COMP3211 Fundamentals of AI **Name:** ____________

Fall 2019 Final

10/12/2018 **Stu ID:** ____________

Time Limit: 1300 Dec 10 – 1259 Dec 11 HKT (GMT+8)

Instructions:

1. This exam contains 12 pages (including this cover page) and 12 questions.
2. This is an open book exam.
3. Observe the honor code. Write the exam on your own. Do not discuss with any other people.
4. Upload your answers as a pdf file (YourStudentID.pdf) to canvas by 1259 on Dec 11 HKT. You can use handwriting or any document editor but must scan or convert your answers to a PDF file.

Grade Table (for teacher use only)

Question	Points	Score
State Machines	6	
Game Theory	6	
Auction	6	
CNF	8	
Representation in PL	10	
Representation in FOL	15	
Uncertainty	12	
MDP	10	
Admissibility	4	
Perceptron Learning and GSCA Rule Learning	12	
Perceptrons/Linearly Separable Functions	6	
Machine Learning/General Discussion	5	
Total:	100	

Question 1: State Machines .... 6 points

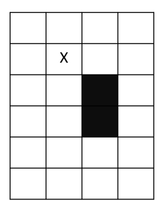
Consider again the boundary-following robot but with only 4 sensors: s_1, s_2, s_3, s_4, which correspond to north, east, south, and west, respectively. The robot again has four actions: *south, north, west, east.* We also define the following features:

- $w_i = s_i$, where $i = 1, 2, 3, 4$.
- $w_5 = 1$ iff at the previous step, the robot moved north.
- $w_6 = 1$ iff at the previous step, the robot moved south.

The production system is as follows:

$$\begin{aligned}
&w_2 w_1 \rightarrow \textit{south},\\
&w_2 w_3 \rightarrow \textit{north},\\
&w_2 w_5 \rightarrow \textit{north},\\
&w_2 w_6 \rightarrow \textit{south},\\
&w_5 \rightarrow \textit{south},\\
&w_6 \rightarrow \textit{north},\\
&w_2 \rightarrow \textit{north},\\
&1 \rightarrow \textit{east}.
\end{aligned}$$

The robot's current position is indicated in the figure below with an "X". Current feature values are: $w_1 = 0, w_2 = 0, w_3 = 0, w_4 = 0, w_5 = 1, w_6 = 0$.



List the next 4 moves of the robot controlled by this production system.

Solution: south, south, south, north, north, north, south, south

The 1st and 4th moves: 2 points each; The 2nd and 3rd moves: 1 point each.

Question 2: Game Theory .... 6 points

There are two firms serving the same market. They have constant average costs of \$2 per unit. The firms can choose either a high price (\$10) or a low price (\$5) for their output. When both firms set a high price, total demand is 10,000 units which is split evenly between the two firms. When both set a low price, total demand is 18,000, which is again split evenly. If one firm sets a low price and the other a high price, the low priced firm sells 15,000 units, the high priced firm only 2,000 units. What pricing decisions would the firms make?

Solve the problem by formalizing the strategic situation as a game in normal form between these two firms (constructing the pay-off matrix) and find a solution by computing the pure Nash equilibria.

Solution: The game in normal form (payoffs are in thousands):

	\$10	\$5
\$10	40, 40	16, 45
\$5	45, 16	27, 27

There is a unique Nash equilibrium: (\$5, \$5), so both will charge \$5.

4 points for game matrix, 2 points for NE.

Question 3: Auction .... 6 points

Consider the single item first-price auction with two bidders B_1 and B_2, and assume that when there is a tie, B_2 will win. Suppose B_i values the item x_i, $i = 1, 2$, and the information is common knowledge (thus B_1 knows that B_2's value is x_2, and vice versa and so on). Suppose further that each can bid with any integers in the interval $[0, 100]$.

- Make this auction into a game in normal form $(\{B_1, B_2\}, R_1, R_2, u_1, u_2)$ by defining R_i (the set of pure strategies for player B_i) and u_i (player B_i's utility function). You can assume that both players are risk neutral.
- Suppose $0 \leq x_i \leq 100$ for both $i = 1, 2$. What are the Nash equilibria of your game? You need to consider all possible cases for x_1 and x_2.

Solution:

- $R_1 = R_2 = \{0, ..., 100\}$. $u_1(b_1, b_2) = x_1 = b_1$ if $b_1 > b_2$ and 0 otherwise. $u_2(b_1, b_2) = x_2 - b_2$ if $b_2 \geq b_1$ and 0 otherwise.
-
  - If $x_1 < x_2$, then (x_1, x_1) and $(x_1 - 1, x_1 - 1)$ are the NE.
  - If $x_1 = x_2$, then (x_1, x_1), $(x_1, x_1 - 1)$ and $(x_1 - 1, x_1 - 1)$ are the NE.
  - If $x_1 = x_2 + 1$, then $(x_1 - 1, x_1 - 2)$, $(x_1 - 1, x_1 - 1)$ and $(x_1, x_1 - 1)$ are the NE. (or $(x_2, x_2 - 1)$, (x_2, x_2) and $(x_2 + 1, x_2)$).
  - If $x_1 > x_2 + 1$, then $(x_2, x_2 - 1)$ and $(x_2 + 1, x_2)$ are the NE.

1 point for R, 1 point for u.
1 point for each case.

Question 4: CNF . . . . 8 points

Consider the following KB with two formulas:

$$(p \supset q) \equiv r,$$
$$(p \supset q) \equiv \neg r.$$

Answer the following questions:

1. (4 pts) Convert the KB to a set of clauses.
2. (4 pts) Use resolution to prove that the KB is contradictory.

Solution: The first formula gives the following clauses:

$$[p \vee r, \neg q \vee r, \neg p \vee q \vee \neg r]$$

and the second the following one:

$$[p \vee \neg r, \neg q \vee \neg r, \neg p \vee q \vee r]$$

Do pairwise resolution to get rid of r in the two set of clauses and we get

$$p, \neg q, \neg p \vee q$$

Do resolution with 1st and 3rd and then with the 2nd to yield the empty clause.

Question 5: Representation in PL . . . . 10 points

Letâs consider a propositional language where

- A=“Angelo comes to the party”,
- B=“Bruno comes to the party”,
- C=“Carlo comes to the party”,
- D=“David comes to the party”.

Formalize the following sentences:

1. If David comes to the party then Bruno and Carlo come too.
2. David comes to the party if and only if Carlo comes and Angelo doesnât come.
3. Carlo comes to the party only if Angelo and Bruno do not come.
4. Angelo, Bruno and Carlo all come to the party if and only if David doesnât come.
5. If neither Angelo nor Bruno come to the party, then David comes only if Carlo comes.

Solution:

1. $D \supset B \wedge C$.
2. $D \equiv (C \wedge \neg A)$.
3. $C \supset \neg A \wedge \neg B$.
4. $(A \wedge B \wedge C \equiv \neg D)$
5. $\neg A \wedge \neg B \supset (D \supset C)$.

2 points each.

Question 6: Representation in FOL . . . . 15 points

Consider a world with balls and boxes, and the following relations:

- $ball(x)$: x is a ball.
- $box(x)$: x is a box.
- $color(x, y)$: x has color y.
- $in(x, y)$: x is in y.

Represent the following statements in first-order logic:

1. (3 pts) Some balls are not in any box.
2. (3 pts) Every box has some balls in it.
3. (3 pts) A ball cannot have more than one color.
4. (3 pts) There can be no more than two balls in any box.
5. (3 pts) A box can only contain balls with the same color as itself (boxes also have colors).

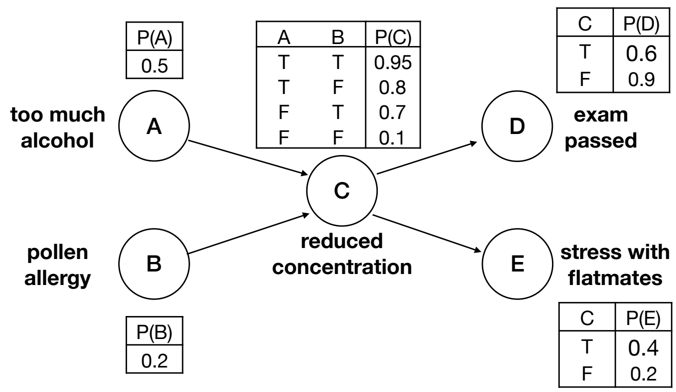
Solution:

1. $\exists x.\ ball(x) \wedge \neg(\exists y. box(y) \wedge in(x, y))$
2. $\forall x.\ box(x) \supset (\exists y. ball(y) \wedge in(y, x))$
3. $\forall x, y, z.\ ball(x) \wedge color(x, y) \wedge color(x, z) \supset y = z$
4. $\forall x, a, b, c.\ box(x) \wedge ball(a) \wedge ball(b) \wedge ball(c) \wedge in(a, x) \wedge in(b, x) \wedge in(c, x) \supset (a = b \vee a = c \vee b = c)$
5. $\forall x, y, a, b.\ box(x) \wedge ball(y) \wedge in(y, x) \wedge color(x, a) \wedge color(y, b) \supset a = b$

3 points each. Remove 1 point for each mistake.

Question 7: Uncertainty . . . . 12 points

Consider the following Bayesian network. Imagine you are at a birthday party of a friend on Sunday and you have an exam on Monday. If you drink too much alcohol at the birthday party, you most likely have problems concentrating the next day, which would reduce the probability that you pass the exam. Another consequence of the reduced concentration might be increased stress with your flatmates, because, e.g., you forget to turn off the radio or the stove. Lack of concentration might also be caused by your pollen allergy, which you suffer from on some summer days.



1. (2 pts) Are A (*too much alcohol*) and B (*pollen allergy*) independent given D (*exam passed*)? Explain your answer using D-separation.
2. (5 pts) Compute the probability of B (*pollen allergy*) given C (*reduced concentration*) is true: $P(B|C)$. There is no need to perform numerical calculations. As long as your formula is right, you will get the full mark.
3. (5 pts) Compute the probability of A (*too much alcohol*) given C (*reduced concentration*) is true and E (*stress with flatmates*) is not true: $P(A|C, \neg E)$. Again, there is no need to perform numerical calculations.

Solution:

1. A and B are not independent given D: there is just one path between them: $A \rightarrow C \leftarrow B$, which cannot be separated by C because the given evidence D is a descendant of C.

2.

$$\begin{aligned}
P(B|C) &= P(B, C)/P(C). \\
P(B, C) &= P(B, C, A) + P(B, C, \neg A) \\
&= P(A)P(B)P(C|A, B) + P(\neg A)P(B)P(C|\neg A, B). \\
P(C) &= P(A, B, C) + P(A, \neg B, C) + P(\neg A, B, C) + P(\neg A, \neg B, C).
\end{aligned}$$

3.

$$\begin{aligned}
P(A|C,\neg E) &= P(A,C,\neg E)/P(C,\neg E). \\
P(A,C,\neg E) &= P(A,C,\neg E,B) + P(A,C,\neg E,\neg B) \\
&= P(A)P(B)P(C|A,B)P(\neg E|C) + \\
&\quad\; P(A)P(\neg B)P(C|A,\neg B)P(\neg E|C). \\
P(C,\neg E) &= P(A,B,C,\neg E) + P(A,\neg B,C,\neg E) + \\
&\quad\; P(\neg A,B,C,\neg E) + P(\neg A,\neg B,C,\neg E).
\end{aligned}$$

For question 2 and 3, two points for Bayes formula, three points for total probability formula.

Question 8: MDP .... **10 points**

Consider a 4×3 stochastic grid world laid out in the figure below (the crossed-out cell is an obstacle). The agent starts in state (1,1), and has four available actions: *North*, *South*, *West*, *East*. For each action, the agent goes forward with 0.8 probability, goes left and right with 0.1 probability respectively. If there is a wall, the agent stays at current location. For example, if the agent move *East* in cell (1,3), then she'll end up with 0.8 probability in cell (2,3), 0.1 probability in (1,2) (goes left instead), and 0.1 probability in the same cell (1,3) (goes right, which is a wall). At the terminating states (4,1) and (4,2), the only action is *Exit*, after which the agent will be in state *END*. The reward function is defined as follows:

$$R(s,a,s') = \begin{cases} -1, & a = \mathit{Exit} \text{ and } s = (4,2) \\ +1, & a = \mathit{Exit} \text{ and } s = (4,1) \\ -0.1, & \text{otherwise} \end{cases}$$

Assume that the discount factor $\gamma = 1.0$.

We perform value iteration on the MDP. The value of each state after 2 iterations are given in the right grid in the following figure.

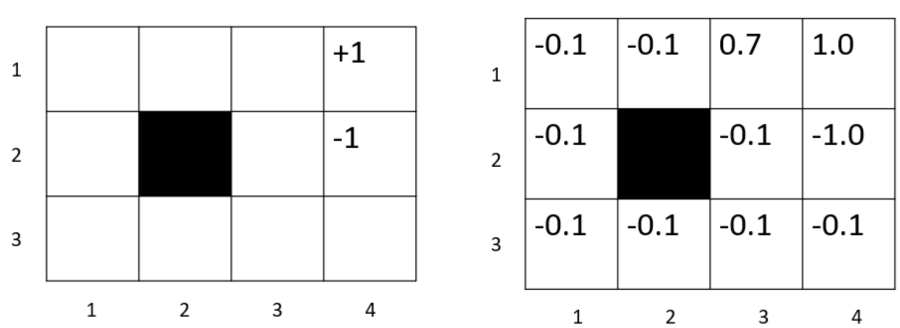


Figure 1: Left: the 4x3 grid world; right the value of each state after 2 iterations of value iteration.

Now calculate the values of the states (2,1), (3,1), (3,2) after one more iteration.

Solution:

$$V^3(s) = \max_{a \in Actions(s)} \sum_{s'} T(s,a,s')(R(s,a,s') + \gamma V^2(s')) \quad (1)$$

At (2,1), the best action is to move east:

$$\begin{aligned}
V^3((2,1)) &= T((2,1), East, (2,1))(R((2,1), East, (2,1)) + V^2((2,1))) && (2)\\
&\quad + T((2,1), East, (3,1))(R((2,1), East, (3,1)) + V^2((3,1))) && (3)\\
&= 0.2 \times ((-0.1) + (-0.1)) + 0.8 \times ((-0.1) + 0.7) && (4)\\
&= 0.44 && (5)
\end{aligned}$$

At (3,1), the best action is to move east:

$$\begin{aligned}
V^3((3,1)) &= T((3,1), East, (4,1))(R((3,1), East, (4,1)) + V^2((4,1))) && (6)\\
&\quad + T((3,1), East, (3,1))(R((3,1), East, (3,1)) + V^2((3,1))) && (7)\\
&\quad + T((3,1), East, (3,2))(R((3,1), East, (3,2)) + V^2((3,2))) && (8)\\
&= 0.8 \times ((-0.1) + 1) + 0.1 \times ((-0.1) + 0.7) && (9)\\
&\quad + 0.1 \times ((-0.1) + (-0.1)) && (10)\\
&= 0.76 && (11)
\end{aligned}$$

At (3,2), the best action is to move north:

$$\begin{aligned}
V^3((3,2)) &= T((3,2), North, (3,1))(R((3,2), North, (3,1)) + V^2((3,1))) && (12)\\
&\quad + T((3,2), North, (3,2))(R((3,2), North, (3,2)) + V^2((3,2))) && (13)\\
&\quad + T((3,2), North, (4,2))(R((3,2), North, (4,2)) + V^2((4,2))) && (14)\\
&= 0.8 \times ((-0.1) + 0.7) + 0.1 \times ((-0.1) + (-0.1)) && (15)\\
&\quad + 0.1 \times ((-0.1) + (-1)) && (16)\\
&= 0.35 && (17)
\end{aligned}$$

0.44; 0.76; 0.35

Since this question has some problems, you will get full marks if the right value update formula is used.

Question 9: Admissibility . . . . 4 points

Assume you are given three admissible heuristic functions for a given search problem, for simplicity, call them h_1, h_2, and h_3. For each of the following combinations, explain whether or not the resulting h function is admissible.

1. (2 pts) $A : h(n) = h_1(n) + h_2(n) + h_3(n)$
2. (2 pts) $B : h(n) = (h_1(n) + h_2(n) + h_3(n))/3$

Solution: Since h_1, h_2, h_3 are admissible, $0 \le h_i(n) \le h^*(n)$, for $i = 1, 2, 3$, where $h^*(n)$ is the true cost to a nearest goal.
Then $0 \le h_1(n) + h_2(n) + h_3(n) \le 3h^*(n)$, but $h_1(n) + h_2(n) + h_3(n)$ can be greater than $h^*(n)$. For example, when $h_1(n) = h_2(n) = h^*(n)$, we can obtain $h_1(n)+h_2(n)+h_3(n) \ge 2h^*(n) \ge h^*(n)$. So A is not always admissible.
Since $0 \le h_i(n) \le h^*(n)$, for $i = 1, 2, 3$, $0 \le (h_1(n) + h_2(n) + h_3(n))/3 \le h^*(n)$, so B is admissible.

2 points each.

Question 10: Perceptron Learning and GSCA Rule Learning . . . . 12 points

Consider the following data set:

ID	x_1	x_2	x_3	x_4	OK
1	1	0	1	0	No
2	0	0	1	0	No
3	1	0	0	0	Yes
4	1	1	0	1	Yes
5	0	1	1	1	No

where x_1, x_2, x_3 and x_4 are some features that should not concern us here.

1. (5 pts) Use these four instances to train a single perceptron using the error-correction procedure. Use the learning rate = 1, and the initial weights all equal to 0. Recall that the threshold is considered to be a new input that always have value "1". Please give your answer by filling in the following table, where weight vector (w_1, w_2, w_3, w_4, t) means that w_i is the weight of input x_i, and t is the weight for the new input corresponding to the threshold. Stop when the weight vector converges. If it doesn't converge, explain why not.

ID	Weight vector (w_1, w_2, w_3, w_4, t)
Initial	(0, 0, 0, 0, 0)
1	
2	
3	
4	
5	
1	
2	
3	
4	
5	
1	
2	
3	
4	
5	

2. (2 pts) What is the Boolean function corresponding to your perceptron?
3. (5 pts) From the same training set, apply the GSCA algorithm to try to learn a set of rules. Give the set of rules if it succeeds. If it fails to learn a set of rules, explain why it failed.

Solution:

1.

ID	Extended input	Weight vector	Sum	Actual	Desired
Initial		0, 0, 0, 0, 0			
1	1, 0, 1, 0, 1	-1, 0, -1, 0, -1	0	1	0
2	0, 0, 1, 0, 1	-1, 0, -1, 0, -1	-2	0	0
3	1, 0, 0, 0, 1	0, 0, -1, 0, 0	-2	0	1
4	1, 1, 0, 1, 1	0, 0, -1, 0, 0	0	1	1
5	0, 1, 1, 1, 1	0, 0, -1, 0, 0	-1	0	0
1		0, 0, -1, 0, 0	-1	0	0
2		0, 0, -1, 0, 0	-1	0	0
3		0, 0, -1, 0, 0	0	1	1

It converges to give the weight vector $(0, 0, -1, 0)$ with the threshold 0.

2. Boolean function: $\neg x_3$, i.e. $\overline{x_3}$.

3. We begin with: $true \supset OK$.
Choose the feature that yields the largest value of r_α:
$r_{x_1} = 2/3, r_{x_2} = 0.5, r_{x_3} = 0, r_{x_4} = 0.5$.
So we choose x_1, this will generate : $x_1 \supset OK$.
This rule still covers the negative instance ID1, so we still need to narrow it. This time we have $r_{x_2} = 1, r_{x_3} = 0, r_{x_4} = 1$. We can add either of $\{x_2, x_4\}$, say x_2,: $x_1 \wedge x_2 \supset OK$. This rule covers only positive instances, so we have learned one rule. To learn the next one, we eliminate data with ID4 and continue like above.
We begin with: $true \supset OK$.
Choose the feature that yields the largest value of r_α:
$r_{x_1} = 0.5, r_{x_2} = 0, r_{x_3} = 0, r_{x_4} = 0$.
So we choose x_1, this will generate : $x_1 \supset OK$.
This rule still covers the negative instance ID1, so we still need to narrow it. This time we have $r_{x_3} = 0$, and x_2 and x_4 are not applicable. So we add x_3: $x_1 \wedge x_3 \supset OK$, but there are no positive examples covered by this rule. So we track back to $x_1 \supset OK$ and replace it by $x_2 \supset OK$ or $x_3 \supset OK$ or $x_4 \supset OK$. All of them cover no positive examples. So it backtrack to the beginning and abort (or loop for ever). So GSCA only leanrs one rule, for the reason that this training set cannot be learned without making use of negative literals like $\neg x_3$ in the rules.

Question 1, 1 point for each entry.

Question 2, your Boolean function should correspond to your perceptron, not necessarily the original data.

Question 3, 3 points for the rule $x_1 \wedge x_2 \supset OK$, 2 points for the failed GSCA.

Question 11: Perceptrons/Linearly Separable Functions . . . . 6 points

Assume three boolean variables x, y and z. If $f(x, y, z)$ is a boolean function of x, y, z, then the projection $f(0, y, z)$ is the boolean function of y and z with x set to 0 (false).

Answer the following questions

1. (3 pts) Is there a linearly separable function $f(x, y, z)$ such that $f(0, y, z)$ is non-linearly separable? If yes, give an example. If no, explain why.
2. (3 pts) Is there a non-linearly separable function $f(x, y, z)$ such that $f(0, y, z)$ is linearly separable? If yes, give an example. If no, explain why.

Solution:

1. No. If $f(x, y, z)$ is linearly separable, then $f(0, y, z)$ must also be linearly separable. Geometrically, if a plane separates $f(x, y, z)$, then the intersection of the plane and the plane of $x = 0$ separates $f(0, y, z)$. Algebraically, if $f(x, y, z)$ is computed by $w_1x + w_2y + w_3z \geq \theta$, then $f(0, y, z)$ is $w_2y + w_3z \geq \theta$.

Remove 1 point if the explanation is not rigorous or clear enough.

2. Yes. Many examples. One of them is: $x(y \oplus z)$ ($\oplus$ is the exclusive or) is non-linearly separable. But $f(0, y, z)$ is $false$ which is of course linearly separable. Notice that the same argument as before would show that if $f(x, y, z)$ is linearly separable, so is $f(1, y, z)$. Since for our example $f(1, y, z)$ is the exclusive or between y and z, which is not linearly separable, so this $f(x, y, z)$ is not linearly separable.

Question 12: Machine Learning/General Discussion .... 5 points

Consider the statement: "The current machine learning models are brittle and their results not explainable". Answer the following questions:

1. (3 pts) Give an example to illustrate this statement.
2. (2 pts) Do you think this impedes the usefulness of the current machine learning models? Explain your answer.

Solution: Give marks for reasonable answers.