Grad AI. 15-780 Fall, 2007

Homework 4

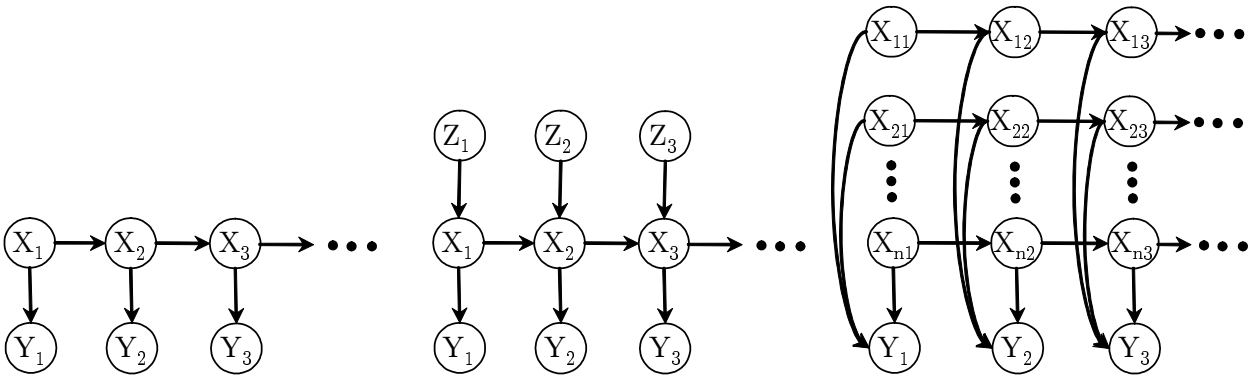
- *Homework deadline: 10:30am on November 13*
- *Please print your code and hand it in with the hard copy of your homework. Also send a copy of your code by e-mail to both TAs (gholling@andrew.cmu.edu and thlin@cs.cmu.edu).*

1. **Dynamic Bayesian Network (30 pts)**

   (a) The hidden Markov models (HMM), input output hidden Markov models (IOHMM), and factorial HMM are all special cases of dynamic Bayesian networks (DBN), with network topologies shown below. Furthermore a DBN can be considered as a large Bayesian network (BN). Based on the BN semantics, we can write the factored joint distribution of the models. For example, the joint distribution of the HMM with m time points is,

   $$p(X_1)p(Y_1|X_1)\prod_{t=2}^{m} p(X_t|X_{t-1})p(Y_t|X_t)$$

   Now write the factored joint distribution of the IOHMM and factorial HMM with m time points in the figure below (5 pts each).



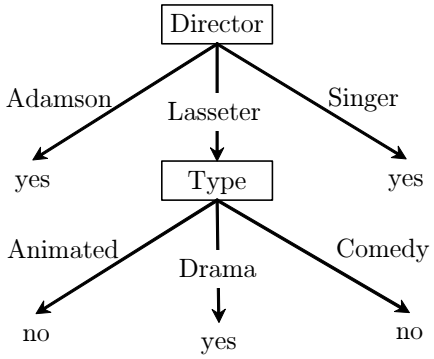
i. HMM ii. IOHMM iii. Factorial HMM

   (b) What is the maximum number of edges in a Bayesian network (BN) with n nodes? Describe how to construct a network containing this number of edges (remember that a BN is a DAG). (10 pts)

   (c) What is the maximum number of edges in a dynamic Bayesian network (DBN) with $2n$ nodes (n in each of the two time slices)? For this answer count the total number of edges in the first and second time slices and the number of edges between the two time slices. Describe how to construct a DBN with this number of edges. (10 pts)

2. **Decision Tree (30 pts)**

During the class, we discussed the movie preference domain: A movie is defined by four attributes, *Type*, *Length*, *Director*, and *FamousActors*; these attributes can help us predict whether a particular viewer likes the movie or not. The figure below shows a data set (right), and the decision tree learned from this data set (left). In the class, we went through the learning process on this data set step by step (if not clear, read the slide and make sure you understand how the conditional entropy and information gain are calculated). Now assume we indeed learned the true model of this viewer. This problem asks you, what could happen when more *correct* training examples are given (i.e. training examples always match this decision tree)? For each question, show a set of possible examples and the information gain in each step if your answer is yes, or explain why if your answer is no.



Movie	Type	Length	Director	Famous actors	Liked?
M1	Comedy	Short	Adamson	No	Yes
M2	Animated	Short	Lasseter	No	No
M3	Drama	Medium	Adamson	No	Yes
M4	Animated	Long	Lasseter	Yes	No
M5	Comedy	Long	Lasseter	Yes	No
M6	Drama	Medium	Singer	Yes	Yes
M7	Animated	Short	Singer	No	Yes
M8	Comedy	Long	Adamson	Yes	Yes
M9	Drama	Medium	Lasseter	No	Yes

Decision trees are always learned by the standard method shown in the class: each time, select the attribute with highest information gain, and when there is a tie, select the attribute by the order above (*Type* first and *FamousActors* last). There is no pruning, random forest, or any other advanced techniques.

(a) Is it possible to add more correct training examples to the 9 movies, so that the decision tree learned from examples will be identical to the true decision tree except having one more node under $Type = Drama$? (10 pts)

(b) Is it possible to add more correct training examples to the 9 movies, so that the decision tree learned from examples will include *Length*, an irrelevant attribute? (10 pts)

(c) Is it possible to add **one** correct training example to the 9 movies, so that the decision tree learned from examples will be different from the decision tree above? (10 pts)

3. **Decision Tree and Neural Network (40 pts)**

(a) Consider the *AND* function and the *OR* function that map n Boolean inputs (0 or 1) to one Boolean output. The *AND* function returns 1 if and only if all inputs are 1. The *OR* function returns 0 if and only if all inputs are 0.

Can a decision tree represent the *AND* function, and the *OR* function? For each function, show the decision tree if your answer is yes, or explain why if your answer if no. (5 pts)

(b) Can a single-layer perceptron represent the *AND* function, and the *OR* function? For each function, show the preceptron if your answer is yes, or explain why if your answer if no. (5 pts)

(c) Now consider any Boolean function that maps n Boolean inputs to a Boolean output, given to you as a truth table. Can you systematically construct a decision tree to represent any Boolean function? If your answer is yes, describe how you can do this. If your answer is no, give an counterexample that cannot be represented by a decision tree, and give a brief explanation why not. (15 pts)

(d) Similarly, can you systematically construct a single-layer perceptron to represent any Boolean function? If your answer is yes, show how you can do this. If your answer is no, provide a counterexample that cannot be represented by a single-layer perceptron, and give a brief explanation why not. (15 pts)