Lecture Pattern Analysis

Part 21: Conditional Independence

Christian Riess
IT Security Infrastructures Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg
June 25, 2021

Overview

- Variable independence enables factorization, and hence training and inference
- Conditional independence is an important special case[1]
- If $p(a|b, c) = p(a|c)$, then a is **conditionally independent** of b given c,

$$a \perp\!\!\!\perp b|c \tag{1}$$

- Conditional independence may occur in more complex expressions, e.g.,

$$p(a, b|c) = p(a|b, c)p(b|c) \tag{2}$$

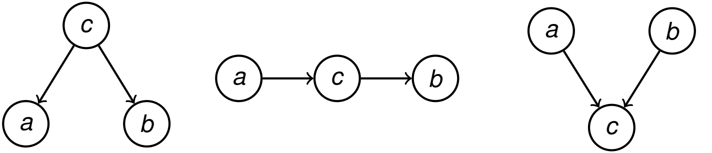
$$\stackrel{!}{=} p(a|c)p(b|c) \tag{3}$$

- Conditional independence is somewhat difficult on directed graphs, simpler on undirected graphs
- The **Markov blanket** makes inference on undirected graphs straightforward, which we will use for the upcoming Markov Random Fields

[1] The reference for this chapter is Bishop Sec. 8.2 and Sec. 8.3.1

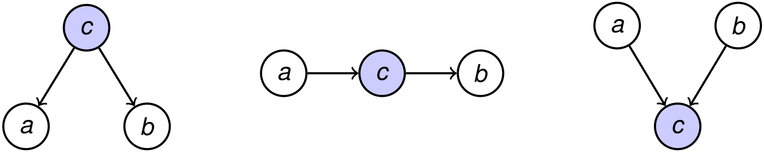
Conditional Independence on Directed Graphs

- We study three directed subgraphs that factorize $p(a, b, c)$:



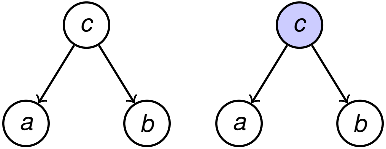
$p(c)p(a, b|c)$ $\quad$ $p(a)p(c|a)p(b|c)$ $\quad$ $p(a)p(b)p(c|a, b)$

- Shading indicates that a variable is observed:



- If c is observed, a and b become cond. **independent** in the first two graphs
- In the third graph, a and b become conditionally **dependent** if c is observed

Conditional Independence on the Tail-to-Tail Graph



- If c is unobserved: $a \not\perp\!\!\!\perp b \mid c$, since $p(a,b) \neq p(a)p(b)$:
  $p(a,b)$ is obtained by marginalizing over the possible values of c,

$$p(a,b) = \sum_c p(a|c)p(b|c)p(c) \neq p(a)p(b) \ , \tag{4}$$

- If c is observed, then $a \perp\!\!\!\perp b \mid c$:

$$p(a,b|c) = \frac{p(a,b,c)}{p(c)} = \frac{p(c)p(a|c)p(b|c)}{p(c)} = p(a|c)p(b|c) \tag{5}$$

Conditional Independence on the Head-to-Tail Graph

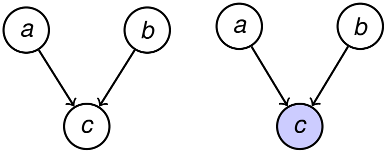
- If c is unobserved: $a \not\perp\!\!\!\perp b \mid c$, since $p(a, b) \neq p(a)p(b)$:

$$p(a, b) = \sum_c p(a)p(c|a)p(b|c) = p(a)p(b|a) \neq p(a)p(b) \ , \tag{6}$$

- If c is observed, then $a \perp\!\!\!\perp b \mid c$:

$$\begin{aligned} p(a, b|c) &= \frac{p(a, b, c)}{p(c)} = \frac{p(a)p(c|a)p(b|c)}{p(c)} = \frac{p(c)p(a|c)p(b|c)}{p(c)} \\ &= p(a|c)p(b|c) \end{aligned} \tag{7}$$

Conditional Independence on the Head-to-Head Graph



- If c is unobserved: $a \perp\!\!\!\perp b \mid c$, since $p(a, b) = p(a)p(b)$:

$$p(a, b) = \sum_c p(a)p(b)p(c|a, b) = p(a)p(b) \tag{8}$$

- If c is observed, then $a \not\perp\!\!\!\perp b \mid c$:

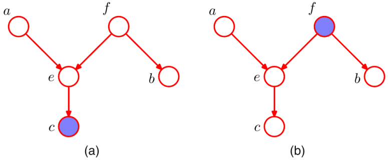
$$p(a, b|c) = \frac{p(a, b, c)}{p(c)} = \frac{p(a)p(b)p(c|a, b)}{p(c)} \neq p(a|c)p(b|c) \tag{9}$$

- Bishop Sec. 8.2.1 provides a numerical example to further illustrate this counter-intuitive case

D-Separation of Variables in Directed Graphs

- On more complex directed graphs, D-separation indicates whether for sets of nodes A, B, C the conditional independence $A \perp\!\!\!\perp B \mid C$ holds
- Variables in C are observed
- Consider all paths between A and B
- A path is **blocked** if it contains a node where
  - the node is in C and its arrows are tail-to-tail or head-to-tail
  - the node and its decendants are not in C and its arrows are head-to-head
- If all paths are blocked, A is d-separated from B by C, i.e., $A \perp\!\!\!\perp B \mid C$

Figure 8.22 Illustration of the concept of d-separation. See the text for details.

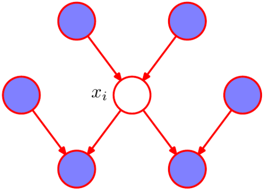


- Example: Left: a and b are dependent. Right: a and b are independent

Markov Blanket on Directed Graphs

- The minimal set of nodes that isolates a node from the rest of the graph is called a **Markov blanket**
- It includes the immediate parents, immediate children, and co-parents of immediate children:

Figure 8.26 The Markov blanket of a node $\mathbf{x}_i$ comprises the set of parents, children and co-parents of the node. It has the property that the conditional distribution of $\mathbf{x}_i$, conditioned on all the remaining variables in the graph, is dependent only on the variables in the Markov blanket.

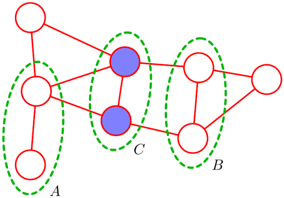


- In HMM training, we (luckily) have a much simpler dependency structure:
  - The iterative parameter updates assume that the previous and future states are tentatively fixed
  - Hence, we can use α and β as aggregated path probabilities to update $\boldsymbol{\pi}, \mathbf{A}, \mathbf{B}$

Conditional Independence and Markov Blanket on Undirected Graphs

- Conditional independence on undirected graphs is simpler: just separate *A* and *B* by observed nodes

Figure 8.27 An example of an undirected graph in which every path from any node in set A to any node in set B passes through at least one node in set C. Consequently the conditional independence property $A \perp\!\!\!\perp B \mid C$ holds for any probability distribution described by this graph.



- Analogously, the Markov blanket just includes the set of neighbors of a node:

Figure 8.28 For an undirected graph, the Markov blanket of a node x_i consists of the set of neighbouring nodes. It has the property that the conditional distribution of x_i, conditioned on all the remaining variables in the graph, is dependent only on the variables in the Markov blanket.