Fundamentals of Artificial Intelligence
COSC 4368
Some Solution Sketches
Final Exam
Monday, May 6, 2p

Name:
SSN:

1. Naïve Bayes and Belief Networks (17 points)
2. FOPL as a Language (10 points)
3. Support Vector Machines (7 points)
4. FOPL Theorem Proving Using Resolution (13 points)
5. Multi-Agent Systems (5 points)
6. Ethics for AI (8 points)
7. Supervised Learning and Neural Networks (13 points)
8. AI in General (7 points)

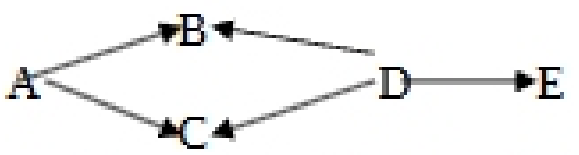
Point Total (out of 73):

Number Grade:

The exam is “open books” and you have 120 minutes to complete the exam.

1) Belief Networks and Naïve Bayes [17]

a) Consider the following belief network that consists of variables A, B, C, D, E all of which have two states {true, false} and whose structure is depicted below is given.



i) Is A|∅ d-separable from E|∅[1]. Given reasons for your answer! [4]

There are two paths:
A-C-D-E and A-B-D-E

The first path is blocked in node C (pattern3 as C in not in evidence) and
The second path is blocked in B (pattern3 as B is not in evidence)
As both paths are block, A and E are independent

ii) Is B|A d-separable from C|A. Give reasons for your answer! [4]

There are two paths: B-D-C and B-A-C
The first path is not blocked in node D, as D is not in evidence
The second path is blocked in A (pattern2 as A is not in evidence) if they do not write that line they still deserve full credit
As the first path is not blocked B|A is not d-separable (independent) from C|A

iii) What advantage you see in using Belief Networks instead of using a naïve Bayesian approach? [4]

Using Belief-networks you can express dependencies between random variables that are not independent of each other, using domain specific knowledge[2]. By doing that BBNs will obtain "better", more accurate predictions than the naïve Bayesian approach, as making conditional independence assumptions that are violated by the observed data will increase prediction errors.

iv) Assume P(D)=0.02, P(S1)=0.2 P(S1|D)=0.4 P(S2)=0.1 P(S2|D)=0.3. Compute P(D|S1,S2) using a Naïve Bayesian approach! [3]

P(D|S1,S2)=0.02*2*3=0.12 No partial credit!

v) What specific conditional independence assumptions are made in your computations for problem iv)? [2]

S1 is independent of S2[1], and S1|D is independent from S2|D[1].

---

1 ∅ represents "no evidence"; question i basically asks if A and E are independent; that is, if P(A∧E)=P(A)*P(E)

2) FOPL as a Language [10]

Express the following natural language statements using first order predicate calculus formulas:

a) Flipper is an intelligent dolphin. [2]
b) There is a dog-catcher in Texas that got bitten by a dog in each county in Texas. [4]
c) No student who took the final exam of COSC 4368 got a grade of 'C' in the final exam. [4]

a) dolpin(Flipper) ^ intelligent(Flipper) no partial credit; using ∃ is wrong!
b) ∃x (dogcatcher(x) ^ lives(x,Texas) ^ ∀c (county(c,Texas)→ ∃d bitten(x,d) ^ located(d,c)))
c) ~∃s (took-final(COSC4368,s) ^ grade(s, COSC4368,Final,C)); might also give full credit for ~∃s grade(s, COSC4368,Final,C))

Can give partial credit up to 1 point for c and up to 2.5 points for b!

3) Support Vector Machines [7]

Assume a support vector machine hyperplane that has been learnt for a dataset having attribute A, B, C is given:

A□2+B□3-C□4-2

e.g. the support vector machine classifies an example ex1 (0,1,1) where 0, 1, 1 are the values for attributes A, B, C as belonging to the negative class as 0*2+1*3-4*1-2=□3.

Assume we have 3 more examples ex2, ex3, ex4 are given for which the hyperplane equation returns -20, 0, +4. What does this tell you about the examples [3].

ex1and ex2 are on the side of the negative class of the hyperplane[1], but ex2 is much further away from the hyperplane than ex1[0.5]; ex3 is located on the hyperplane[1], and ex4 is one the other (positive class) side if the hyperplane [0.5]

Can the fact that the support vector machine equation returns a negative or positive number—and not just a class label of the predicted class—be used for anything useful? [4]

If the values returned for the hyperplane equation for a testing example e is very close to 0 we are much less confident about the correctness of the SVM's prediction compared to the case where plugging the attribute values of e into the hyperplane equation returns large negative or positive values.

Other way to say it: Closeness to 0 can be used to assess a degree of confidence the SVM has with respect to the correctness of the prediction it made for e!