CS717: Midsem

30 Marks. No doubts please! There should be no need for clarifications or any additional assumptions. But if there is any need for additional assumptions, please make them and state them explicitly in your answer sheets.

1. **Maximum Weighted Clause Satisfiability and Graphical Models:** Recall the maximum weighted clause satisfiability problem (MWCSP): Given a set of clauses and their associated weights, $\{(C_1, w_1), (C_2, w_2), \ldots, (C_n, w_n)\}$, the problem is to determine the set of true propositions that maximize the sum of the weights of the clauses satisfied by setting those propositions to true. For this problem, we only discussed a heuristic called Maxwalksat which was not guaranteed to yield the optimal result. Now we want to solve MWCSP optimally using graphical model representation and algorithms on graphical models. How do we do that? Here are suggestions:

   - Cast MWCSP as an equivalent graphical model problem that we discussed in the course so that the weighted clause satisfiability problem can be solved exactly.
   - Specify this equivalent graphical model (directed/undirected) exactly, including exact specification of the parameters (CPTs or clique potentials). Explain how the solution to the problem on the graphical model is equivalent to the solution to the MWCSP.
   - What procedure(s) discussed in class will be applicable on the graphical model to efficiently get the exact solution to the MWCSP? What will be the complexity of these algorithms with respect to the MWCSP?
   - Comment on what algorithm(s) could be used to learn the weights $w_1, w_2, \ldots, w_n$ from training data.

   (**8 Marks**)

2. In the class, we presented an example of a simple 3 node directed graphical model for which there exists no equivalent undirected graphical model. In this exercise you need to prove this by giving a concrete example of such a 3 node directed graphical model. Here are suggested steps:

- Assume each of the 3 nodes is binary.
- Present corresponding conditional probability tables for that directed graph (with precise numerical values in the tables) such that the marginal independence and conditional dependence can be represented by that directed graph but not by any undirected graph containing just those three nodes.
- You need to numerically show the marginal independence and the conditional dependence for the parameters chosen by you.

(**6 Marks**)

3. **Noisy-or representation:** Let $X_1, X_2, \ldots X_M$ be parents of Y in a bayesian network. Let $Y, X_1, \ldots, X_M \in \{0, 1\}$. The conditional probability table (CPT) for $p(Y|X_1, \ldots, X_M)$ should ordinarily consist of lot of entries. How many are they?

   Now, here is a representation of the same CPT using just M parameters! This representation, called the noisy-or representation, is as follows:

$$p(Y = 1|X_1, \ldots, X_M) = 1 - \prod_{i=1}^{M}(1 - \mu_i)^{X_i}$$

   where the parameters μ_i ($i = 1, 2, \ldots, M$) represent the probabilities $p(X_i = 1)$. Explain why this representation for the CPT for $p(Y|X_1, \ldots, X_M)$ might be called a 'noisy-or' representation.

   (**4 Marks**)

4. **This exercise is on learning with missing data:**

   (a) Consider the directed graph $\mathcal{G} = \langle \mathcal{V}, \mathcal{G} \rangle$ where $\mathcal{V} = \{H, A, B, C\}$ and $\mathcal{E} = \{(H, A), (H, B), (H, C)\}$. Consider the dataset $\mathcal{D}$ provided in the table below:

H	A	B	C	w
?	a_0	b_0	c_0	14
?	a_0	b_0	c_1	11
?	a_0	b_1	c_0	20
?	a_0	b_1	c_1	20
?	a_1	b_0	c_0	5
?	a_1	b_0	c_1	5
?	a_1	b_1	c_0	11
?	a_1	b_1	c_1	14

   where ? stands for missing observations. Consider the following random initial probabilities for the example execution of the expectation maximization algorithm.

p_H	h_0	h_1
	0.3	0.7

$p_{A\|H}$	h_0	h_1
a_0	0.4	0.6
a_1	0.6	0.4

$p_{B\|H}$	h_0	h_1
b_0	0.7	0.8
b_1	0.3	0.2

$p_{C\|H}$	h_0	h_1
c_0	0.2	0.5
c_1	0.8	0.5

Compute the probabilities (parameters) of the bayesian network after one iteration of the EM algorithm.
(**6 Marks**)

(b) Consider the algorithms for decision tree learning (or even decision list learning which several of you presented). Now consider a dataset that has some attribute values missing, such as the above dataset reproduced below[1] with some modification, where C can be treated as the class label and where values of H and A are ocassionally missing. Present atleast 4 different strategies for handling this missing attribute value problem. Your 4 alternatives **should not be specific to this dataset** and should be as different as possible.

Present the pros and cons of each choice. **Make sure that at least one of the strategies specifically takes advantage of the fact that you are learning decision tree (or decision list) and not any other arbitrary machine learning model**.

H	A	B	C	w
?	a_0	b_0	c_0	14
h_1	a_0	b_0	c_1	11
?	a_0	b_1	c_0	20
?	a_0	b_1	c_1	20
h_0	a_1	b_0	c_0	5
?	a_1	b_0	c_1	5
h_1	?	b_1	c_0	11

(**6 Marks**)

[1] The dataset is shown only as a sample and the answer to this question should not be specific to this dataset alone, though you can illustrate your idea on it.