上 海 交 通 大 学 卷

(2019−2020 Academic Year/Fall Semester)

Class No. ____________ Student ID No. ____________

Name in English/Pinyin: ____________ Name in Hanzi, if applicable: ____________

VE492 Introduction to Artificial Intelligence

Final Exam

August 6, 2020, 8:00am-9:40am

The exam paper has **9** pages in total.

Exam rules and information:

- The four main sections of this exam are independent.
- This is an open-book and open-note exam.
- No communication devices/apps are allowed. These include cell phones and smart watches.
- The points given to each question are indicative. They may be adjusted.

Instructions for submission:

- You will provide your answers in the file `answer.txt`. In that file, you will add your student ID number and your answers by replacing and removing the items in brackets **(brackets included)**. Also, read very carefully the instructions in each question, which give additional information to format your answers.
- When the time is up, submit this txt file on Canvas. You will have 10 minutes to submit on Canvas (Quiz/Final Exam) and accept the Honor Code pledge. Late submission will be penalized.

You are to abide by the University of Michigan-Shanghai Jiao Tong University Joint Institute (UM-SJTU JI) honor code. Please sign below to signify that you have kept the honor code pledge.

THE UM-SJTU JI HONOR CODE

I accept the letter and spirit of the honor code:
I have neither given nor received unauthorized aid on this examination, nor have I concealed any violations of the Honor Code by myself or others.

Signature: Log in on Canvas to sign via the corresponding Quiz.

1 Bayes Net

1.1 Small Bayes Net

1. (2 points) Provide the formula of the joint distribution over all the variables given by the Bayes net depicted in Figure 1. Your formula should list the terms in alphabetic order for CPTS and the variables that we conditioned on. Here is an example of format for the answer:

```
P(A|C)P(B)P(C)P(D|C)P(E|B,D)P(F|D,E)P(H|B)
```

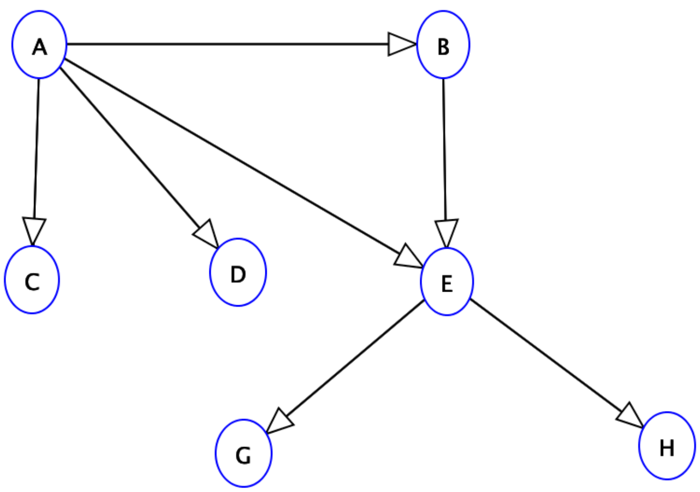

Figure 1: Small Bayes net

2. (2 points) Assuming that all the variables are ternary (i.e., their domain is of size 3), provide the number of degrees of freedom of the Bayes net depicted in Figure 1. Here is an example of format for the answer:

```
42
```

3. (5 points) Run the variable elimination (VE) algorithm on the Bayes net depicted in Figure 1 to compute $\mathbb{P}(A|H = h)$ using the alphabetic order to choose the hidden variables. Provide the list of the sizes (i.e., number of variables) of the factors obtained at the end of each iteration in VE. We do not consider the factor generated after the join operation. Here is an example of format for the answer:

```
2 1 2 1 2
```

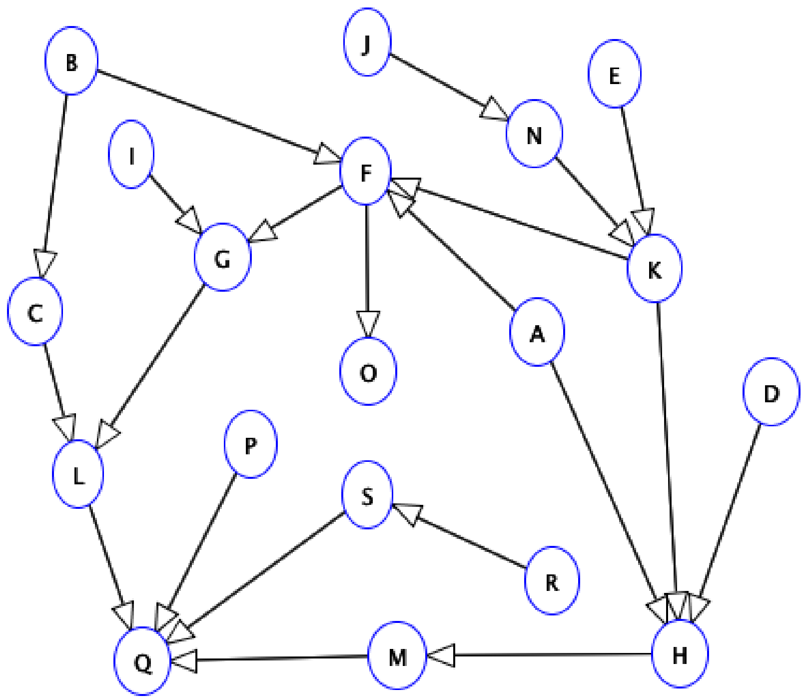

Figure 2: Bayes net for (un)conditional independences.

1.2 Conditional Independences

Answer the following questions about (un)conditional independences based on the Bayes net depicted in Figure 2. Your answer needs to be "Yes" or "No" (without the quotes). If your answer is "Yes", you don't need to justify furthermore. If your answer is "No", provide a path that is active. Therefore, the answer for one question should be either:

```
Yes
```

or

```
No BCLQM
```

where BCLQM represents an active path that goes from B to M, which are the two variables that we are checking for independence. Provide all the answers in the same file, which should therefore contain exactly 5 lines. If you decide to skip one question, leave the corresponding line blank.

1. (3 points) $B \perp\!\!\!\perp H$?
2. (3 points) $F \perp\!\!\!\perp M|L$?
3. (3 points) $B \perp\!\!\!\perp D|A, L, O$?
4. (3 points) $C \perp\!\!\!\perp H|F, L, Q$?

5. (3 points) $L \perp\!\!\!\perp K, R | C, G$?

6. (3 points) What is the cutset of the Bayes net of Figure 2? Provide your answer as a sequence of random variables in alphabetic order.

2 Markov Models

Consider the hidden Markov model with three states and two observations, which is defined by the following elements:

- initial distribution: (1/2 1/2 0)
- transition probability matrix:

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \end{pmatrix}$$

  where P_{ij} is the probability of moving to jth state if we were in the ith state.

- emission probability matrix:

$$Q = \begin{pmatrix} 1/2 & 1/2 \\ 1/3 & 2/3 \\ 1/4 & 3/4 \end{pmatrix}$$

  where Q_{ij} is the probability of observing the jth observation when in the ith state.

We assume that we start obtaining observations from the initial state.

1. (4 points) Compute the probability of being in each state after two transitions. You will give the three probabilities as irreducible fractions separated by a space. For instance,

```
1/2 1/3 1/6
```

2. (4 points) Compute its stationary distribution. Use the same format for the answer as in the first question.

3. (3 points) Compute the probability of observing each observation after three transitions. You can provide the values with two decimals (possibly rounded).

4. (5 points) Assume that we have observed $h = (o_1, o_2, o_1, o_1)$ after three transitions. What is the probability of the current state given the observations? You can provide the values with two decimals (possibly rounded).

5. (2 points) Using the sequence of observations h, what is the probability of the next state given h? You can provide the values with two decimals (possibly rounded).

6. (3 points) Using the sequence of observations h, what is the probability of the state after two transitions given h? You can provide the values with two decimals (possibly rounded).

7. (5 points) Using the sequence of observations h, what is the most likely sequence of states given h? List the indices of the states separated by a space. For instance,

```
1 2 3 2
```

3 Machine Learning

3.1 Bayes Rule

Bayes rule plays a fundamental role in probabilistic reasoning and machine learning. Let's apply it in a simple problem:

A company has three machines M1, M2, and M3, which produce identical items. From the engineers' past experience, 5% of items from M1 are faulty, 3.5% of items from M2 are faulty, and 2.5% of items from M3 are faulty. On a given day, M1 produces 15% of the total output, M2 produces 40%, and M3 produces the remainder. A randomly selected item is found to be faulty.

1. (2 points) What is the probability that an item is faulty? You should provide your answer as a percentage with 2 decimals (possibly rounded), for instance: `2.51`.

2. (2 points) What is the probability that the faulty item comes from M1? Provide your answer in the same format as the previous question.

3. (2 points) What is the probability that the faulty item comes from M2? Provide your answer in the same format as the previous question.

4. (2 points) What is the probability that the faulty item comes from M3? Provide your answer in the same format as the previous question.

3.2 Learning an HMM

We want to estimate the parameters of an HMM that has 3 states and 2 observations. We have collected the following data listed in Table 1, where the numbers correspond to the indices of the states or observations.

state	2	1	3	3	1	2	1	2	2	1	3	1	1	2	2
observation	1	2	1	2	1	1	1	2	2	1	2	1	1	2	1

Table 1: Data for HMM

1. (3 points) Provide the maximum likelihood estimator of the transition matrix. You will write the matrix in the following format:

   ```
   (a b c ; d e f ; g h i)
   ```

   where `a, b, c` corresponds to the first row of the matrix. Add a space before and after the semi-colons. You will provide the values with two decimals (possibly rounded).

2. (3 points) Provide the maximum likelihood estimator of the emission matrix. Use the same format as the first question.

3. (3 points) Provide the Laplace estimator with $k = 1$ of the transition matrix. Use the same format as the first question.

4. (3 points) Provide the Laplace estimator with $k = 1$ of the emission matrix. Use the same format as the first question.

3.3 Perceptron

Assume that we have collected the following dataset of 2D points:

x	y	Class
0	1	1
2	0	-1
1	0	1
1	1	-1

Table 2: Simple dataset

1. (6 points) Provide the weights of a perceptron model that classifies correctly all the points of Table 2. You will give the weight for x first, then y, and finally for the bias, separated by a space. For each value, provide up to two decimals. Here is an example of format for the answer:

   ```
   1 0 -3.24
   ```

2. (4 points) Assume that the class of $(1, 0)$ becomes -1 and that of $(2, 0)$ becomes 1. What is the smallest multi-layer perceptron that can classify correctly this new problem? Provide first the minimal number of **hidden** layers and then the minimal numbers of nodes per layer, separated by a space. Here is an example format for the answer:

   ```
   2 3 1
   ```

where 2 is the number of hidden layers, 3 is the number of nodes in the first hidden layer, and 1 is the number of nodes in the second hidden layer.

3. (2 points) More generally, assume that you have a dataset of 1512 data points in dimension 25, distributed in 5 classes. How many nodes do you need in the output layer of a multi-layer perceptron using a softmax for its output?

4 Logic

4.1 Propositional Logic

1. (2 points) Write the following sentence in CNF:

$$(A \vee B) \Leftrightarrow ((C \wedge \neg D) \Rightarrow (\neg A \vee D))$$

2. (2 points) Is the previous sentence satisfiable? Answer ”yes” or ”no” without the quotes. An incorrect answer will receive a one-point deduction.

3. (1 point) Is the previous sentence valid? Answer ”yes” or ”no” without the quotes. An incorrect answer will receive a one-point deduction.

4. (1 point) Is the previous sentence unsatisfiable? Answer ”yes” or ”no” without the quotes. An incorrect answer will receive a one-point deduction.

4.2 First Order Logic

We consider a 8x8 grid world where there are three types of objects: triangle, square, or pentagon. An object can be small, medium or large. Objects are positioned in one of the 64 cells of the grid. Each cell is indexed by (i, j) where i is the row's index and j is the column's index. The cell on the lower left has index $(1, 1)$ and indices increase when moving up or right. Translate the sentences below in first-order logic using the following predicates (you may not need to use all of them):

- triangle(x) is true if x is a triangle
- square(x) is true if x is a square
- pentagon(x) is true if x is a pentagon
- small(x) is true if x is small
- medium(x) is true if x is medium
- large(x) if true if x is large
- leftof(x,y) is true if the index of the row of x is smaller than that of y

- rightof(x,y) is true if the index of the row of x is larger than that of y
- behindof(x,y) is true if the index of the column of x is larger than that of y
- frontof(x,y) is true if the index of the column of x is smaller than that of y

All the relation (e.g., smaller or larger) are understood as strict.

To write your sentences, you will follow the syntax defined in class. In addition, you will respect the following rules:

- The logic connectives $\neg, \wedge, \vee, \Rightarrow, \Leftrightarrow$ are written respectively `~`, `/\`, `\/`, `=>`, `<=>`.
- The quantifiers $\forall, \exists$ are written respectively A, E.
- If you need variables, use first x, then y, and finally z.
- We assume the following order of precedence (in decreasing order): $\neg, \wedge, \Rightarrow, \forall$. Operation $\vee$ has the same precedence as $\wedge$, $\Leftrightarrow$ has the same precedence as $\Rightarrow$, and $\exists$ the same as $\forall$. Therefore, parentheses may be needed when the previous pairs both appear in a sentence).

Here is an example of a random well-formed sentence:

```
A x,y triangle(x) /\ triangle(y) <=> ~E z pentagon(z) /\ (behindof(x,z) => leftof(y,z))
```

Translate all the following statements in sentences in first-order logic:

1. (2 points) A square is medium.
2. (2 points) All squares are medium.
3. (2 points) Some squares are not medium.
4. (2 points) There are at least two triangles.
5. (2 points) There are at most two triangles.
6. (2 points) There are exactly two triangles.
7. (2 points) Every small square is behind a large square.
8. (2 points) Any object that doesn't have anything behind it is large.
9. (2 points) Using the existing predicates, define lessfig(x, y), which means that "x is a smaller figure than y" (in terms of number of faces). The definition should look like this:

   ```
   lessfig(x,y) <=> ...
   ```

   In your answer, provide only the part written as "...".