Bayesian Machine Learning - Lecture 6

Guido Sanguinetti

Institute for Adaptive and Neural Computation
School of Informatics
University of Edinburgh
gsanguin@inf.ed.ac.uk

March 2, 2015

Today's lecture

1. Structure and conditional independence
2. Directed Graphical Models
3. Undirected Graphical Models

Product rule - revisited

- Consider a joint probability distribution over random variables $x_1, \dots, x_N$ $p(x_1, \dots, x_N)$
- As probability distributions are symmetric, picking an ordering of the variables is arbitrary
- Given an arbitrary ordering, I can use the product rule to rewrite the joint as

$$p(x_1, \dots, x_N) = p(x_N | x_{N-1}, \dots, x_1) p(x_{N-1} | x_{N-2}, \dots, x_1) \dots p(x_1) \quad (1)$$

- Think of linking all random variables with segments and then stretching the resulting graph
- Useful to draw samples, but does not encode any deeper structure

Factorisations and graphs

- The product rule gives us a generic but rather useless factorisation of joint probabilities
- Interesting things arise when special factorisations are present, i.e. when you can remove some variables from the conditioning sets in (1)
- Today we will formalise what interesting means by mapping distributions onto graphs
- Graphs are also a very useful tool to model systems, as they abstract the dependence structure between variables without bothering with the details of the specific distributions

Conditional independence

- Two variables x_1 and x_2 are *conditionally independent* given x_3 $(x_1|x_2|x_3)$ if

$$p\left(x_1|x_2, x_3\right) = p\left(x_1|x_3\right).$$

- Equivalently

$$p(x_1, x_2|x_3) = p(x_1|x_3)p(x_2|x_3)$$

- Two non-overlapping sets of nodes A and B are *d-separated* by a set C if every node in A is conditionally independent of every other node in B given the nodes in C
- Given a variable (node) x_i, its *Markov blanket* is the set of nodes M that *d*-separate x_i from the rest of the network

DAGs and distributions

- Let us assume we have a factorisation of joint probability distribution in terms of conditionals
- A factorization defines a *Directed Acyclic Graph* (DAG) in the following way: variables are associated with nodes, and variables to the right of the conditioning sign (parents) are linked with an arrow to variables to the left (children)
- Warning: the graph is not unique (*Markov equivalence*)!!!
- Cool thing, start with a DAG and get a distribution with structure

Graphical notation

- Random variables are denoted by circles with the name inside
- Observed random variables are often shaded
- I.i. d. variables are denoted through plates
- Sometimes, additional parameters which we are not doing posterior inference over are denoted as squares
- Let's draw the graphical model for the Gaussian Mixture Model

Francesco's example

Francesco works on modelling users from twitter data. Let's draw a graphical model for him!

Working out conditional independences from graphs

- Graphically, two variables a and b are d-separated by C if every path between a and b is blocked
- A path is blocked if the arrows meet head to tail or tail to tail at a node in C
- A path is blocked also if there exists a node x on the path where arrows meet head to head, and neither x nor its descendants are in C
- The Markov blanket of a node are its children, parents and co-parents
- Notice that not all conditional independence can be represented by a DAG (rather pathological counter examples)

Dependencies with no direction

- Sometimes we wish to encode dependencies which do not have an obvious precedence
- Discuss some examples
- In these cases, we just draw lines between nodes without arrows
- Hybrid examples (some directed edges, some undirected) also possible
- One particular example from my work (Sanguinetti et al, Bioinformatics 2008)

Conditional independence and Markov Random Fields

- Formally, we can associate an undirected graphical model (Markov Random Field) to a general factorisation of a joint distribution (not necessarily in terms of conditional distributions)
- Nodes are connected if they appear together in a factor
- Maximal cliques in the graph correspond to factors
- Conditional independence is equivalent to separation in the graph (easy!)

Simple example

- Consider the linear dynamical system

$$\begin{aligned} x_t &= a x_{t-1} + \epsilon_t \\ \epsilon_t &\sim \mathcal{N}(0, \sigma^2) \qquad x_0 \sim \mathcal{N}(0, 1) \end{aligned} \tag{2}$$

- Write the joint distribution
- Draw the corresponding directed and undirected graphical models

From directed to undirected graphs

- In general, it is possible to go from a directed to an undirected graphical model (why?)
- Remove all arrows
- Join all nodes that share a child (moralisation)
- Are conditional independences preserved?