## 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**

**Problem 1 – Bayesian Networks (15 points)**

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

**P(B=true)** = 0.8

**P(A=true)** = 0.3

# A

# B

# C

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

# D

1. 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.]

1. What is the probability that ***A*** is *false*, **B** is *true*, and ***D*** is *true*? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
2. What is the probability that ***C*** is *true* given that ***A*** is *false*, ***B*** is *true*, and ***D*** is *true*?   
    \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Problem 2 – Naïve Bayes (10 points)**

1. 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.

Prob(Output = True | A = False, B = False, C = False)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ = \_\_\_\_\_\_\_\_\_\_\_

Prob(Output = False | 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.

**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.]

**Problem 4 – Multiple-Choice Questions (20 points)**

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

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

***TRUE FALSE***

1. Compared to a full-joint probability table, a Bayesian Network will
2. always be strictly less accurate, though possibly only by a small amount
3. be able to answer a large subset of possible queries, but not all
4. possibly use significantly less computer memory
5. require more training examples to reach the same testset accuracy
6. *Drop Out* is used in
7. A\* search
8. alpha-beta pruning
9. genetic algorithms
10. neural networks
11. A well-formed formula (wff) that is always true regardless of the truth values of the atomic predicates it contains is called
    1. commensurate
    2. equivalent
    3. satisfiable
    4. valid
12. *Ensembles* are
13. a collection of trained models for a given prediction task
14. a set of sound logical inference rules
15. used to address the horizon problem
16. what the fashionable robot wears
17. If expression1 and expression2 unify and expression2 and expression3 unify,   
    then expression1 and expression3
18. definitely also unify
19. definitely do not unify
20. might unify, but not necessarily
21. Some neural networks are called ‘deep’ because
22. large numbers of examples are used to train them
23. they contain a large number of hidden units
24. they contain many layers of hidden units
25. they have been trained on challenging tasks that require substantial intelligence
26. *Kernels* (i.e., similarity functions) used in support-vector machines play a role most similar to  
     i) heuristic functions in search
27. hidden units in neural networks
28. resolution in theorem proving
29. tuning sets in machine learning
30. *Situation calculus* is used to  
     i) calculate gradients for backpropagation  
     ii) define the singularity
31. represent predicates whose truth value can change over time

iv) search for good moves when computers play games like chess

1. Circle all those search strategies that can suffer from OPEN getting too large   
   even when the maximum number of legal actions per node is small
2. A\*
3. breadth-first
4. hill climbing
5. simulated annealing

**Problem 5 – Logical Reasoning (15 points)**

1. 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 = { }*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Given the following background knowledge
2. P
3. ¬ Q
4. P → R
5. ¬ Q ∨ W
6. W → P
7. ¬ R ∨ 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

1. Given the following clauses, show *P(John)* must be true by adding *¬ 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.)

W(?z)

Q(John)

P(?y) ∨ ¬ R(?y) ∨ ¬ W(?y)

R(?x) ∨ ¬ Q(?x)

**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.*

1. 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

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

**Problem 7 – Miscellaneous Questions (15 points)**

1. How should *k* be chosen in the *k*-NN algorithm? Circle the answer you feel is best.
   1. a tuning set should be used to evaluate a set of candidates for the value of *k*
   2. any odd number between 1 and 10% of the number of examples   
      in the data set will suffice
   3. one should perform gradient descent to find a local minimum of error
   4. use the value of *k,* chosen from a set of candidates, that leads to the   
      best test-set accuracy

Briefly justify your answer below.

1. 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.

1. Qualitatively draw a (2D) picture of *weight space* where the backpropagation algorithm   
   is likely to
2. do well
3. do poorly

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

Have a good break!