6.892 Machine learning seminar

Prof. Tommi S. Jaakkola
tommi@ai.mit.edu

February 18, 1999

Decimation

Decimation as an inference algorithm is restricted to a particular subclass of undirected graphs called decimatable graphs. For simplicity we will assume that the graph is connected (there is a path between any pair of nodes in the graph). If it is not, then each of the connected components can be handled separately.

Definition 1 *An undirected graph G is decimatable if it can be triangulated in such a way that the maximal clique size is at most 3 in the triangulated graph.*

Trees and chains are decimatable graphs, for example.

Consider now a decimatable connected graph G with a vertex set V and an edge set E. Note that the above definition does not require that this graph is triangulated. Let P be any probability model consistent with this graph and consisting of only pairwise potentials, i.e.,

$$P(x_1, \ldots, x_n) = \prod_{(i,j) \in E} \phi_{ij}(x_i, x_j) \tag{1}$$

where the product contains a potential function for each edge in the graph. Assuming that the number of values the variables x_i can take is finite, then this probability model can be decimated as described below.

The decimation algorithm

1. Find a node j in the graph with at most two neighbors. Such a node must exist if the graph is decimatable. Any such node will do.

2. Let $nb(j)$ be the set of neighbors of j. Compute

$$\tilde{\phi}(x_{nb(j)}) = \sum_{x_j} \prod_{i \in nb(j)} \phi_{ij}(x_i, x_j) \tag{2}$$

where the product can have at most two terms (but at least one since we assumed that the graph is connected).

3. Connect the neighbors of j and remove j and its associated edges from the graph.

4. If the neighbors of j were already connected, update one of the potential functions $\phi_{ik}(x_i, x_k)$ for which $nb(j) \subseteq \{i, k\}$ according to

$$\phi_{ik}(x_i, x_k) \leftarrow \phi_{ik}(x_i, x_k)\tilde{\phi}(x_{nb(j)}) \tag{3}$$

for all values of x_i and x_k. Note that $nb(j)$ can be $\{i\}$, $\{k\}$ or $\{i, k\}$.

If the neighbors of j were not connected then introduce $\tilde{\phi}(x_{nb(j)})$ as the new pairwise potential corresponding to the added edge.

5. Repeat until the graph is over only the variables of interest.

After each round in this iterative method, the joint distribution over the remaining nodes in the graph is the correct marginal distribution where the marginalization is performed over the eliminated variables. The marginal distribution has the same form as the original distribution, i.e., consists of pairwise marginals.

Some justification

Here we show that decimatable graphs remain decimatable if we eliminate any node from the graph with at most two neighbors by first connecting the neighbors and then eliminating the node and its associated edges. The possible cases are (for connected graphs):

1. The node j has only one neighbor. Eliminating such a node does not introduce any further edges and cannot change whether the remaining graph is decimatable.

2. The node j has two neighbors which are already connected. Again, no further edges are added when removing the node.

3. The node j has two neighbors which do not have an edge between them. By definition there must be an ordering so that the nodes can be eliminated (which includes

connecting their neighbors) without introducing cliques of size larger than 3. Eliminating j would violate this ordering only if either or both of its neighbors, say i and k, should be eliminated before j. Any neighbor of j, say k, can be eliminated only when it has at most one additional neighbor (besides j); otherwise the elimination would create a clique of size at least 4. So when k can be eliminated the graph looks like

```
      i   j   k
...--O--O--O--O--...
```

where the symmetry guarantees that there can be no difference between having eliminated j earlier or eliminating k now. Thus the elimination ordering is irrelevant for decimatable graphs.