CS 540: Introduction to Artificial Intelligence

Final Exam: 8:15-9:45am, December 21, 2016
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 <u>six</u> problems on eight pages.
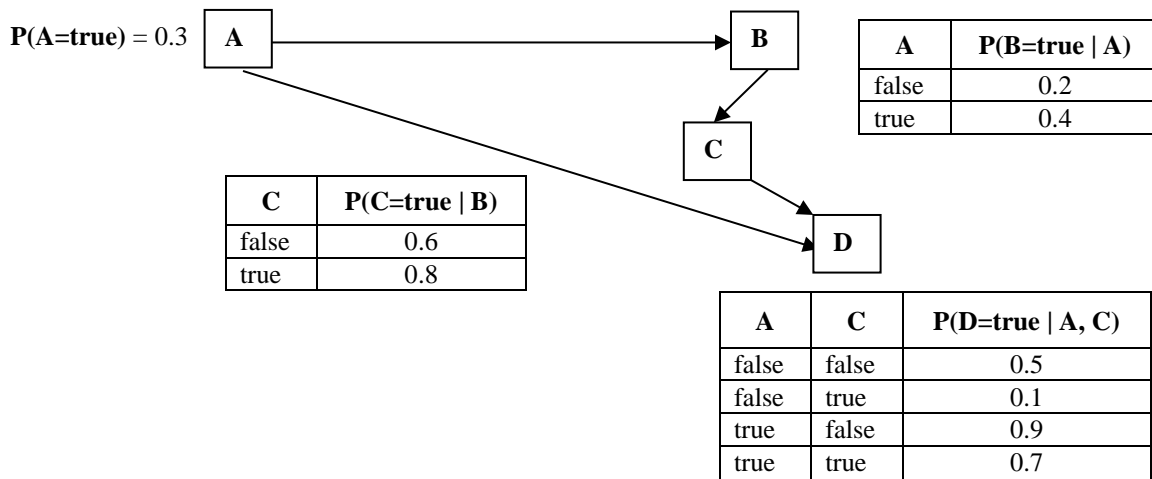
Name _____

UW Net ID _____

Problem	Score	Max Score
1	_____	17
2	_____	10
3	_____	18
4	_____	15
5	_____	15
6	_____	25
TOTAL	_____	100

Problem 1 – Bayesian Networks (17 points)

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

$P(A=true) = 0.3$

A	P(B=true | A)
false	0.2
true	0.4

C	P(C=true | B)
false	0.6
true	0.8

A	C	P(D=true | A, C)
false	false	0.5
false	true	0.1
true	false	0.9
true	true	0.7

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

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

c) What is the prob that **A** is *true* <u>given</u> that **B** is *false*, **C** is *false*, and **D** is *true*? _____

Problem 2 – Naïve Bayes (10 points)

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

Ex #	A	B	Output
1	True	False	1
2	True	True	0
3	True	True	1

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 <u>explicitly show</u> in your work the counts due to pseudo examples.

$$\frac{Prob(Output = 1 \ | \ A = False, B = False)}{Prob(Output = 0 | \ A = False, B = False)} \ = \ _____$$

Problem 3 – Representing Knowledge with First-Order Logic (18 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.)

There exists a dog that is older than all other dogs.

Students who do not understand kernels can learn about them by taking cs540.
[You must use situation calculus here. Do not use Markov Logic, though.]

Between every pair of real numbers there is another real number.

Most integers are not prime.

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

Problem 4 – Logic (15 points)

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

$$Q(?x, f(?x), f(?y))\qquad\qquad Q(?w, f(g(?z)), ?w)$$

$$\Theta = \{\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\}$$

_____

b) Put the following into clausal form (write your answer on the line below):

$$\forall x\ [\ p(x) \leftrightarrow q(2)\ ]$$

_____

c) Given the following clauses, show $\exists w\ C(w)$ must be true by assuming its negation and using <u>only</u> the <u>resolution inference rule</u> to derive a contradiction. In the clauses below all the variables are universally quantified.

Use the notation presented in class (and in the book) where the resulting clause is connected by lines to the two clauses resolved and indicate any variable bindings needed.

| A(3) | ¬ A(x) ∨ B(x) | ¬ B(y) ∨ A(y) | ¬ B(z) ∨ C(z) | C(1) ∨ D(2) |

Problem 5 – ANNs and SVMs (15 points)

Consider a *perceptron* that has <u>two</u> 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 0.1. Assume the teacher has said that the output should be 0 for the input:

$$in1 = 0.0 \qquad in2 = 1.2$$

a) Show how the perceptron learning rule (also called the delta rule) would alter this neural network upon processing the above training example. Let η (the learning rate) be 0.3.

Perceptron BEFORE Training

Perceptron AFTER Training

a) Assume you are given this initial dataset and wish to use the <u>Gaussian kernel</u> with variance (i.e., σ^2) equal to 1 to create a <u>new</u> dataset. (If you do not recall the Gaussian kernel, use the dot-product kernel for partial credit.)

Ex #	A	B	Output
1	0.1	0.9	0
2	0.2	0.8	0
3	0.7	0.4	1

<u>Show the new dataset below</u>. Be sure to clearly label the columns and rows.

Problem 6 – Miscellaneous Questions (25 points)

a) You have two Markov Logic Network rules: $wgt = 2 \; P \rightarrow Q$ $wgt=1 \; Q \rightarrow \neg P$

What is the probability P is true? Show your work below. Answer: _____

b) Two methods for avoiding *overfitting* in neural networks are:

    i.   _____

    ii.   _____

c) Two methods covered this semester for *clustering* <u>un</u>labeled training examples are:

    i.   _____

    ii.   _____

d) Two machine learning approaches covered this semester that do <u>not</u> require users to represent examples as *fixed-length feature vectors* are:

    i.   _____

    ii.   _____

e) *Circle* the phrase(s) below that is (are) *not* a valid logical inference rule

    i.   AND Elimination

    ii.   OR Elimination

    iii.   AND Introduction

    iv.   OR Introduction

    v.   Modus Facto

f) *Circle* the topic(s) below that was (were) <u>not</u> covered this semester.

    i.   Byzantine Agreements

    ii.   Max Pooling

    iii.   Reinforcement Learning

    iv.   Searle's Chinese Room

    v.   Simulated Annealing

g) Consider the kernel function below, where examples are Boolean-valued features:

$$K(ex_i, ex_j) = 1 \text{ if the two feature vectors are exactly equal}$$
$$= 0 \text{ otherwise}$$

Briefly describe <u>below</u> the main reason why this would <u>*not*</u> be a good kernel function.

Happy holidays and best wishes for 2017.