Bayesian Inference

Viktor

March 18, 2018

1 Graph theory

We recap the basics of graph theory. [2, 1]

Graph. A graph $G = (V, E)$ consistes of a set $V = \{1, \dots, m\}$ of vertices and a set $E \subseteq V \times V$ of edges. So each edge consists of a pair of vertices $(s, t) \in E$. For an *undirected* graph there is no distinction between (s, t) and (t, s). In a *directed* graph the edge orientation is distinguished, which is emphasised by the notation $(s \to t) := (s, t)$.

Remark: By definition a graph does not contain self-loops as edges $(s, s) \notin E,\ \forall s \in V$ nor does it contain multiple copies of the same vertex. This may be included within the framework of multigraphs.

Subgraph. A subgraph $G' = (V', E')$ of a given graph $G = (V, E)$ is a graph such that $V' \subseteq V$ and $E' \subseteq E$. Given a vertex subset $V' \subseteq V$ of $G = (V, E)$, the *vertex-induced subgraph* $G'(V') = (V', E(V'))$ has the edge set $E(V') = \{(s, t) \in E | s, t \in V'\}$. Given an edge subset $E' \subseteq E$ of $G = (V, E)$, the *edge-induced subgraph* $G'(E') = (V(E'), E')$ has vertex set $V(E') = \{s \in V | (s, t) \in E'\}$

Path. A path $P = (V(P), E(P))$ is a graph with vertex set $V(P) = \{v_0, \dots v_k\}$ and edge set $V(P) = \{(v_0, v_1), (v_1, v_2) \dots (v_{k-1}, v_k)\}$. The path joins vertex v_0 to vertex v_k. Of special interest are paths that are subgraphs of a given Graph $G = (V, E) :\ V(P) \subseteq V,\ E(P) \subseteq E$.

Cycle. A cylcle $C = (V(C), E(C))$ is a graph with vertex set $V(C) = \{v_0, \dots v_k\}$ and edge set $E(C) = \{(v_0, v_1), (v_1, v_2) \dots (v_{k-1}, v_k), \dots (v_k, v_0)\}$. An undirected graph is acyclic if it contains no cycles.

Bipartite. A graph $G = (V, E)$ is bipartite if its vertex set can be partitioned as a disjoint union $V = V_a \dot{\cup} V_b$ such that $(s, t) \in E \Rightarrow s \in V_a, t \in V_b$ (or vice versa).

Clique. A clique of a graph $G = (V, E)$ is a subset of vertices $V' \subseteq V$ that are all joined by vertices, $(s, t) \in E\ \forall s, t \in V'$. A clique V' is maximal if there is no vertex $v \in V \setminus V'$ such that $V \cup \{v\}$ is a clique.

Remark: Sometimes maximal cliques are just called cliques and non-maximal cliques are calle cliquos.

Chord. Given a cylce C with vertex set $V(C) = \{v_0, \ldots v_k\}$ and edge set $E(C) = \{(v_0, v_1), (v_1, v_2) \ldots (v_{k-1}, v_k), \ldots (v_k, v_0)\}$. A chord is an edge that is not part of $E(C)$. Given a Graph $G = (V, E)$ and a cycle C of length four or greater. C is *chordless* if the edge set E of G contains no chords for C.

Triangulated. A graph is triangulated if it contains no chordless cycles (of length four or greater).

Connected component. A connected component of a graph $G = (V, E)$ is a subset of vertices $V' \subseteq V$ such that $\forall\ s, t \in V'$, there exists a path in G joining s to t. A graph is *singlyconnected* if it consists of a single connected component.

Tree. A tree is an acyclic singly connected graph. It can be shown that a tree with m verices must have $m - 1$ edges.

Forest. A forest is an acyclic graph consisting of one or more connected components.

Hypergraph. A hypergraph is $G = (V, E)$ consists of a vertex set $V = \{1, 2, ..., m\}$, and a set E of hyperedges, with $E \ni h \subset V$. So, each hyperedge is a particular subset of V. In particular, an ordinary graph is a hypergraph with $|h| = 2$.

Factor graph. Given a hypergraph $G = (V, E)$. A Factor graph is a bipartite graph $F = (V', E')$ with $V' = V \cup E$ and $E' = \{(s, h) \in V \times E \mid s \in h\}$.

2 Probability distributions on graphs

In order to define a probabilistic graphical model over a graph $G = (V, E)$, each *single vertex* $s \in V$ is associated with a random variabel X_s. The state space of X_s is denoted by $\mathcal{X}_s$ It defines all possible values X_s may take. For example, in the contineous case $\mathcal{X}_s \subseteq \mathbb{R}$ and in the discrete case $\mathcal{X}_s = \{1, \ldots, r\}$. Lower case letters are used to denote an element of the state space, $x_s \in \mathcal{X}_s$. The notation $\{X_s = x_s\}$ corresponds to an event where the random variable X_s takes the specific value $x_s \in \mathcal{X}_s$ over its state space $\mathcal{X}_s$.

For a *subset of vertices* $A \subset V$, the above notations may be generalized: $X_A := (X_s, s \in A)$. The joint state space is the cartesian product of the individual state spaces, $\mathcal{X}_A := \otimes_{s \in A} \mathcal{X}_s$. A state space element $x_A \in \mathcal{X}_A$ is then given by $x_A = (x_s, s \in A)$. This can be also interpreted as the marginal pdf of $|A|$ random variables with respect to a joint pdf of $|V|$ random variables.

2.1 Directed graphical models

Given a directed graph $G = (V, E)$. For an edge $(s \rightarrow t)$ s is called parent of t and s is called child of t. For a given vertex $s \in V$ denote the set of its parents $\pi(s) = \{t \in V | s \rightarrow s\}$.

- DAG
- interpretation
- not unique (bayes thm)
- plate notation
- model building
- conditional independence

Undirected graphical models

- Markov random fields
- interpretation
- conditional independence
- Moralization

2.2 Factor graphs

Junction tree

-

References

[1] Bela Bollobas. *Modern Graph Theory*. Springer, 1998.

[2] Martin J. Wainwright and Michael I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1–2):1–305, 2008.