Reading Notes for ch8 Graphical Models

Xiang Pan

September 16, 2021

1 Graphical Models

We can trun the probability dependency of a random variable to a graphical model.

Note: We use the notations from the bishop book.

$$p(\mathbf{w} \mid \mathbf{T}) \propto p(\mathbf{w}) \prod_{n=1}^{N} p\left(t_n \mid \mathbf{w}\right) \tag{1}$$

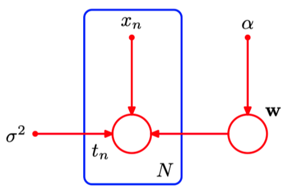

Figure 1: Graphical Model of observation

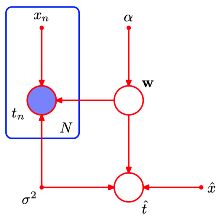

Figure 2: Prediction

$$p\left(\widehat{t}, \mathbf{t}, \mathbf{w} \mid \widehat{x}, \mathbf{x}, \alpha, \sigma^2\right) = \left[\prod_{n=1}^{N} p\left(t_n \mid x_n, \mathbf{w}, \sigma^2\right)\right] p(\mathbf{w} \mid \alpha) p\left(t \mid \widehat{x}, \mathbf{w}, \sigma^2\right) \tag{2}$$

$$p\left(t \mid \widehat{x}, \mathbf{x}, \mathbf{t}, \alpha, \sigma^2\right) \propto \int p\left(\widehat{t}, \mathbf{t}, \mathbf{w} \mid \widehat{x}, \mathbf{x}, \alpha, \sigma^2\right) \mathrm{d}\mathbf{w} \tag{3}$$

An alternative way to reduce the number of independent parameters in a model is by sharing parameters (also known as tying of parameters).

2 Conditional Independence

[1] Conditional Independence is widely used in causal learning [1]. We use the name of three types of conditional independence in causal learning.

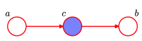

Figure 3: V-Structure (Chain Structure)

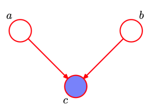

Figure 4: Collider Structure

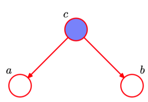

Figure 5: Fork Structure

[1] You can get the full version note at https://github.com/Xiang-Pan/NYU_Bayesian_Machine_Learning/blob/master/reading_notes/ch8/build/note4.pdf

V-Structure

$$p(a, b, c) = p(a)p(c \mid a)p(b \mid c) \tag{4}$$

$$a \not\perp\!\!\!\perp b \mid \emptyset \tag{5}$$

$$p(a, b \mid c) = \frac{p(a, b, c)}{p(c)} = p(a \mid c)p(b \mid c) \tag{6}$$

$$a \perp\!\!\!\perp b \mid c \tag{7}$$

Collider Structure

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

$$a \perp\!\!\!\perp b \mid \emptyset \tag{9}$$

$$p(a, b \mid c) = \frac{p(a, b, c)}{p(c)} = \frac{p(a)p(b)p(c \mid a, b)}{p(c)} \tag{10}$$

$$a \not\perp\!\!\!\perp b \mid c \tag{11}$$

Fork Structure

$$p(a, b) = \sum_{c} p(a \mid c)p(b \mid c)p(c) \tag{12}$$

$$a \not\perp\!\!\!\perp b \mid c \tag{13}$$

$$p(a, b \mid c) = \frac{p(a, b, c)}{p(c)} = p(a \mid c)p(b \mid c) \tag{14}$$

$$a \perp\!\!\!\perp b \mid c \tag{15}$$

3 D-Separation

Consider a general directed graph in which A, B, and C are arbitrary nonintersecting sets of nodes. To evaluate whether $A \perp\!\!\!\perp B|C$, we consider all possible paths from any node in A to any node in B: blocked paths:

(a.) the arrows on the path meet either head-to-tail or tail-to-tail at the node, and the node is in the set C, or

(b.) the arrows meet head-to-head at the node, and neither the node, nor any of its descendants, is in the set C.

If all paths are blocked, then A is said to be d-separated from B by C, and the joint distribution over all of the variables in the graph will satisfy $A \perp\!\!\!\perp B|C$. In summary, observation (given exact value) will block the path.

4 Markov Random Fields

We can easily factorize directed graphs,

$$p(x) = \prod_{i=1}^{p} p\left(x_i \mid x_{prior^{(i)}}\right) \tag{16}$$

4.1 Conditional Independence

Global CI

$X_A \perp\!\!\!\perp X_C | X_B$

Local CI

$a \perp\!\!\!\perp$ {nodes - a - neighbour-a}|neighboar of a

Pair CI

$x_i \perp\!\!\!\perp x_j | x_i j (i \neq j)$

5 Inference in Graphical Models

References

[1] Judea Pearl, Madelyn Glymour, and Nicholas P Jewell. *Causal inference in statistics: A primer*. John Wiley & Sons, 2016.