8 Graphical Models

8.1 Bayesian Networks

Definition. ***Concepts***

1. ***Graphical Model*** *The joint distribution defined by the graph is given by the product, over all nodes of the graph, of a conditional distribution for each node conditioned on the variable corresponding to the parents of that node in the graph*

$$p(\mathbf{x}) = \prod_{k=1}^{K} p(x_k|parent(x_k))$$

   which represents ***factorization properties*** *of the joint distribution for a directed graphical model*

   (a) ***Node***

      i. ***Open Circle*** *Random variable*

         A. ***Shaded*** *Observed variable, i.e. variable $\{t_n\}$ from the training set, for setting the variable to some observed value*

         B. ***Unshaded*** *Latent variable*

      ii. ***Solid Circle (Dot)*** *Deterministic parameter*

   (b) ***Link*** *probabilistic relationships between these variables*

   (c) ***Plate*** *labelled with N indicate N nodes of a certain kind by drawing a single representative node and surround it with a box*

2. ***Bayesian Polynomial Regression***

   (a) ***Joint Distribution***

$$p(\mathbf{t}, \mathbf{w}) = p(\mathbf{w}) \prod_{n=1}^{N} p(t_n|\mathbf{w}) \qquad p(\mathbf{t}, \mathbf{w}|\mathbf{x}, \alpha, \sigma^2) = p(\mathbf{w}|\alpha) \prod_{n=1}^{N} p(t_n|\mathbf{w}, x_n, \sigma^2)$$

   (b) ***Posterior distribution***

   (c) ***Predictive distribution***

3. ***Generative Models***

4. ***Discrete Variable***

8.2 Conditional Independence

Definition. ***Conditional Independence***

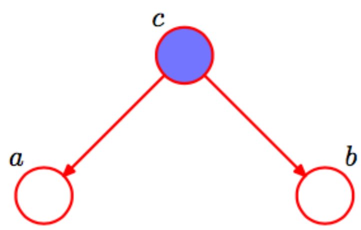
1. ***Goal*** *Infer conditional indepedence from graph structure*
2. ***Conditional Independence*** *a is conditionally independent of b given c, denoted as $a \perp\!\!\!\perp b \mid c$, if either is true*

$$p(a|b,c) = p(a|c) \qquad or \qquad p(a,b|c) = p(a|b,c)p(b|c) = p(a|c)p(b|c)$$

in other words, joint distribution of a and b factorizes into product of marginal distribution of a and marginal distribution of b

Definition. ***3 examples***

1. *For **tail-to-tail** node c, presence of path connecting a and b via a tail-to-tail node causes a, b be dependent, however the conditioned node 'blocks' the path from a to b and causes a and b to become conditionally independent*



Joint distribution can be derived from the graph

$$p(a,b,c) = p(a|c)p(b|c)p(c)$$

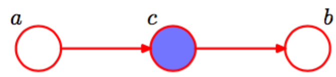
Note $a \not\!\perp\!\!\!\perp b \mid \emptyset$ (a, b are dependent) when c not observed

$$p(a,b) = \sum_c p(a,b,c) = \sum_c p(a|c)p(b|c)p(c) \overset{normally}{\neq} p(a)p(b)$$

But $a \perp\!\!\!\perp b \mid c$ (a, b are conditionally independent) when c is observed (conditioned)

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

2. *For **head-to-tail** node c, presence of path connecting a and b causes a, b to be dependent. If we observe c, this observation blocks the path from a to b so we obtain conditional independence*



Joint distribution can be derived from the graph

$$p(a,b,c) = p(a)p(c|a)p(b|c)$$

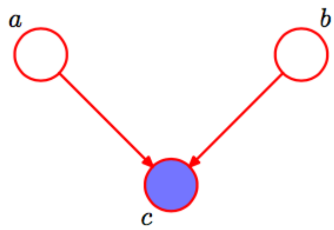
Note $a \not\perp\!\!\!\perp b \mid \emptyset$ *(a, b are dependent) when c not observed*

$$p(a,b) = \sum_c p(a,b,c) = p(a) \sum_c p(c|a)p(b|c) = p(a)p(b|a) \overset{normally}{\neq} p(a)p(b)$$

But $a \perp\!\!\!\perp b \mid c$ *(a, b are conditionally independent) when c is observed (conditioned)*

$$p(a,b|c) = \frac{p(a,b,c)}{p(c)} = \frac{p(a)p(c|a)}{p(c)} p(b|c) \overset{bayes}{=} p(a|c)p(b|c)$$

3. *For **head-to-head** node c. When c is unobserved, it blocks the path, and a, b independent. However conditioning on c unblocks the path and renders a, b dependent. In general, a head-to-head path will become unblocked if either the node, or any of its descendents, is **observed***



Joint distribution can be derived from the graph

$$p(a,b,c) = p(a)p(b)p(c|a,b)$$

Note $a \perp\!\!\!\perp b \mid \emptyset$ *(a, b are independent) when c not observed*

$$p(a,b) = \sum_c p(a,b,c) = p(a)p(b) \sum_c p(c|a,b) = p(a)p(b)$$

But $a \not\perp\!\!\!\perp b \mid c$ *(a, b are conditionally dependent) when c is observed (conditioned)*

$$p(a,b|c) = \frac{p(a,b,c)}{p(c)} = \frac{p(a)p(b)p(c|a,b)}{p(c)} \overset{normally}{\neq} p(a|c)p(b|c)$$

Summary *tail-to-tail or head-to-tail node leaves path unblocked unless it is observed in which case it blocks the path. By contrast, head-to-head node blocks a path if it is unobserved, but once the node, and/or at least one of its descendents, is observed the path becomes unblocked*

Definition. ***d-separation property*** *For directed graphs, where A, B, C are nonintersecting sets of nodes. We want to know if $A \perp\!\!\!\perp B \mid C$ is implied by the graphical model. We consider **all paths** from any node in A to any node in B. A path is **blocked** if it includes either*

1. ***head-to-tail*** *or* ***tail-to-tail*** *node in set* C *(observed)*

2. ***head-to-head*** *node and all of its descendent are not in set* C *(not observed)*

If all paths are blocked, then A *is said to be d-separate from* B *by* C*, and the joint distribution over all variables in the graph satisfies* $A \perp\!\!\!\perp B \mid C$*. We consider parameters of model as behaving in the same way as observed nodes, but they do not have parents, so all paths through these nodes will be tail-to-tail and hence blocked, so play no role in d-separation*

Definition. ***Markov Blanket*** *(derivation) of a node* $\mathbf{x}_i$ *is the set of nodes comprising the parents, the children and the co-parents. 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.*

$$p(\mathbf{x}_i|\mathbf{x}_{j\neq i}) = \frac{p(\mathbf{x}_1, \cdots, \mathbf{x}_D)}{\sum_{\mathbf{x}_i} p(\mathbf{x}_1, \cdots, \mathbf{x}_D)} = \frac{\prod_k p(\mathbf{x}_k|parent(\mathbf{x}_k))}{\sum_{\mathbf{x}_i} \prod_k p(\mathbf{x}_k|parent(\mathbf{x}_k))}$$

where terms $p(\mathbf{x}_k|parent(\mathbf{x}_k))$ *that doesnt involve* $\mathbf{x}_i$ *directly in the summation can be factored out and canceled with numerator. The terms that cannot be factored are either* $p(\mathbf{x}_i|parent(\mathbf{x}_i))$ *and* $p(\mathbf{x}_k|parent(\mathbf{x}_k))$ *where* $\mathbf{x}_i \in parent(\mathbf{x}_k)$ *Equivalently, the conditional distribution* $\mathbf{x}_i$ *conditioned on its non-descendents is dependent only on* $parent(\mathbf{x}_i)$