6.892 Machine learning seminar

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

February 16, 1999

This assignment is due **February 22**. You can either email the assignment to me as plain text (use subject "HOMEWORK 2") or give it to me in class (the latter may be easier in this case).

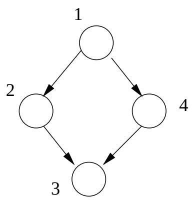
Consider a directed graph G.

- Let $An(i)$ denote the minimal ancestral set of node i. Recall that $An(i)$ must contain i, the parents of i, the parents of the parents of i and so on. For sets of nodes

$$An(A) = \cup_{i \in A} An(i) \tag{1}$$

- Suppose now that we have three disjoint sets of nodes A, B and S, where A and B are non-empty. We say that A and B are *D-separated* by S if A and B are separated by S in $\mathcal{M}(G_{An(A \cup B \cup S)})$, the moral graph of the subgraph induced by the ancestral sets of A, B, and S.
- We use $A \perp_G B | S$ to denote a graph separation; for directed graphs this refers to D-separation while for undirected graphs it is the ordinary graph separation.

---

1. 

   a) List all the D-separation properties that hold for this graph.

   b) Explain how/if you were able to avoid exhaustive search.

2. a) Prove that for any directed graph G,

$$A \perp_{\mathcal{M}(G)} B | S \Rightarrow A \perp_G B | S \tag{2}$$

where A, B, and S are arbitrary disjoint sets of nodes so long as A and B are non-empty. In other words, show that any separation property in the moral graph of a directed graph implies the corresponding D-separation property for the underlying directed graph.

b) Moralization is used to transform directed graphs into undirected graphs. Interpret the above property in terms of probability models underlying the graphs.

c) What conditions should a directed graph satisfy so that the converse, i.e.,

$$A \perp_{\mathcal{M}(G)} B | S \Leftarrow A \perp_G B | S, \tag{3}$$

also holds? Your answer need not be complete. Simply explain your reasoning.

3. **(Semi-)research level:** Consider a simple Markov model represented as an undirected graphical model:

```
... -O--O--O--O- ...
```

where the nodes correspond to state variables s_t. We can extend this model by allowing at each time step an input x_t that can influence the value of the state s_t. The graph would be

```
... -O--O--O--O- ...
     |  |  |  |
     O  O  O  O
```

We would like to further extend this model. Suppose the input variables x_t no longer simply bias the states s_t but they also affect which other state variables s_t can depend on. The "indexing" of the variables s_t in this setting is no longer effectively relative to time but relative to the inputs x_t. For example, s_t and $s_{t'}$ would be connected in the graph if the corresponding x_t and $x_{t'}$ were close. The model is dynamic in the sense that the graph itself changes in response to the inputs. Assume s_t are discrete variables and that the connectivity is e.g. a function of the distance between the inputs. Explicate such a process and the associated inference calculations. Where could you use this process? How about a similar extension to HMMs?

Note: if the state variables are Gaussian and the inputs define the covariance between the state variables rather than dependence, then the resulting process is known as a Gaussian process.