Zoubin Ghahramani - Introduction to Graphical Models

Camilo López

02/28/2008

1 Introducción

En esta conferencia se presentan los tres modelos principales de modelos gráficos (factor, dirigido y no dirigido). Tienen en común que cada nodo corresponde a una variable aleatoria mientras que los arcos representan las dependencias estadísticas entre las variables.

La ventaja con los modelos gráficos es que representan y permiten visualizar de una forma intuitiva las relaciones entre un gran número de variables. Esta facilidad en el análisis de un modelo permite responder preguntas como "Es A dependiente de B dado que se conoce el valor de C?" con tan solo mirar el grafo.

2 Independencia condicional

$X \perp Y \mid V$ Esta notación nos dice que X es condicionalmente independiente de Y dado un evento V.

La independencia condicional entre estas dos variables se da cuando la probabilidad de ocurrencia de X no depende de Y dado que se conoce V

$$X \perp Y \mid V \Leftrightarrow p\left(X \mid Y, V\right) = p\left(X \mid V\right)$$

De otra forma,

$$X \perp Y \mid V \Leftrightarrow p\left(X, Y \mid V\right) = p\left(X \mid V\right) p\left(Y \mid V\right)$$

y en general se puede expresar la independencia condicional entre un conjunto de variables:

$$\mathbf{X} \perp \mathbf{Y} \mid \mathfrak{V} \Leftrightarrow \{X \perp Y \mid \mathfrak{V},\, \forall X \in \mathbf{X}\ and\ \forall Y \in \mathbf{Y}\}$$

Por su parte, la independencia marginal es una independencia condicional dado el conjunto vacío, es decir, si no hay más información. Esta es la noción de independencia estadística con la que estábamos acostumbrados a trabajar.

$$X \perp Y \mid \emptyset \Leftrightarrow p\left(X, Y \mid \emptyset\right) = p\left(X\right) p\left(Y\right)$$

Algunos ejemplos:

- La habilidad de dos equipos A y B es marginalemente independiente, cuando los jugadores son escogidos aleatoriamante de un gran número de posibilidades.

$$Ability\ of\ Team\ A \perp Ability\ of\ Team\ B$$

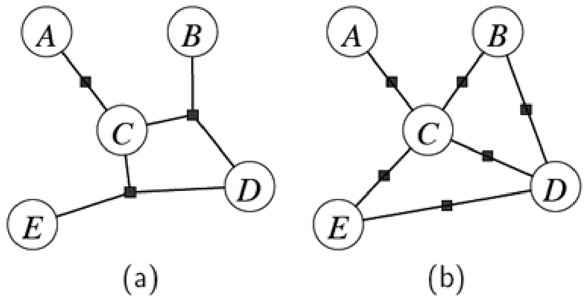

Figure 1: Factor Graphs

- Esta independencia se pierde cuando se condicionan las variables con nuevo conocimiento que posiblemente relaciona las variables, por ejemplo al conocer los resultados entre estos dos equipos se tiene una independencia condicional.

$$Ability\ of\ Team\ A \perp Ability\ of\ Team\ B \mid Outcome\ of\ A\ vs\ B\ Game$$

- El costo de una multa por exceder la velocidad no depende del tipo de carro cuando se conoce la velocidad

$$Amount\ of\ Speeding\ Fine \perp Type\ of\ Car \mid Speed$$

- Los genes de unos niños son independientes de los genes de los abuelos dado que se conocen los genes de los padres

$$Child's\ Genes \perp Grandparents's\ Genes \mid Parents's\ Genes$$

3 Grafos de Factores (Factor Graphs)

La característica principal de los grafos de factores es la existencia de dos tipos de nodos:

- Los círculos representan variables aleatorias (e.g. A).
- Los puntos representan los factores en la distribución conjunta (e.g. $g(\cdot)$).

Un grafo de este tipo representa la factorización de la distribución de probabilidad conjunta donde cada factor es una función no-negativa de sus argumentos, de esta forma la probabilidad conjunta para los ejemplos de la figura 1 es:

(a) $P(A,B,C,D,E) = \frac{1}{Z} g_1(A,C)\, g_2(B,C,D)\, g_3(C,D,E)$

(b) $P(A,B,C,D,E) = \frac{1}{Z} g_1(A,C)\, g_2(B,C)\, g_3(C,D)\, g_4(B,D)\, g_5(C,E)\, g_6(D,E)$

Z es una constante de normalización. Por ejemplo en (a), si todas las variables son discretas y toman valores en $\mathfrak{A} \times \mathfrak{B} \times \mathfrak{C} \times \mathfrak{D} \times \mathfrak{E}$:

$$Z = \sum_{a\in\mathfrak{A}} \sum_{b\in\mathfrak{B}} \sum_{c\in\mathfrak{C}} \sum_{d\in\mathfrak{D}} \sum_{e\in\mathfrak{E}} g_1(A=a, C=c)\, g_2(B=b, C=c, D=d)\, g_3(C=c, D=d, E=e)$$

Se dice que dos nodos son vecinos si comparten un factor común y un *camino* se define como una secuencia de nodos vecinos.

Hecho: $X \perp Y \mid \mathfrak{V}$ si todos los caminos entre X y Y contienen algún nodo $V \in \mathfrak{V}$

Corolario: Dados los vecinos de X, la variable X es condicionalmente independiente de todas las demás variables:

$$X \perp Y \mid \ vecinos\,(X)\,, \qquad \forall Y \notin \{X \cup \ vecinos\,(X)\}$$

La independencia condicional se prueba de la siguiente forma:

$$X \perp Y \mid V \Leftrightarrow p\,(X \mid Y, V) = p\,(X \mid V) \qquad (a)$$

Se asume una factorización para la probabilidad conjunta:

$$P\,(X, Y, V) = \frac{1}{z} g_1\,(X, V)\, g_2\,(Y, V) \qquad (b)$$

Al sumar para todo X se obtiene:

$$P\,(Y, V) = \frac{1}{z}\left[\sum_X g_1\,(X, V)\right] g_2\,(Y, V) \qquad (c)$$

Dividiendo (b) por (c) se obtiene:

$$P\,(X \mid Y, V) = \frac{g_1\,(X, V)}{\sum_X g_1\,(X, V)} \qquad (d)$$

Y esta no depende de Y. Por lo tanto la fctorización (b) implica independencia condicional (a).

4 Modelos de grafos no dirigidos (Undirected Graphical Models)

De forma similar al modelo visto anteriormente, en los modelos de grafos no dirigidos se puede expresar la probabilidad conjunta como una factorización, en este caso cada factor es una función de un subconjunto de variables que están completamente conectadas.

$$P\,(X) = \frac{1}{Z}\prod_j g_j\left(\mathbf{X}_{\mathbf{C_j}}\right)$$

donde X es el conjunto de todas las variables, $\mathrm{X} = (\mathrm{x}_1, \ldots, \mathrm{x}_\mathrm{K})$, y

$$C_j \subseteq \{1, 2, \ldots, K\}$$

son subconjuntos del conjunto de todas las variables, y $\mathrm{X_S} \equiv (\mathrm{x_K} : \mathrm{K} \in \mathrm{S})$

Para la especificación del grafo se debe crear un nodo por cada variable y se tiene que un conjunto C_j es aquel que forma un subgrafo completamente conectado, o clique.

En la figura 2 se puede ver un ejemplo de grafo no dirigido. El cálculo de la probabilidad conjunta es:

$$P\,(A, B, C, D, E) = \frac{1}{Z} g_1\,(A, C)\, g_2\,(B, C, D)\, g_1\,(C, D, E)$$

Hecho: $X \perp Y \mid \mathfrak{V}$ si todos los caminos entre X y Y contienen algún nodo $V \in \mathfrak{V}$

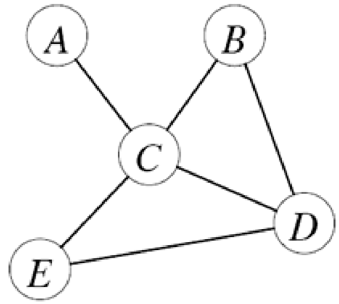

Figure 2: Undirected Graphical Models

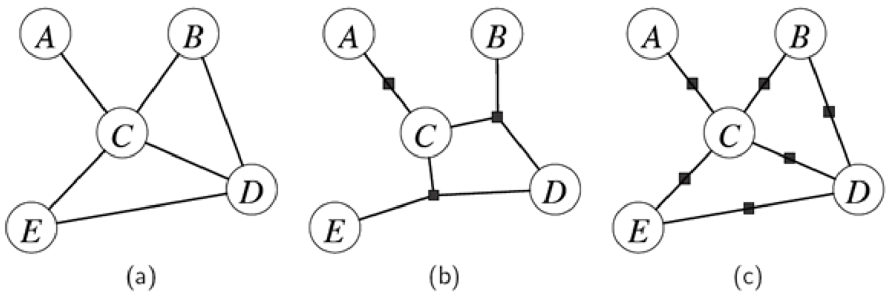

Figure 3: Comparing Undirected Graphs and Factor Graphs

Corolario: Dados los vecinos de X, la variable X es condicionalmente independiente de todas las demás variables:

$$X \perp Y \mid \textit{vecinos}(X), \qquad \forall Y \notin \{X \cup \textit{vecinos}(X)\}$$

Manta de Markov (Markov Blanket): $\mathfrak{V}$ es una Manta de Markov para X si y solo si $X \perp Y \mid \mathfrak{V} \forall\, Y \notin \{X \cup \mathfrak{V}\}$

Frontera de Markov (Markov Boundary): Es la mínima Manta de Markov $\equiv \textit{vecinos}(X)$

5 Comparación entre Grafos no dirigidos y Grafos de factores

En la figura 3 pueden verse tres grafos que representan las mismas relaciones de independencia condicional, ya que los nodos tienen los mismos vecinos. Sin embargo los grafos de factores muestran más información, por ejemplo, a pesar de que los grafos (b) y (c) son del mismo tipo los cliques tienen diferentes factores dentro de ellos mismos.

El problema con estos modelos es que hay situaciones de dependencia que no pueden ser explicadas. En el modelo en el cual intervienen: Lluvia (R), Suelo húmedo (G) y Aspersor (S) R y S son marginalmente independientes y son condicionalmente dependientes dado G. Si se utiliza uno de estos modelos, cuando se representa la dependencia entre R y G, y entre S y G (Figura 4a) existe un camino

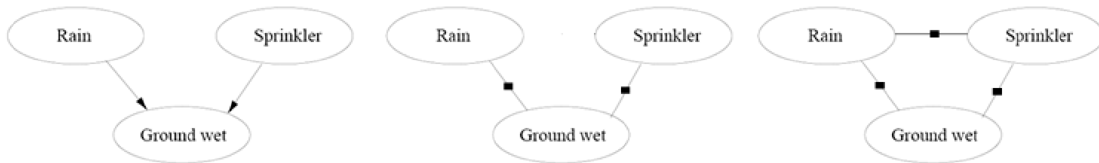

Figure 4: Problems with Undirected Graphs and Factor Graphs

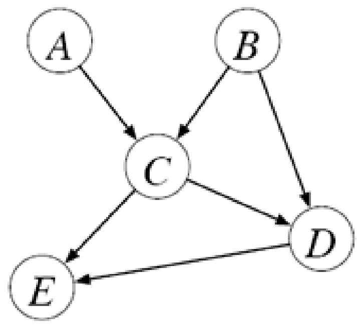

Figure 5: Directed Acyclic Graphical Models

entre R y S, lo que permite decir dos cosas. La primera es que R y S son dependientes por la existencia del camino. La segunda es que R y S son independientes dado G pues todos los caminos entre R y S incluyen a G. Las dos son incorrectas. Para tratar de solucionar la segunda se podría pensar en agregar un camino entre R y S (Figura 4b) pero esto los hace marginalmente dependientes. Entonces no hay forma de representar correctamente el modelo.

6 Modelo de Grafos Acíclicos Dirigidos (Redes Bayesianas)

La probabilidad conjunta puede factorizarse como la probabilidad de cada nodo dados sus padres.

$$p\left(X_1, \ldots, X_n\right) = \prod_{i=1}^{n} p\left(X_i \mid X_{pa(i)}\right)$$

donde pa(i) son los padres del nodo i.

Para el ejemplo de la figura 5 se tiene que

$$p\left(A, B, C, D, E\right) = p\left(A\right) p\left(B\right) p\left(C \mid A, B\right) p\left(D \mid B, C\right) p\left(E \mid C, D\right)$$

Semántica: $X \perp Y \mid \mathfrak{V}$ if $\mathfrak{V}$ d-separates (dependency-separates) X from Y

Definición: $\mathfrak{V}$ d-separates X de Y si cualquier camino, no dirigido, entre X y Y es bloqueado por $\mathfrak{V}$. Un camino es bloqueado por V si existe un nodo W en el camino tal que se cumpla una de las siguientes condiciones:

1. W tiene flechas que convergen ($\rightarrow W \leftarrow$)y, ni W ni sus descendientes son observados ($in\ \mathfrak{V}$), o
2. W no tiene flechas que convergen ($\rightarrow W \rightarrow$) o ($\leftarrow W \rightarrow$)y W es observado ($W \in \mathfrak{V}$)

Corolario: La frontera de Markov para X: $\{parents\left(X\right) \cup children\left(X\right) \cup parents - of - children\left(X\right)\}$

Cabe resaltar que las redes bayesianas son un caso particular de Grafos Acíclicos Dirigidos.

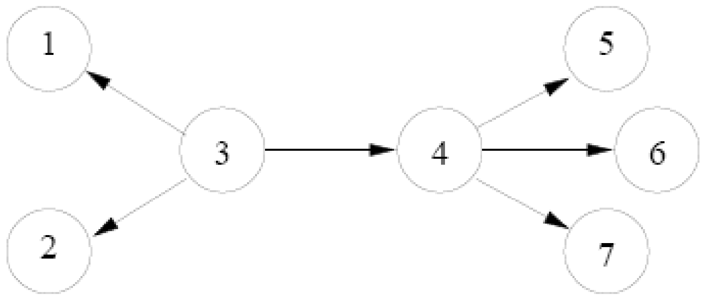

Figure 6: From Directed Trees to Undirected Trees.

6.1 Algunos ejemplos

Siguiendo con la figura 5, se busca analizar los siguientes casos:

$A \perp B$ $A \to C \leftarrow B$ es bloqueado por C ya que tiene flechas que convergen, lo mismo pasa para $A \to C \to D \leftarrow B$, bloqueado por D y $A \to C \to E \leftarrow D \leftarrow B$, bloqueado por E. Como $\mathfrak{V}$ es el conjunto vacío ni los nodos que bloquean ni sus descendientes pueden pertenecer a él por lo que se concluye que A es marginalmente independiente de B.

$(A \perp B \mid C)$ El camino $A \to C \leftarrow B$ no es bloqueado ya que $C \in \mathfrak{V}$, por lo tanto $\mathfrak{V}$ no "d-separates" a A y B, entonces $not\,(A \perp B \mid C)$.

$A \perp D \mid \{B, C\}$ Es correcto ya que $A \to C \leftarrow D$ es bloqueado por C, $A \to C \leftarrow B \to D$ es bloquado por B y $A \to C \to E \leftarrow D$ es bloqueado por E.

$(A \perp B \mid E)$ Es incorrecto ya que $A \to C \leftarrow B$ no es bloqueado (Las flechas convergen en C, pero E es descendiente de C y$E \in \{E\}$).

6.2 De árboles dirigidos a árboles no dirigidos

Se tiene un grafo dirigido (figura 6), que representa la siguiente probabilidad conjunta:

$$p(x_1, x_2, \ldots, x_7) = p(x_3)\, p(x_1 \mid x_3)\, p(x_2 \mid x_3)\, p(x_4 \mid x_3)\, p(x_5 \mid x_4)\, p(x_6 \mid x_4)\, p(x_7 \mid x_4)$$

y aplicando de nuevo el teorema de Bayes:

$$p(x_1, x_2, \ldots, x_7) = \frac{p(x_1, x_3)\, p(x_2, x_3)\, p(x_4, x_3)\, p(x_5, x_4)\, p(x_6, x_4)\, p(x_7, x_4)}{p(x_3)\, p(x_3)\, p(x_4)\, p(x_4)\, p(x_4)}$$

$$p(x_1, x_2, \ldots, x_7) = \frac{\textit{producto de cliques}}{\textit{producto de intersecciones}}$$

Esto puede expresarse como un producto de factores,

$$p(x_1, x_2, \ldots, x_7) = g(x_1, x_3)\, g(x_2, x_3)\, g(x_4, x_3)\, g(x_5, x_4)\, g(x_6, x_4)\, g(x_7, x_4)$$

Lo cual corresponde a la expresión de un grafo no dirigido cuya estructura es la misma que el grafo dirigido solo que sin las direcciones de las flechas.

$$p(x_1, x_2, \ldots, x_7) = \prod_i g_i(C_i)$$

De esta forma un árbol dirigido puede ser convertido en un árbol no dirigido que represente la mismas mismas relaciones de independencia condicional y viceversa.

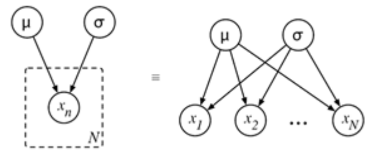

Figure 7: Plate notation

7 Grafos dirigidos para modelos estadísticos: Notación de plato (Plate notation)

Este es un método de representación muy usado cuando un subgrafo se presenta muchas veces con las mismas dependencias, lo que se busca es, en lugar de dibujarlo cada vez, representarlo una vez dentro de una caja, usualmente de línea punteada, con un subíndice que representa el número de repeticiones del subgrafo.