(\mathrm{x}, \mathrm{y})}$$

MRF para quitar ruido en una imagen binaria (4)

- Algoritmo *iterated conditional modes* (ICM) para obtener $P(\mathrm{x}|\mathrm{y})$
    1. Inicializa $x_i = y_i$
    2. Calcula la energía dado que un nodo $x_i = +1$ y $x_i = -1$ y fija el valor con menor energía.
    3. Repite 2 hasta cumplir criterio de paro

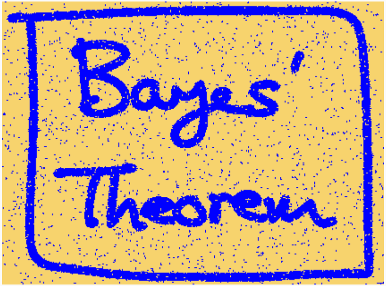


Imagen tomada de Bishop, *Pattern Recognition and Machine Learning*, 2006.

Conversión de redes bayesianas a MRFs

- Nos debemos asegurar que el conjunto de variables que aparecen en cada distribución condicional sean miembros de al menos un clique

- Para nodos de una red bayesiana con más de un padre es necesario agregar aristas entre los nodos padre en el MRF (**moralización**)

Conversión de redes bayesianas a MRFs

- Cadenas en redes bayesianas

$$P(\mathbf{x}) = P(x_1)P(x_2|x_1)P(x_3|x_2)\cdots P(x_n|x_{n-1})$$

Tomada de PRML (Bishop 2009)

Conversión de redes bayesianas a MRFs

- Cadenas en redes bayesianas

$$P(\mathbf{x}) = P(x_1)P(x_2|x_1)P(x_3|x_2)\cdots P(x_n|x_{n-1})$$

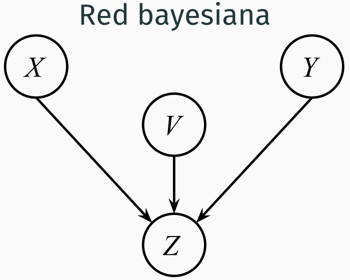
Tomada de PRML (Bishop 2009)

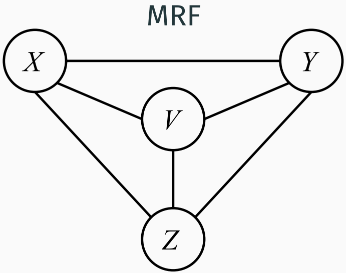
- Cadenas en MRFs

$$P(\mathbf{x}) = \frac{1}{Z}\psi_{1,2}(x_1, x_2)\psi_{2,3}(x_2, x_3)\cdots\psi_{n,n-1}(x_{n-1}, X_n)$$

Tomada de PRML (Bishop 2009)

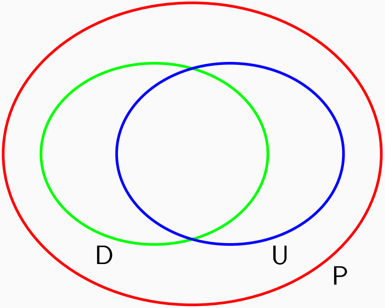
Conversión: otra topología





¿Que distribuciones podemos representar?

- Redes bayesianas vs MRFs



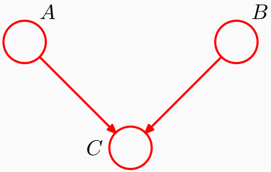
Tomada de PRML (Bishop 2009)

Conversión: limitaciones

- Algunas redes bayesianas no se pueden representar como MRFs

$$A \perp\!\!\!\perp B \mid \varnothing$$

$$A \not\perp\!\!\!\perp B \mid C$$



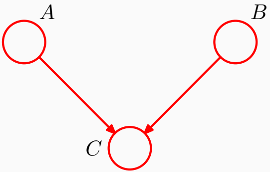
Tomada de PRML (Bishop 2009)

Conversión: limitaciones

- Algunas redes bayesianas no se pueden representar como MRFs

$$A \perp\!\!\!\perp B \mid \varnothing$$

$$A \not\!\perp\!\!\!\perp B \mid C$$



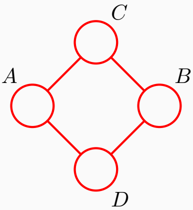
Tomada de PRML (Bishop 2009)

- Y viceversa

$$A \not\!\perp\!\!\!\perp B \mid \varnothing$$

$$C \perp\!\!\!\perp D \mid A \cup B$$

$$A \perp\!\!\!\perp B \mid C \cup D$$



Tomada de PRML (Bishop 2009)