Probabilistic Machine Learning
Lecture 8 – Graphical models

Lawrence Murray, Uppsala University
2018-05-15

Outline – Lecture 8

Aim: To understand the language of graphical models for representing probabilistic models.

1. Languages for probabilistic models
2. Directed graphical models
3. Undirected graphical models
4. Factor graphs
5. Converting between the different types.

Languages for probabilistic models

There are many ways of expressing probabilistic models, and each offers a different, and useful, perspective.

1. Mathematical
   Provides the equations that we need for a precise understanding of a model.

2. Graphical
   Visualises the structure of a model for quick communication and high-level calculations.

3. Programmatical
   Provides the means of performing the computations necessary.

Example: Polynomial regression

• Let \( t_{1:N} \) be the values of a function at the points \( x_{1:N} \).
• Find the \( M \)th degree polynomial approximating the function, with coefficients \( w \in \mathbb{R}^{M+1} \).
Example: Polynomial regression

For $n = 1, \ldots, N$, and $\phi(x) = (1, x, x^2, \ldots, x^M)^T$:

$$t_n = w^T \phi(x_n) + v_n \quad v_n \sim \mathcal{N}(0, \sigma^2),$$

with

$$w \sim \mathcal{N}(0, \Sigma).$$

The joint distribution can be written:

$$p(t_{1:N}, w) = p(t_{1:N} | w)p(w) = p(w) \prod_{n=1}^{N} p(t_n | w).$$

What is the reason for this last equality?

Types of graphical models

1. **Directed graphs** (a.k.a. Bayesian networks) represent a set of random variables and their conditional dependence structure.
2. **Undirected graphs** (a.k.a. Markov random fields) represents a set of random variables and their Markov structure.
3. **Factor graphs** are a more convenient form obtained from the above two for the purposes of inference and learning.

Common elements

- These conditional independencies are usually what we refer to when talking about the **structure** of a model.
- Graphical models are visual languages that draw particular focus to this structure over other aspects of a model.

The different types have common elements:

- Nodes represent random variables.
- Filled nodes represent observed random variables.
- Empty nodes represent unobserved random variables.
- Edges represent relationships between random variables.
The space of all probabilistic models is $P$.
The directed ($D$) and undirected ($U$) classes of graphical models intersect.
Probabilistic programs attempt to fill the whole space $P$.

A directed graphical model represents the decomposition of a joint distribution into a product of conditional distributions.
Edges are directed from a parent node to a child node and represent conditional dependencies.
It is the absence of an edge that conveys interesting information regarding conditional independencies.

In directed graphical models, factorisation of the joint distribution into conditional probability distributions is clear.
Directed graphical models and conditional independence

- In the above graph, is $a$ conditionally independent of $b$ given $c$? (We write this $a \perp b \mid c$.)
- To answer this, we will need a concept called d-separation.
- This is the difficult part with directed models: factors are clear, but conditional independencies not so much.

Tail-to-tail nodes

- Is $a \perp b \mid c$? i.e. does $p(a, b \mid c) = p(a \mid c)p(b \mid c)$?
- Is $a \perp b$? i.e. does $p(a, b) = p(a)p(b)$?

Rule for tail-to-tail nodes
For conditional independence of two nodes, the tail-to-tail nodes between them must be observed.

Head-to-tail nodes

- Is $a \perp b \mid c$? i.e. does $p(a, b \mid c) = p(a \mid c)p(b \mid c)$?
- Is $a \perp b$? i.e. does $p(a, b) = p(a)p(b)$?

Rule for head-to-tail nodes
For conditional independence of two nodes, the head-to-tail nodes between them must be observed.

Head-to-head nodes

- Is $a \perp b \mid c$? i.e. does $p(a, b \mid c) = p(a \mid c)p(b \mid c)$?
- Is $a \perp b$? i.e. does $p(a, b) = p(a)p(b)$?

Rule for head-to-head nodes
For conditional independence of two nodes, the head-to-head nodes between them must be unobserved.
Definition: d-separation

Consider a directed acyclic graph in which $A$, $B$ and $C$ are arbitrary non-intersecting sets of nodes. We have the property

$$A \perp B \mid C$$

if, on all possible paths from any node in $A$ to any node in $B$:

- all tail-to-tail and head-to-tail nodes are in $C$, and
- neither head-to-head nodes nor any of their descendants are in $C$.

Examples: d-separation

- The path from $a$ to $b$ is not blocked by $f$, since it is a tail-to-tail node and $f$ is unobserved.
- Nor is it blocked by $e$, which is a head-to-head node, and has an descendant $c$ which is observed.
- Hence, $a \perp b \mid c$ does not follow from this graph.

- The path from $a$ to $b$ is blocked by $f$, since it is a tail-to-tail node and $f$ is observed.
- It is also blocked by $e$, since it is head-to-head node and neither it nor any of its descendants are observed.
- Hence, $a \perp b \mid c$ does follow from this graph.

Definition: Markov blanket for directed graphical models

In a directed graphical model, a node is conditionally independent of all other nodes given its parents, its children, and its co-parents.

(Co-parents are the other parents of its children.)

Undirected Graphical Models
Definition: Undirected Graphical Model

- Undirected models are often used in imaging and spatial applications.
- Nodes are random variables.
- Conditional independence is straightforward, no d-separation!
- That’s the easy bit: now the factors are tricky.

Definition: Markov blanket for undirected graphical models

In an undirected graphical model, a node is conditionally independent of all other nodes given its neighbours.

Undirected graphical models and conditional independence

If $A$, $B$, and $C$ are arbitrary non-intersecting sets of nodes, we have $A \perp \perp B \mid C$ if all paths between nodes in $A$ and nodes in $B$ are blocked by nodes in $C$.

Factors of an undirected model

- The joint distribution is a product of potential functions over the maximal cliques of the graph.
- A clique is a subset of the nodes of the graph that are fully connected (i.e. edges between all pairs).
- A maximal clique is a clique for which it is not possible to add any additional nodes without it no longer being a clique.
Factors of an undirected model

- \{x_1, x_2\} is a clique (there are several others).
- \{x_2, x_3, x_4\} is a maximal clique (there is one other).

Example: image de-noising

Suppose we have a noisy image and want to remove the noise.

Choose the potential functions:

\[
\psi_y(x_{ij}, y_{ij}) = \exp\left(-\frac{1}{\beta}(y_{ij} - x_{ij})^2\right)
\]
\[
\psi_x(x_{i_1j_1}, x_{i_2j_2}) = \exp\left(-\min\left(\frac{1}{\alpha^2}(x_{i_1j_1} - x_{i_2j_2})^2, \gamma\right)\right)
\]

Example: image de-noising

Let the true pixel values be \(x_{ij}\) and the observed pixel values be \(y_{ij}\).

The joint distribution is given by:

\[
p(x) = \frac{1}{Z} \prod C \psi_C(x_C),
\]

where \(C\) indexes the maximal cliques, and \(Z\) is the normalizing constant (a.k.a. partition function):

\[
Z = \sum_x \prod C \psi_C(x_C).
\]

- If we restrict the potential functions to be non-negative, then \(p(x) \geq 0\).
- But it is possible that \(Z\) is not finite, in which case the graph does not represent a probability distribution at all.
Example: Road surface estimation

Estimate road surface using images from a stereo camera.


Factor Graphs and Converting Between Types

Directed graphical model $\rightarrow$ undirected graphical model

You can’t just drop the arrow heads!

Although in the special case of tree-structured graphs, you can: each node has at most one parent, so there are no co-parents to join anyway.
Undirected graphical model $\implies$ directed graphical model

- No one actually does this.
- There is a deeper reason for this. Conversions between graphical model types are not one-to-one, and may discard information.
- If your factors are conditional probability distributions, you would use a directed graphical model in the first place. Undirected graphical models are usually used in situations where you can’t do this.

Directed graphical model $\implies$ factor graph

1. Create a variable node for each node in the original graph.
2. Create a factor node for each node in the original graph, where this factor expresses its conditional probability distribution.

Factor graphs

$p(x_{1:3}) = f_a(x_1, x_2)f_b(x_1, x_2)f_c(x_2, x_3)f_d(x_3)$

- We can convert both directed and undirected graphical models to factor graphs.
- These have both variable nodes and factor nodes, which form a bipartite graph.
- The motivation is to make the factors more explicit, and facilitate inference algorithms based on message passing.
A few concepts to summarize lecture 8

**Three types of graphical model:** These are directed graphical models, undirected graphical models, and factor graphs. Conversions between them are not one-to-one.

**Graphical models draw focus to structure:** By structure, we mean the conditional independencies of a probabilistic model.

**D-separation** is used to determine conditional independencies in a directed graphical model.

**Markov blanket** is easier to remember, but less general.

**Next lecture:** Inference algorithms for graphical models, and probabilistic programming.