CS 760
Machine Learning
Fall 2016 Exam

Name _____

Write your answers on these pages and show your work. You may use the back sides of pages as necessary. Before starting, make sure your exam has every page (numbered **1** through **8**).

Problem	**Score**	**Max Score**
1.	_____	15
2.	_____	15
3.	_____	15
4.	_____	15
5.	_____	15
6.	_____	15
7.	_____	10
Total		100

1. Information gain and feature selection (15 points):

(a) Show how you would use information gain to select <u>two</u> features from the given training set that has two Boolean features and one three-valued nominal feature.

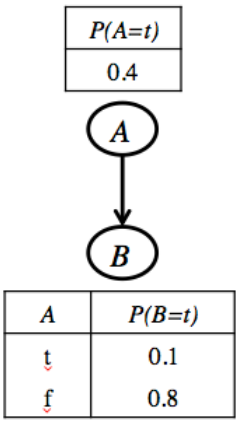
	A	*B*	*C*	*Class*
Instance 1	F	T	red	neg
Instance 2	T	F	red	neg
Instance 3	T	T	red	neg
Instance 3	T	T	blue	pos
Instance 4	F	F	green	pos
Instance 5	F	T	red	pos

(b) Briefly describe <u>two</u> limitations of using information gain for feature selection.

(c) For <u>one</u> of the limitations you listed above, briefly describe an alternative feature selection approach that would not be as susceptible to this limitation.

2. Bayesian network parameter learning (15 points): Suppose you are given the training set shown below on left, and the Bayes net structure shown on the right, along with initial parameters. Show the estimated parameters for the network when using m-estimates with $m = 2$ and uniform priors, and only one iteration of the EM algorithm. The symbol '?' is used to indicate a value that is missing.

A	B
t	f
?	t
t	f
t	t
f	?
f	f
f	t
f	t

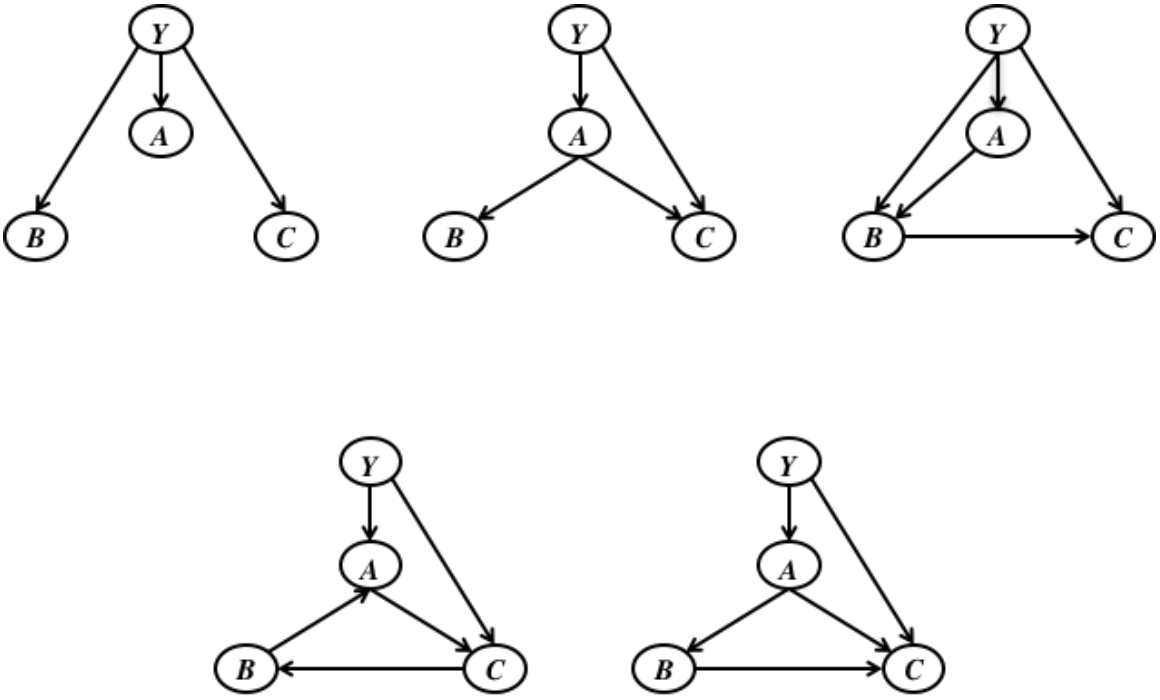


A	P(B=t)
t	0.1
f	0.8

3. Bayesian network structure learning (15 points):

(a) For each of the Bayes net structures show below, indicate which of the following methods could have returned it as a learned model:

- TAN (assume that Y is the class variable, and all conditional mutual information values > 0)
- Sparse Candidate with k=2
- Hill-climbing search with add, delete and reverse edge operators
- Chow-Liu (assume all mutual information values > 0)

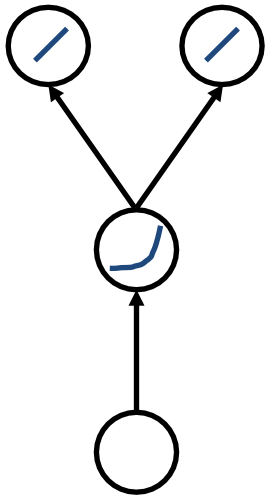


(b) Briefly describe the inductive bias of the TAN algorithm.

4. Neural Networks (15 points):

Suppose you are given the neural network below that maps an input value x to two output values y_1 and y_2. The output units use linear activation functions, and the hidden unit uses an exponential activation function. You can assume that these units do not use biases. Show the specific update rule that would be used for each weight in the network when doing online training with the following error function: $\frac{1}{2}(y_1 - o_1)^2 + \frac{1}{2}(y_2 - o_2)^2$ where y_1 and y_2 are the target values, and o_1 and o_2 are the activations of the two output units.

Hint: the derivative of $\exp(a)$ is $\exp(a)$



5. PAC Learning Theory (15 points):

(a) Consider the concept class C, in which each concept is represented by a pair of circles centered at the origin, $(0, 0)$. Let r be the radius of the inner circle and $r+a$ be the radius of the outer circle (a is a positive number). Each training instance is represented by two real-valued features x_1 and x_2, and a binary class label $y \in \{0, 1\}$. The concept predicts y=1 for instances that are outside the radius of the inner circle and inside the radius of the outer circle, and y=0 otherwise. Show that C is PAC learnable.

(b) Consider the concept class C in which each concept is represented by $sign(\sin(\omega x + \theta))$ where x is a real number and $sign(a)$ predicts $y = 1$ for positive values of a and $y = -1$ for negative values of a. Note that $sign(\sin(\omega x + \theta))$ can shatter any set of m instances with a sufficiently large value of ω (frequency) and appropriate θ (phase). What does this observation tell us about the PAC learnability of the class C?

6. Randomization in Supervised Learning (15 points):
Several supervised learning approaches employ randomization, including *bagging*, *random forests*, and *dropout*. Briefly describe each of these three approaches in terms of the following aspects.

(a) What type of model is used in the approach?

(b) What is randomly selected by the approach?

(c) When during the learning process are the random selections made?

(d) What, if anything, is randomly selected at classification time?

7. Short Answer (10 points): Briefly define each of the following terms.

autoencoder

lesion study

agnostic PAC learning

lasso regression

hinge loss