CS-171, Intro to A.I. — Quiz#4 — Fall Quarter, 2013 — 20 minutes

YOUR NAME: _____

YOUR ID: _____ ID TO RIGHT: _____ ROW: _____ NO. FROM RIGHT: _____

1. (5 pts) Definition of conditional probability. Write down the definition of P(H | D) in terms of P(H), P(D), P(H $\wedge$ D), and P(H $\vee$ D).

P(H | D) = P(H $\wedge$ D) / P(D)

2. (5 pts) Bayes' Rule. Write down the result of applying Bayes' Rule to P(H | D).

P(H | D) = P(D | H) P(H) / P(D)

3. 15 pts total, 3 pts each) Machine Learning. Label the following statements T (true) or F (false).

3a. T A decision tree can learn and represent any Boolean function.

3b. F The information gain from an attribute A is how much classifier accuracy improves when attribute A is added to the example feature vectors in the training set.

3c. T Overfitting is a general phenomenon that occurs with all types of learners.

3d. F Cross-validation is a way to improve the accuracy of a learned hypothesis by reducing over-fitting using Ockham's razor.

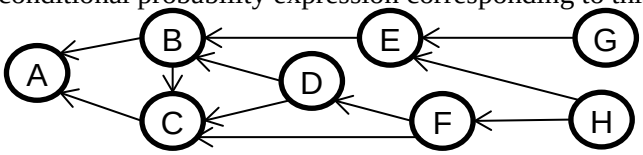
3e. T An agent is learning if it improves its performance on future tasks after making observations about the world.

4. (30 pts total, 2 pts each) Machine Learning concepts.
For each of the following items on the left, write in the letter corresponding to the best answer or the correct definition on the right. The first one is done for you as an example.

A	Learning	A	Improves performance of future tasks after observing the world
J	Information Gain	B	Fixed set, list, or vector of features/attributes paired with a value
M	Decision Boundary	C	Agent learns patterns in the input with no explicit feedback
P	Support Vector Machine	D	Agent observes input-output pairs & learns to map input to output
L	Cross-validation	E	Example input-output pairs, from which to discover a hypothesis
N	Linear Classifier	F	Examples distinct from training set, used to estimate accuracy
B	Factored Representation (Feature Vector)	G	Supervised learning with a discrete set of possible output values
D	Supervised Learning	H	Supervised learning with numeric output values
F	Test Set	I	Internal nodes test a value of an attribute, leaf nodes=class labels
O	Naïve Bayes Classifier	J	Expected reduction in entropy from testing an attribute value
G	Classification	K	Choose an over-complex model based on irrelevant data patterns
I	Decision Tree	L	Randomly split the data into a training set and a test set
H	Regression	M	Surface in a high-dimensional space that separates the classes
E	Training Set	N	Tests **w**·**f**>0, where **w** is a weight vector and **f** is a feature vector
C	Unsupervised Learning	O	Tests P (C) Π_i P(X_i | C), where C is a class label and X_i are features
K	Overfitting	P	Current most-popular "off-the-shelf" supervised learning method

****** TURN QUIZ OVER. QUIZ CONTINUES ON THE REVERSE. ******

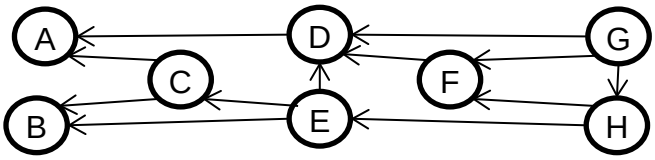
5. (15 pts total, -5 for each error, but not negative) Bayesian Networks. Write down the factored conditional probability expression corresponding to this Bayesian Network.



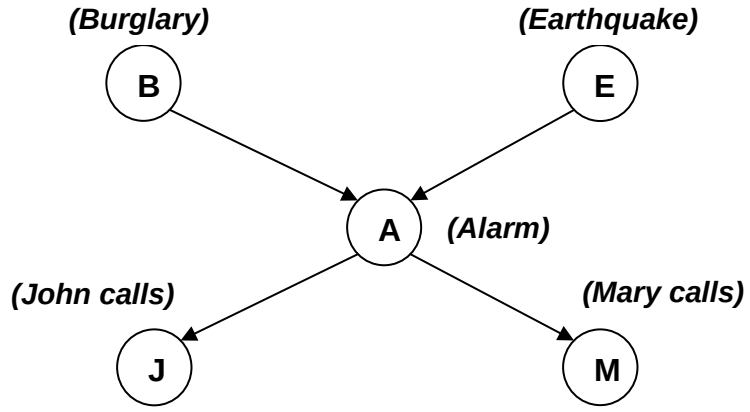
P(A | B, C) P(B | D, E) P(C | B, D, F) P(D | F) P(E | G, H) P(F | H) P(G) P(H)

6. (15 pts total, -5 for each error, but not negative) Bayesian Networks. Draw the Bayesian Network corresponding to this factored conditional probability expression. Draw left-to-right, as in Problem 5.

P(A | C, D) P(B | C, E) P(C | E) P(D | E, F, G) P(E | H) P(F | G, H) P(G) P(H | G)



7. (15 pts total, -5 for each error, but not negative) Bayesian Networks. Shown below is the Bayesian network corresponding to the Burglar Alarm problem, P(J | A) P(M | A) P(A | B, E) P(B) P(E).



P(B)
.001

P(E)
.002

B	E	P(A)
t	t	.95
t	f	.94
f	t	.29
f	f	.001

A	P(J)
t	.90
f	.05

A	P(M)
t	.70
f	.01

Write down an expression that will evaluate to P(j=T $\wedge$ m=F $\wedge$ a=T $\wedge$ b = F $\wedge$ e = T). Express your answer as a series of numbers (numerical probabilities) separated by multiplication symbols. You do not need to carry out the multiplication to produce a single number (probability). **SHOW YOUR WORK.**

P(j=T $\wedge$ m=F $\wedge$ a=T $\wedge$ b = F $\wedge$ e = T)

= P(j=T | a=T) * P(m=F | a = T) * P(a=T | b=F $\wedge$ e=T) * P(b=F) * P(e=T)

= .9 * .3 * .29 * .999 * .002