CS 540: Introduction to Artificial Intelligence

Final Exam: 5:30-7:30pm, December 17, 2015
Beatles Room at Epic

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 ten problems on eleven pages.

Name ______________________________

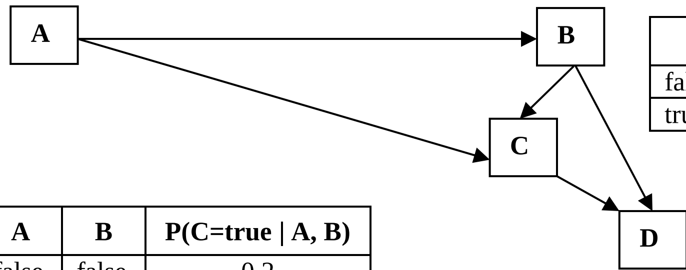
UW Net ID ______________________________

Problem	Score	Max Score
1	______	12
2	______	10
3	______	13
4	______	8
5	______	12
6	______	10
7	______	10
8	______	10
9	______	10
10	______	5
TOTAL	______	100

Problem 1 – Bayesian Networks (12 points)

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

P(A=true) = 0.8



A	P(B=true | A)
false	0.1
true	0.7

A	B	P(C=true | A, B)
false	false	0.2
false	true	0.4
true	false	0.3
true	true	0.5

B	C	P(D=true | B, C)
false	false	0.6
false	true	0.8
true	false	0.1
true	true	0.9

a) What is the probability that **A**, ***B*** and ***D*** are *true* but ***C*** is *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 *true*, **C** is *false*, and ***D*** is *true*? ______________

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

Initials: ______________________

Problem 2 – Naïve Bayes (10 points)

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

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

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.

$$\frac{Prob(Output = True \mid A = True, B = False)}{Prob(Output = False \mid A = True, B = False)} = \underline{\qquad\qquad}$$

b) What is $Prob(Output = True \mid A = True, B = False)$? ______________________
Show your work below.

Problem 3 – Representing Knowledge with First-Order Logic (13 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 is this one history student who has read every history book in the library.

Washing a car does not change who owns it. [You must use situation calculus here.]

Most students who get an A in cs302 also get an A in cs367.
Triangles have three sides.

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

Problem 4 – Resolution (8 points)

Given the following propositional-logic clauses, show *E* must be true by adding $\neg E$ 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. (If you don't recall that notation, use the notation appearing in Problem 5 below for partial credit.)

$\boxed{A \vee \neg C}$ $\boxed{B \vee \neg A}$ $\boxed{B \vee \neg D}$ $\boxed{E \vee \neg A \vee \neg B}$ $\boxed{C}$ $\boxed{\neg D}$

Problem 5 – Unification and Natural Deduction (12 points)

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

R(?x, ?y, ?x, f(?y)) *R(f(?z), ?z, f(g(2)), ?w)*

Θ = { ____ }

b) Given the following background knowledge and use the 'natural deduction' inference rules presented in class and in the textbook (do not convert these WFFs to CNF clauses).

1. $P(1)$
2. $Q(1) \vee Q(2)$
3. $\forall c\ P(c) \rightarrow \neg Q(c)$
4. $\forall d,e\ \ P(d) \wedge Q(e) \rightarrow A(e) \vee A(d)$
5. $\forall f,g\ \ P(f) \wedge Q(g) \rightarrow R(g, f)$
6. $\forall h,i\ \ R(h, i) \rightarrow A(h)$

Show $\exists a\ A(a)$ by filling out the table below, using as many lines as needed.

Number	WFF	Justification
7		
8		
9		
10		
11		
12		
13		
14		

Initials: ______________________

Problem 6 – Artificial Neural Networks (10 points)

Consider a *perceptron* that has three 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 2. Assume the teacher has said that the output should be 1 for the input:

$$in1 = 4 \qquad in2 = 0 \qquad in3 = -5.$$

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 0.10.

Perceptron BEFORE Training

Perceptron AFTER Training

b) Using the two pictures of feature space below, illustrate one important advantage ANNs with hidden units have over perceptrons. Briefly explain your drawings below each of them.

Perceptrons

ANN with THREE (3) HUs

Problem 7 – Support Vector Machines (10 points)

a) Assume you are given this initial dataset and wish to use a kernel to create a new dataset.

Ex #	A	B	C	Output
1	1	0	0	1
2	1	0	1	1
3	0	1	1	0

Your *kernel function* is the number of feature values in common.
Show the new dataset below. Be sure to clearly label the columns and rows.

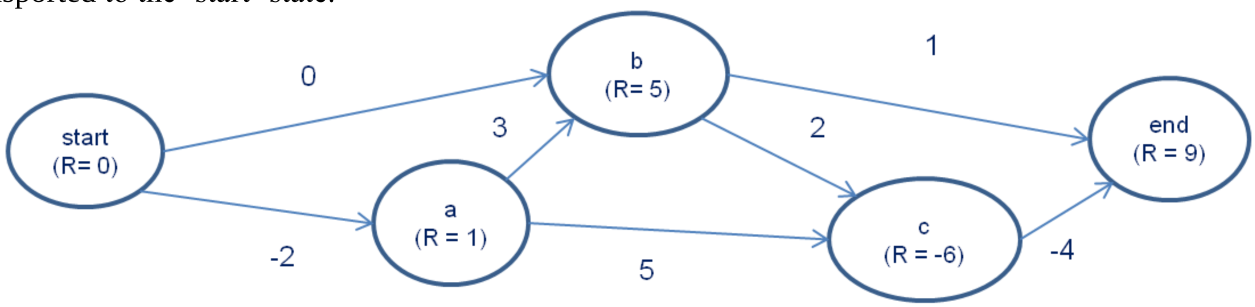
b) Assume a one-norm SVM puts weight = 4 on feature *P*, weight = -3 on feature *Q*, and weight = 1 on feature *R*. It sets $\gamma = 5$. What would the cost of this solution be, based on the three training examples below, assuming $\mu = 2$? Show your work below.

	P	*Q*	*R*	*Output*
Ex1	0	1	2	0
Ex2	-1	-1	4	0
Ex3	1	0	-1	1

Cost: ______________

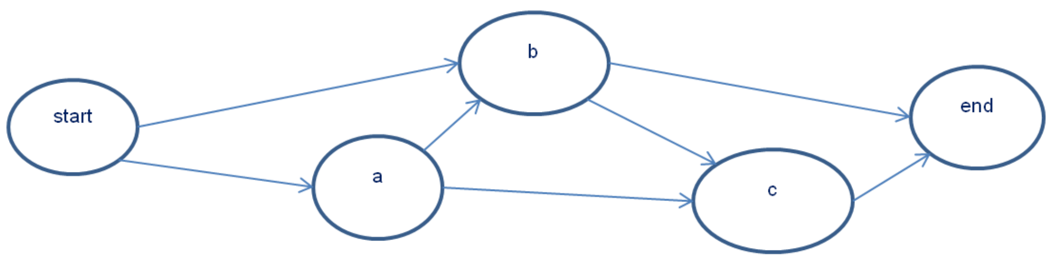
Problem 8 – Reinforcement Learning (10 points)

Consider the *deterministic* reinforcement environment drawn below, where the current state of the *Q* table is indicated on the arcs. Let γ=*0.9*. I*mmediate* rewards are indicated inside nodes. Once the agent reaches the 'end' state the current episode ends and the agent is magically transported to the 'start' state.

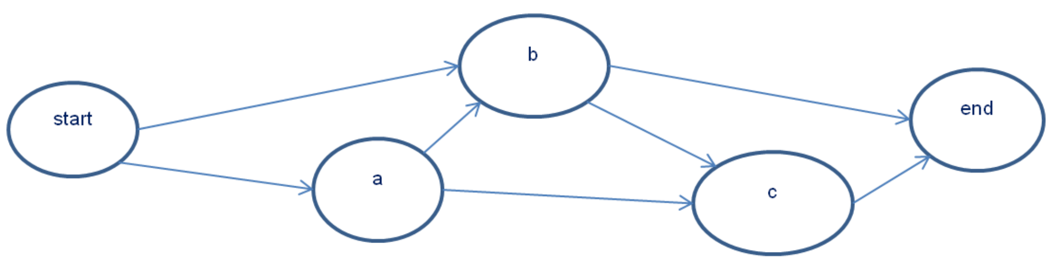


a) Assuming our RL agent *exploits* its *policy* (with learning turned off), what is the path it will take from *start* to *end?* Answer: ______________________________________
Briefly explain your answer:

b) Assuming the RL agent is using *one*-step Q learning and moves from node *a* to node *b*. Report below the changes to the graph above (only display what changes). Show your work.



c) Show the final state of the *Q* table after a very large number of training episodes (i.e., show the *Q* table where the Bellman Equation is satisfied everywhere). No need to show your work nor explain your answer.



Problem 9 – Miscellaneous Questions (10 points)

a) Assume we have a binary classification problem where Feature *A* has 3 values, Feature *B* has 5 values, and Feature *C* has 4 values. How big is a *full joint probability* table for this problem? Show your work.

Answer: ___________________

b) An unavoidable weakness of SVMs is that the kernel matrix produced when using kernels is of size N^2, where N is the number of examples: **TRUE** or **FALSE** ? (circle one)
Briefly explain your answer below.

c) You have two Markov Logic Network rules: $wgt = 2 \;\; true \rightarrow P \qquad wgt=3 \;\; true \rightarrow \neg P$
(Not surprisingly, the propositional '*true*' is true in all world states.)
What is the probability *P* is true? Show your work below. Answer: ___________________

d) Often data sets have missing values for some features in some examples. Circle the method below that is the best way to 'fill in' these missing values. No need for an explanation.

i. Drop In
ii. Expectation-Maximization
iii. *K*-Means
iv. Transfer Learning

e) A 'complete world state' that makes every WFF in a set of WFFs true is called a/an (circle one and no need to explain your answer):

i. Interpretation
ii. Model
iii. Skolemizer
iv. Tautology

Problem 10 – Searching for FOPC Rules (5 points)

Assume we are doing Inductive Logic Programming to learn to predict the two-argument predicate POS and have popped off OPEN this rule:

$$\forall x,y \quad P(x) \land Q(y) \rightarrow POS(x, y)$$

Our only applicable search operator is "add the two-argument predicate R" and assume we wish to *not* use constants in our rules. Show five (5) possible next states that might be added to OPEN. No need to explain your answers. Write your answers slightly above the lines below.

a) ____

b) ____

c) ____

d) ____

e) ____

Happy holidays and best wishes for 2016.