6.867 Machine learning and neural networks

Problem set 5

Deadline: when you get it done... do not return

Problem 1: Medical diagnosis

In figure 1 we have two diseases d_1 and d_2 as well as two potential findings f_1 and f_2. The diseases and findings are assumed to be binary. We assume further that the conditional probabilities of $P(f_1|d_1, d_2)$ and $P(f_2|d_2)$ are noisy-OR models as described in the lectures.

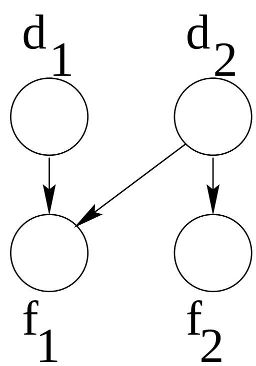


Figure 1: A Bayesian network for medical diagnosis. We assume that the conditional probabilities are noisy-OR models.

1. Now, suppose we know that the outcome of finding f_1 was positive ($f_1 = 1$). What happens to the disease probabilities? Justify your answer.

2. Suppose later on we learn that finding f_2 was positive as well. How will the probability of disease d_1 being present change as a result? Justify your answer.

3. Construct a junction tree for the graph model in figure 1. Also initialiaze the potential functions for the junction tree. Are there many ways of initializing the potentials?

4. Roughly speaking, how many operations do we need to perform to complete both "collect" and "distribute" steps? You can assume here that we will blindly apply both propagation steps whatever the evidence might be.

5. Suppose we would like to determine which finding to query (which of the corresponding tests to carry out). For this we need to evaluate the mutual information between the disease configurations and the possible values of each of the findings. In other words, we have to compute $I(f_1; d_1, d_2)$ and $I(f_2; d_1, d_2)$, where, for example,

$$I(f_1; d_1, d_2) = \sum_{d_1, d_2, f_1 = 0,1} P(d_1, d_2, f_1) \log \frac{P(d_1, d_2, f_1)}{P(d_1, d_2) P(f_1)} \tag{1}$$

Now, show that the graph structure (original or the junction tree) implies that $I(f_2; d_1, d_2) = I(f_2; d_2)$.

6. Based on your results for 3) and 5) show that the marginal probabilities that we compute in the junction tree suffice for evaluating which test we should query next ("active learning").

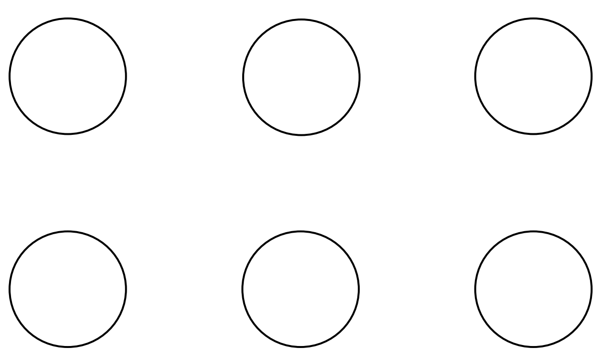
Problem 2 (from last year's final)

Your task here is to identify the relevant variables and the graph structure that captures the following (imaginary) setting. There may be multiple "correct" answers.

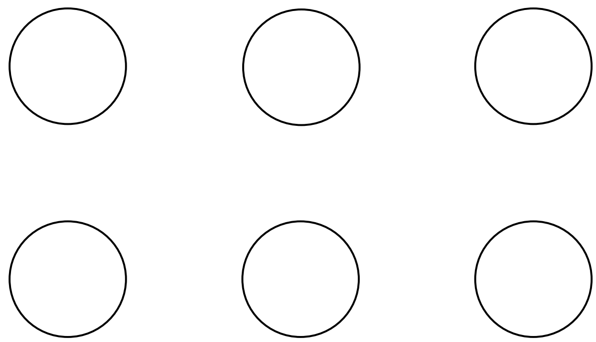
"A panel of three judges determines the outcome of presidential elections. Each judge can vote for one of the two possible candidates and the outcome is obtained by a majority rule. Two of the judges are impartial in the sense that they will listen to arguments from two spokespersons each working for one of the candidates while the remaining judge consistently pays attention to only one of the spokespersons. Each spokesperson will ask a judge to vote for a specific candidate. The spokespersons never talk nor listen to each other directly."

1. Identify the relevant variables based on the above description. For each variable state the possible values that it can take. If you use abbreviations to identify the variables make sure they are not ambiguous.

2. Draw a Bayesian network that captures the interactions between the variables. *Avoid any assumptions that you cannot make on the basis of the above description.* Please indicate which variables correspond to which nodes.



3. The graph might change if the above description had started with "A panel of three *independent* judges...". If the graph would change, please draw the new graph. Otherwise state that there are no changes.



4. Explain under what circumstances (setting of some of the variables etc.) we might observe "explaining away" in the graph you just drew. If none exists, briefly explain why not.

   (For your convience, here's a brief description of "explaining away": *When we have multiple possible causes for a single known effect, explaining away refers to the phenomenon where acquiring further evidence about the presence of one of the causes makes the other ones less likely.*)

5. **(T/F – 2 points)** The graph structure is useful *only if* it captures all the independence properties present in the underlying probability distribution

6. **(T/F – 2 points)** Given any probability distribution, we can find a Bayesian network as well as a Markov random field that is consistent with the distribution

7. **(T/F – 2 points)** A Boltzmann machine where *all* the variables are observable can *only* capture second order statistics (means and covariances) between the variables