CS 540: Introduction to Artificial Intelligence

Final Exam: 12:25-2:25pm, December 17, 2014
Room 132 Noland

CLOSED BOOK
(two sheets of notes and a calculator allowed)

Write your answers on these pages and show your work. If you feel that a question is not fully specified, state any assumptions that you need to make in order to solve the problem. You may use the backs of these sheets for scratch work.

Write your name on this page and initial all other pages of this exam. Make sure your exam contains seven problems on eleven pages.

Name _____

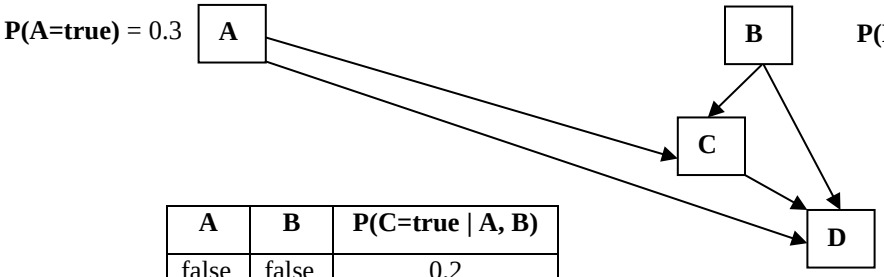
Student ID _____

Problem	**Score**	**Max Score**
1	_____	**15**
2	_____	**10**
3	_____	**15**
4	_____	**20**
5	_____	**15**
6	_____	**10**
7	_____	**15**
TOTAL	_____	**100**

Initials: _____

Problem 1 – Bayesian Networks (15 points)

Consider the following Bayesian Network, where variables **A-D** are all Boolean-valued:



A	B	P(C=true | A, B)
false	false	0.2
false	true	0.4
true	false	0.6
true	true	0.8

A	B	C	P(D=true | A, B, C)
false	false	false	0.1
false	false	true	0.2
false	true	false	0.3
false	true	true	0.4
true	false	false	0.5
true	false	true	0.6
true	true	false	0.7
true	true	true	0.8

a) What is the probability that ***A*** and ***C*** are *true* but ***B*** and ***D*** are *false*? _____
[Be sure to show your work for Parts a-c. Put your (numeric) answers on the lines provided.]

b) What is the probability that ***A*** is *false*, **B** is *true*, and ***D*** is *true*? _____

c) What is the probability that ***C*** is *true* given that ***A*** is *false*, ***B*** is *true*, and ***D*** is *true*?

_____

Initials: _____

Problem 2 – Naïve Bayes (10 points)

a) Consider the following training set, where three Boolean-valued features are used to predict a Boolean-valued output. Assume you wish to apply the Naïve Bayes algorithm.

Ex #	A	B	C	Output
1	True	False	True	True
2	False	True	False	True
3	False	True	True	True
4	False	False	True	False

Calculate the ratio below, showing your work below it and putting your final (numeric) answer on the line to the right of the equal sign. Be sure to explicitly show in your work the counts due to pseudo examples.

$$\frac{\text{Prob(Output = True} \mid \text{A = False, B = False, C = False)}}{\text{Prob(Output = False} \mid \text{A = False, B = False, C = False)}} = \_\_\_\_\_$$

b) What is the most likely output for this example (A = False, B = False, C = False)? _____

Briefly explain your answer below.

Initials: _____

Problem 3 – Representing Knowledge with First-Order Logic (15 points)

Convert each of the following English sentences into *First-Order Predicate Calculus* (FOPC), using reasonably named predicates, functions, and constants. If you feel a sentence is ambiguous, clarify which meaning you're representing in logic. (Write your answers in the space *below* each English sentence.)

John's sister is married to someone who works at UW-Madison.

People allergic to pollen need to get a flu shot.

Occasionally small packages sent to Mary are valuable.
Large packages sent to Mary almost always are valuable.

[You must use the notation of *Markov Logic Networks* here and write one wff for each of these two sentences.]

Initials: _____

Problem 4 – Multiple-Choice Questions (20 points)

For each question, circle your answer(s). Choose the one *best* answer (except for Part *j*).

a) Simulated annealing can find goal states in cases where hill climbing fails.

TRUE ***FALSE***

b) Compared to a full-joint probability table, a Bayesian Network will

i) always be strictly less accurate, though possibly only by a small amount
ii) be able to answer a large subset of possible queries, but not all
iii) possibly use significantly less computer memory
iv) require more training examples to reach the same testset accuracy

c) *Drop Out* is used in

i) A* search
ii) alpha-beta pruning
iii) genetic algorithms
iv) neural networks

d) A well-formed formula (wff) that is always true regardless of the truth values of the atomic predicates it contains is called

i) commensurate
ii) equivalent
iii) satisfiable
iv) valid

e) *Ensembles* are

i) a collection of trained models for a given prediction task
ii) a set of sound logical inference rules
iii) used to address the horizon problem
iv) what the fashionable robot wears

f) If expression$_1$ and expression$_2$ unify and expression$_2$ and expression$_3$ unify,
then expression$_1$ and expression$_3$

i) definitely also unify
ii) definitely do not unify
iii) might unify, but not necessarily

g) Some neural networks are called ‘deep’ because

i) large numbers of examples are used to train them
ii) they contain a large number of hidden units
iii) they contain many layers of hidden units
iv) they have been trained on challenging tasks that require substantial intelligence

h) *Kernels* (i.e., similarity functions) used in support-vector machines play a role most similar to

i) heuristic functions in search
ii) hidden units in neural networks
iii) resolution in theorem proving
iv) tuning sets in machine learning

i) *Situation calculus* is used to

i) calculate gradients for backpropagation
ii) define the singularity
iii) represent predicates whose truth value can change over time
iv) search for good moves when computers play games like chess

j) Circle <u>all</u> those search strategies that can suffer from OPEN getting too large
even when the maximum number of legal actions per node is small

i) A*
ii) breadth-first
iii) hill climbing
iv) simulated annealing

Initials: _____

Problem 5 – Logical Reasoning (15 points)

a) What is the *most-general unifier* (mgu), if any, of these two wff's?

P(?x, ?x, ?y) *P(f(1,3), ?z, ?z)*

Θ = { }

_____

b) Given the following background knowledge

1. P
2. $\neg$ Q
3. P $\rightarrow$ R
4. $\neg$ Q $\vee$ W
5. W $\rightarrow$ P
6. $\neg$ R $\vee$ W

Show W by filling out the table below, using as many lines as needed.

Number	WFF	Justification
7		
8		
9		
10		
11		
12		
13		
14		
15		

Initials: _____

c) Given the following clauses, show *P(John)* must be true by adding $\neg P(John)$ and using only the resolution inference rule to derive a contradiction.

Use the notation presented in class (and in the book) where the resulting clause is connected by lines to the two clauses resolved and any necessary unifications are listed. (If you don't recall that notation, use the notation appearing in Part (b) above for partial credit.)

$R(?x) \vee \neg Q(?x)$	$P(?y) \vee \neg R(?y) \vee \neg W(?y)$	$W(?z)$	$Q(John)$

Initials: _____

Problem 6 – Artificial Neural Networks (10 points)

Consider a *perceptron* that has two real-valued inputs and an output unit that uses a *step function* as its output function. All the initial weights and the output unit's threshold equal 0.3. Assume the teacher has said that the output should be 0 for the input *in1 = 2* and *in2 = 4*.

a) Show how the perceptron learning rule (also called the delta rule) would alter this neural network upon processing this training example. Let η (the learning rate) be 0.5.

Perceptron BEFORE Training

Perceptron AFTER Training

b) Briefly describe below one important advantage support vector machines have over perceptrons.

Initials: _____

Problem 7 – Miscellaneous Questions (15 points)

a) How should *k* be chosen in the *k*-NN algorithm? Circle the answer you feel is best.

i) a tuning set should be used to evaluate a set of candidates for the value of *k*

ii) any odd number between 1 and 10% of the number of examples in the data set will suffice

iii) one should perform gradient descent to find a local minimum of error

iv) use the value of *k*, chosen from a set of candidates, that leads to the best test-set accuracy

Briefly justify your answer below.

b) Assume you are given this initial dataset.

Ex #	A	B	C	D	Output
1	1	2	0	5	0
2	3	5	4	8	1
3	2	1	2	6	0

You wish to use the "- Euclidean distance" kernel to create a new dataset for use by a support-vector machine (the '-' is a minus sign because similarity is the opposite of distance).

Recall that Euclidean distance is the square root of the sum of the squared differences between each of two examples' corresponding feature values.

Show the new dataset below. Be sure to label the columns and rows.

c) Qualitatively draw a (2D) picture of *weight space* where the backpropagation algorithm is likely to

i) do well

ii) do poorly

Be sure to label your axes and briefly explain your answers.

Have a good break!