

CS281A/Stat241A recitation 3: towards the junction tree algorithm

September 17, 2007

Percy Liang

Review

- We have a joint distribution specified by a graph. To compute marginals (conditionals are just ratios of marginals), we have to sum out variables. The techniques we have been developing provide efficient ways of doing the summation.
- In increasing order of sophistication
 - Brute-force summation
 - Elimination algorithm: works on any graph, but only get one marginal distribution
 - Sum-product algorithm (works on trees): get all the node marginal distributions
 - Junction tree algorithm (works on arbitrary graphs, but need some preprocessing)
- If have a directed graph, computing marginals upstream in the graph is easy, but conditioned on nodes downstream, the marginals become non-trivial to compute

Steps for applying the junction tree algorithm (Hugin's algorithm)

1. Moralization of directed graphs:
 - Convert directed graph into to undirected graph
 - Treat conditional probabilities as arbitrary potentials
 - Absorb non-maximal clique potentials into maximal cliques
2. Introduction of evidence:
 - Take slices of potentials
 - Treat observed nodes as constants
 - Rewrite potentials as distribution over non-observed nodes
3. Triangulation of undirected graphs
 - Triangulated graph: every cycle of length 4 has a chord
 - Triangulation: add edges to a graph to make it triangulated
 - Example: need to add $n - 3$ edges to triangulate a n -node cycle

- Finding best triangulation is NP-hard
- Elimination algorithm implicitly does triangulation (via creation of elimination cliques)
- Tree-width:
 - For triangulated graphs, is size of largest clique minus one
 - For general graphs, is minimum tree-width over all possible triangulations
- Finding best triangulation (tree-width) is NP-hard

4. Construction of junction tree

- Clique tree
 - Nodes: cliques
 - Edges: separator sets
- Junction tree property (equivalent characterizations)
 - For every pair of cliques V, W , all cliques on the (unique) path contain $V \cap W$.
 - Continuity: for every node, the set of cliques containing it are connected
 - Decomposition: for every clique, no variable not in the clique appears on different “sides” of the clique in the clique tree
 - Junction tree: a clique tree satisfying the junction property
- Use maximum spanning tree algorithm with size of clique intersections separators as edge weights:
 - Triangulated graph \Rightarrow clique tree

5. Junction tree algorithm

- Pass messages to compute marginals
- Input: a junction tree
- Output: marginal probabilities on cliques
- Special case: elimination algorithm