Graphical Models

Dmitrij Lagutin, dlagutin@cc.hut.fi

T-61.6020 - Machine Learning: Basic Principles
12.3.2007
Contents

- Introduction to graphical models
- Bayesian networks
- Conditional independence
- Markov random fields
- Inference in graphical models
Introduction to graphical models

• This presentation concentrates on *probabilistic graphical models*, they have several benefits
  – They provide a simple, clear way to visualize the probabilistic model, this helps e.g. designing new models
  – Different properties of the probabilistic model, like conditional independence can be seen from the graphical model
  – Complex computations in sophisticated models can be expressed in terms of graphical manipulations
Introduction to graphical models

- A graph consists of nodes (vertices) that are connected by edges (links, arcs)
- The graph can be directed (edges have arrows to indicate the direction) or undirected (edges do not have arrows)
- In probabilistic graphical models each node in a graph represents a random variable and the edges of the graph represent probabilistic relationships between these variables
Bayesian networks

• Also known as directed graphical models
• Each node is associated with a random variable or with a group of variables
• Edges describe conditional relationships between variables
  – For example, if the distribution is $p(b|a)$ then the corresponding graphical model will have a directed edge from $a$ to $b$
Bayesian networks: example

• Let the distribution be:
  \[ p(x_1) p(x_2) p(x_3) p(x_4 \mid x_1, x_2, x_3) p(x_5 \mid x_1, x_3) p(x_6 \mid x_4) p(x_7 \mid x_4, x_5) \]

• Then the corresponding graphical model is:
Bayesian networks: polynomial regression

• Lets build a graphical model for polynomial regression

• The polynomial regression model contains
  – Vector of polynomial coefficients \( \mathbf{w} \)
  – Observed data \( \mathbf{t} = (t_1, \ldots, t_N)^T \)
  – Input data \( \mathbf{x} = (x_1, \ldots, x_N)^T \)
  – Noise variance \( \sigma^2 \)
  – Hyperparameter \( \alpha \)
Bayesian networks: polynomial regression

• Lets concentrate first on random variables, their distribution can be expressed as

\[ p(t, w) = p(w) \prod_{n=1}^{N} p(t_n \mid w) \]

• The corresponding graphical model (the box denotes that there are \( N \) values inside) is:
Bayesian networks: polynomial regression

- Machine learning problems usually have a training set
  - In graphical models these observed variables are denoted by shading
- Sometimes it is better to show parameters of a model explicitly
  - In graphical models deterministic parameters are denoted by small, solid circles (variables are denoted by larger, open circles)
Bayesian networks: polynomial regression

• With parameters the distribution becomes:

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

• And the graphical model:
Bayesian networks: polynomial regression

• The aim of the polynomial regression model is: given a new input value $\hat{x}$ find the corresponding probability distribution $\hat{t}$ conditioned on the observed data

• Thus, the final form of the model is:

$$p(\hat{t}, t, w | \hat{x}, x, \alpha, \sigma^2) = \prod_{n=1}^{N} p(t_n | w, x_n, \sigma^2) p(w | \alpha) p(\hat{t} | \hat{x}, w, \sigma^2)$$
Bayesian networks: polynomial regression

• The corresponding final graphical model is:
Conditional independence

• If $p(a|b,c) = p(a|c)$ then $a$ is conditionally independent of $b$ given $c$: $a \perp b | c$

• We can write above as:

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

  – Now the joint distribution factorises into the product of marginal distribution of $a$ and $b$, both of these marginal distributions are conditioned on $c$
Conditional independence: d-separation

• D-separation can be used to determine whether or not there exist a conditional independence in the underlying model

• Some concepts
  – Node c is a tail-to-tail node:

  – Node c is a head-to-tail node:
Conditional independence: d-separation

- Node c is a head-to-head node:

- Node y is a descendant of node x if there is a path from x to y in which each step of the path follows the directions of the arrows
Conditional independence: d-separation

• Let A, B and C be nonintersecting sets of nodes
• Let's consider all possible paths from any node in A to any node in B, the path is blocked if it includes a node such that either
  – 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
  – 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 d-separated from B by C and thus condition $A \perp \perp B \mid C$ is satisfied
Conditional independence: example

- Node $f$ does not block the path between $a$ and $b$ because it is a tail-to-tail node and is not observed.
- Node $e$ does not block the path because it has a descendant $c$.
- Thus, $a$ and $b$ are not conditionally independent given $c$. 
Conditional independence: second example

• Node $f$ blocks the path between $a$ and $b$ because it is a tail-to-tail node that is observed, thus there exist a conditional independence:

\[ a \perp b \mid f \]

• In this case, also node $e$ blocks the path, because $e$ does not have descendants in the conditioned set
Markov random fields

• Also known as Markov networks or undirected graphical models
• In Markov random fields each node corresponds to a variable or to a group of variables just like in the Bayesian networks
• However, edges between nodes are undirected in Markov random fields
Markov random fields: conditional independence

- Let A, B and C be nonintersecting sets of nodes.
- Let’s consider all possible paths from any node in A to any node in B, the path is blocked if it includes a node from C.
- If all such paths are blocked, then A and B are conditionally independent given C: $A \perp B \mid C$.
Markov random fields: conditional independence

- In this example A and B are conditionally independent because all paths between A and B go through C.
Inference in graphical models

- Lets look at the Bayes’ theorem’s rule: \( p(x, y) = p(x)p(y|x) \), this rule can be represented by a graph (a) on the next slide.
- If the \( y \) is observed (graph (b)), then it is possible to infer the corresponding posterior distribution of \( x \): 
  \[
  p(y) = \sum_{x'} p(y | x') p(x')
  \]
  \[
  p(x | y) = \frac{p(y | x) p(x)}{p(y)}
  \]
Inference in graphical models

• Now the joint distribution $p(x,y)$ can be expressed in terms of $p(y)$ and $p(x|y)$ as graph (c) shows: $p(x,y) = p(y)p(x|y)$

• This is a simple example of inference problem in graphical models
Inference in graphical models: factor graphs

• Graphical models allow a function of several variables to be expressed as a product of factors over subsets of those variables.

• Factor graphs make this separation explicit by introducing additional nodes for factors.

• Let's write the joint distribution in the form of a product of factors ($x_s$ denotes a subset of variables):

$$ p(x) = \prod_s f_s(x_s) $$
Inference in graphical models: factor graphs, example

• Let the distribution be
  \[ p(x) = f_a(x_1, x_2) f_b(x_1, x_2) f_c(x_2, x_3) f_d(x_3) \]
• The corresponding factor graph is
Inference in graphical models: factor graphs

• Factor graphs are bipartite graphs, because they consist of two different kinds of nodes, and all edges go between nodes of opposite type

• Factor graphs are useful for building more complex models that are not covered here