CPSC420-500 Final Exam (12/13/2006, Wed)[1]

Last name: _______________, First name: _______________

Time: **8:00am–10:00am (2 hours)**, Total Points: **100**

Subject	Score
First-order logic	/30
Uncertainty	/20
Probabilistic reasoning	/20
Learning	/40
Natural language processing	/10
Total	/120

- The total adds up to 120, but your final score will be $\min(raw\_score, 100)$. For example, if you solve all problems and get 85 points total, then your score will be $\min(85, 100) = 85$. If you get 112 points total, then your score will be $\min(112, 100) = 100$. So, if you have time, solve all problems!
- Be as **succinct** (i.e., brief) as possible.
- Read the questions carefully to see what kind of answer is expected (*explain blah* **in terms of ...** *blah*).
- Total of __10__ pages, including this cover and the Appendix at the end. **Before starting, count the pages and see if you have all __10__.**
- Problems are on **both** sides of each sheet.
- This is a closed-book, closed-note exam.
- You may rip off the last page (Appendix) to view it while solving the logic problems.

[1] Instructor: Yoonsuck Choe.

1 First-order Logic

Question 1 (10 pts):

Convert the following in to prenex normal form, conjunctive normal form, and then skolemize. Show all steps.

$$\forall x\,[(\forall y(P(x) \to Q(x,y))) \to (\exists z(S(x,z) \wedge P(x)))]$$

Question 2 (10 pts): For each pair of predicates below, give the most general unifier if it exists. (P, Q, R, S are predicates, f, g, h are functions, A, B, C are constants, and w, x, y, z are variables.)

- $P(x, f(B), g(x, f(x)))$
  $P(A, z, g(w, z))$

- $Q(f(x), g(x, w), f(B))$
  $Q(y, g(B, h(A)), y)$

Question 3 (10 pts): Susan, Clyde, and Oscar are elephants (constants). $Pink(\cdot), Gray(\cdot)$, and $Likes(\cdot, \cdot)$ are predicates. x and y are variables.

We know the following:

1. Susan is Pink.
   $Pink(Susan)$
2. Clay is gray and he likes Oscar.
   $Gray(Clyde) \wedge Likes(Clyde, Oscar)$
3. Oscar is either pink or gray (but not both) and he likes Susan.
   $[(\neg Pink(Oscar) \wedge Gray(Oscar)) \vee (Pink(Oscar) \wedge \neg Gray(Oscar))] \wedge Likes(Oscar, Susan)$

Prove that $\exists x \exists y\, [Gray(x) \wedge Pink(y) \wedge Likes(x, y)]$ is a logical consequence of the above, **using resolution**.

2 Uncertainty

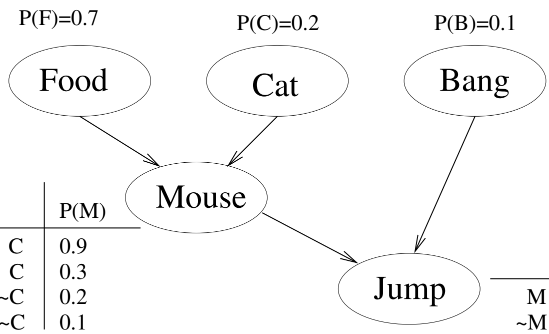
Question 4 (10 pts): Explain why Bayes rule can be helpful when doing diagnostic inference (e.g., inferring the disease causing a symptom: $P(Disease|Symptom)$). Explain in terms of how (1) causal knowledge and (2) prior knowledge can be used.

Question 5 (10 pts): When combining multiple evidence, under what assumption does Bayesian updating become efficient? Example: $P(Cavity|Toothache)$ vs. $P(Cavity|Toothache \wedge Catch)$.

3 Probabilistic Reasoning

Question 6 (10 pts): The elephant jumps ($Jump$) when it hears a bang ($Bang$) or sees a mouse running across the floor ($Mouse$). The mouse come outside when there's food ($Food$), and it will run when the cat appears (Cat).

Given the following belief network describing this scenario, calculate the following joint probability: $P(\neg Bang, \neg Food, Cat, Mouse, Jump)$. You don't need to reduce your answer to a single number.



	P(M)
F C	0.9
~F C	0.3
F ~C	0.2
~F ~C	0.1

	P(J)
M B	0.8
~M B	0.5
M ~B	0.7
~M ~B	0.05

Question 7 (10 pts): (1) Why is $P(Alarm|Earthquake) >> P(Alarm|Earthquake \wedge Burglary)$? (2) What is this called?

4 Learning

Question 8 (10 pts): In decision tree learning, suppose you started with 13 positive and 14 negative examples. After testing with attribute Foo that can have values x, y, and z, you got the following breakdown:

Foo	# Positive	# Negative
x	2	7
y	8	3
z	3	4

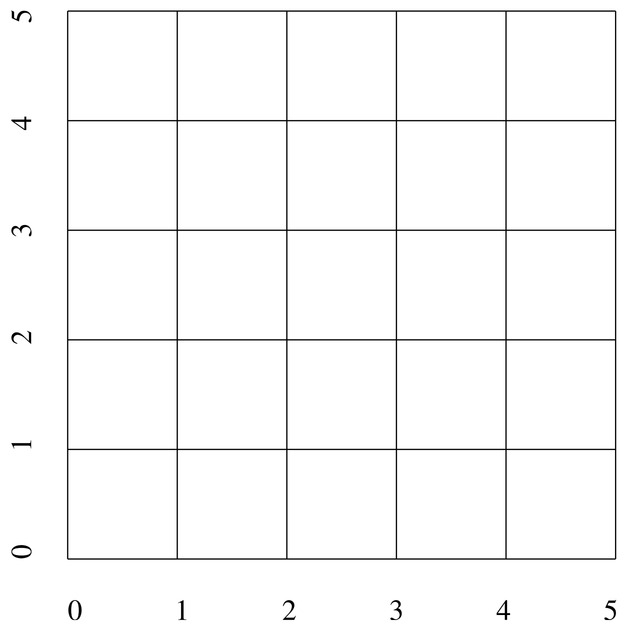
If you test for attribute Bar that can have values 1 or 2, you would get the following breakdown:

Bar	# Positive	# Negative
1	7	7
2	6	7

Testing which attribute will result in higher information gain? Explain in terms of the uncertainty before testing and after testing with each attribute. (**Hint:** You don't need to calculate the exact information gain to answer this question.)

Question 9 (10 pts): Given the following input–target pairs, do you think a single perceptron unit with two inputs and one bias unit will be able to learn the input–target mapping without any error? Use the grid on the right to visualize your answer.

Input 1	Input 2	Target
0	0	Y
2	3	N
4	2	N
3	0	Y
0	2	Y
3	2	N
1	1	Y
0	4	N
1	5	N
2	0	Y
3	1	N



Question 10 (10 pts): Explain how you can compress images using a backpropagation network.

Question 11 (10 pts): Answer the following questions, given the two 50×50 image below. Think of each pixel location as a random variable, and suppose you have a huge collection of images that look like Image A (all being different), and the same for Image B.

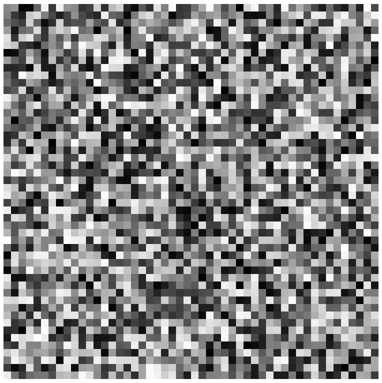


Image A

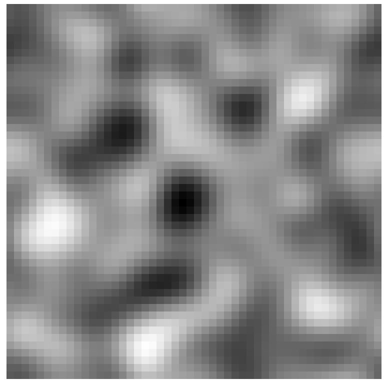


Image B

1. Which one has more structure: Image A or Image B?

2. Which one has more redundancy: Image A or Image B? Explain why.

3. Which one has more independence across pixels: Image A or Image B?

4. Which one can be more ideal for unsupervised learning: Image A or Image B?

5 Natural Language Processing

Question 12 (10 pts): In binary spatter code (BSC), merging is defined as follows, using bit-wise majority rule:

$$M = \langle A + B + C \rangle.$$

(1) What is one useful property of the merging operator? Explain in terms of the relationship within and across M, A, B, C and a random vector X.

(2) In BSC, you can combine the binding operator $\otimes$ and the merging operator $\langle ... + ... + ... \rangle$. What do you get when you do:

$$x \otimes \langle A + B + C \rangle ?$$

(3) Why are the above related to the need for an item memory/clean-up memory? Explain in terms of the following example representing a relation $\mathbf{r}(\mathbf{A}, \mathbf{B})$: Here, we're using $\mathbf{r1}$ to probe the merged structure $\mathbf{R}$ and hope to retrieve $\mathbf{A}$. (**Hint:** Try calculating $\mathbf{X}$.)

$$\mathbf{R} = \langle \mathbf{r} + \mathbf{r1} \otimes \mathbf{A} + \mathbf{r2} \otimes \mathbf{B} \rangle$$

$$\mathbf{X} = \mathbf{r1} \otimes \mathbf{R}$$

Appendix

Note: There is no exam question on this page.

Logic:

- $P \vee Q = Q \vee P$,
  $P \wedge Q = Q \wedge P$ (commutative)
- $(P \vee Q) \vee H = P \vee (Q \vee H)$,
  $(P \wedge Q) \wedge H = P \wedge (Q \wedge H)$,(associative)
- $P \vee (Q \wedge H) = (P \vee Q) \wedge (P \vee H)$,
  $P \wedge (Q \vee H) = (P \wedge Q) \vee (P \wedge H)$ (distributive)
- $P \vee \mathbf{F} = P, P \wedge \mathbf{F} = \mathbf{F}$ (**F**: False)
- $P \vee \mathbf{T} = \mathbf{T}$
  $P \wedge \mathbf{T} = P$ (**T**: True)
- $P \vee \neg P = \mathbf{T}$
  $P \wedge \neg P = \mathbf{F}$
- $\neg(P \vee Q) = \neg P \wedge \neg Q$,
  $\neg(P \wedge Q) = \neg P \vee \neg Q$ (DeMorgan's law)
- $P \rightarrow Q = \neg Q \rightarrow \neg P$ (contrapositive)
- $P \rightarrow Q = \neg P \vee Q$
- $(Qx, F(x)) \vee G = Qx, (F(x) \vee G)$
  $(Qx, F(x)) \wedge G = Qx, (F(x) \wedge G)$
- $\neg(\forall x, F(x)) = \exists x, (\neg F(x))$
  $\neg(\exists x, F(x)) = \forall x, (\neg F(x))$
- $(\forall x, F(x)) \wedge (\forall x, G(x)) = \forall x, (F(x) \wedge G(x))$
  $(\exists x, F(x)) \vee (\exists x, G(x)) = \exists x, (F(x) \vee G(x))$
- $(Q_1 x, F(x)) \vee (Q_2 x, H(x)) = Q_1 x, Q_2 z, (F(x) \vee H(z))$
  $(Q_1 x, F(x)) \wedge (Q_2 x, H(x)) = Q_1 x, Q_2 z, (F(x) \wedge H(z))$

Probability (for Boolean random variables):

1. $P(A, B) = P(A \wedge B) = P(A = \mathbf{True} \wedge B = \mathbf{True})$
2. $P(A|B) = \frac{P(A,B)}{P(B)}$
3. $P(A, B) = P(A|B)P(B)$
4. Bayes rule: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$
5. Condition for independence: $P(A, B) = P(A)P(B)$
6. $P(A, B|C) = P(\underbrace{A, B}|C)$, **not** $P(A, \underbrace{B|C})$