FINAL EXAM
FIRST SEMESTER SESSION 2011/2012

COURSE CODE / NAME : STIN2063/MACHINE LEARNING
DATE : 31 DECEMBER 2011 (SATURDAY)
TIME : 9.00 AM – 11.30 AM (2 ½ HOURS)
VENUE : DSB K.MAS

INSTRUCTION :

1. This book script contains TWO (2) sections. SECTION A contains ELEVEN (11) questions and SECTION B contains TWO (2) questions in TWELVE (12) printed pages, excluding the cover page.
2. Answer ALL the questions in Section A in the spaces provided in this booklet and Section B in separate answer sheets.
3. You are NOT ALLOWED to remove the exam paper from the examination hall.

MATRIC NO : ________________________________  (with word)

IDENTIFICATION CARD NO. :

LECTURER :

GROUP :  TABLE NO. :

DO NOT OPEN THIS EXAMINATION PAPER UNTIL INSTRUCTED

CONFIDENTIAL
SECTION A: STRUCTURED QUESTIONS (60 MARKS)

1. A computer program is said to learn from experience E with respect to some task T and some performance measure P if its performance on T, as measured by P, improves with experience E. Suppose we feed a learning algorithm with a lot of historical weather data, and have it learn to predict weather. What would be a reasonable answer for P?

   (2 marks)

2. Here are TWO (2) potential real-world application tasks for machine learning:

   a) The amount of rain that falls in a day is usually measured in either millimeters (mm) or inches. Suppose you use a learning algorithm to predict how much rain will fall tomorrow.

   b) Suppose you are working on stock market prediction, and you would like to predict whether or not a particular stock's price will be higher tomorrow than it is today. You want to use a learning algorithm for this.

Pick ONE (1) of the tasks and state how you would define it as a well-posed machine learning problem.

   (3 marks)
3. Write an algorithm called "FIND-G" to find a maximally-general consistent hypothesis. You can assume the data will be noise-free and that the target concept is in the hypothesis space. (3 marks)

4. Outline the steps to proof that Find-G will never fail to cover a positive example in the training set. (3 marks)
5. Suppose we have a consistent learner with a hypothesis space restricted to conjunctions of exactly 8 attributes, each with values {true, false, don't care}.

   a) What is the size of this learner's hypothesis space? (2 marks)

   b) Compute the solution for the number of examples sufficient to learn with probability at least 95% an approximation of any hypothesis in this space with error of at most 10%. (3 marks)

6. Describe THREE (3) main steps in the basic decision tree learning algorithm. (3 marks)
7. Let the weights of a two-input perceptron be: \( w_0 = 0.2, \ w_1 = 0.5 \) and \( w_2 = 0.5 \). Assuming that \( x_0 = 1 \), what is the output of the perceptron when:

a) \( X_1 = -1 \) and \( x_2 = -1 \)?  

(1 mark)

b) \( X_1 = -1 \) and \( x_2 = 1 \)?  

(1 mark)

c) Letting \( W_0 = -0.2 \) and keeping \( x_0 = 1 \), \( w_1 = 0.5 \) and \( w_2 = 0.5 \), what is the perceptron output when:

i) \( X_1 = 1 \) and \( x_2 = -1 \)?  

(1 mark)

ii) \( X_1 = 1 \) and \( x_2 = 1 \)?  

(1 mark)
d) Here is a regression tree with leaf nodes denotes A, B and C:

\[\begin{align*}
X \leq 5 : & \text{ A} \\
X > 5 : & \\
| \quad X \leq 9 : & \text{ B} \\
| \quad X > 9 : & \text{ C}
\end{align*}\n
This is the training set from which the regression tree was learned:

<table>
<thead>
<tr>
<th>X</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>14</td>
<td>15</td>
</tr>
</tbody>
</table>

Write down the output (class) values and number of instances that appear in each of the leaf nodes A, B and C of the tree.

(6 marks)
8. In general, feedforward neural networks (multi-layer perceptrons) trained by error back-propagation are:
   (i) fast to train, and fast to run on unseen examples
   (ii) slow to train, and fast to run on unseen examples
   (iii) fast to train, and slow to run on unseen examples
   (iv) slow to train, and slow to run on unseen examples

   In ONE (1) sentence, explain your choice of answer. 

   (1 mark)

9. Suppose you have a decision tree (DT) and a multi-layer perceptron (MLP) that have been trained on data sampled from a two-class target function, with all attributes numeric. You can think of both models as graphs whose edges are labelled with numbers: weights in the MLP and threshold constants for attribute tests in the DT. Compare and contrast the roles of these numbers in the two models.

   (4 marks)
10. Suppose a genetic algorithm uses chromosomes of the form $x = abcddefgh$ with a fixed length of eight genes. Each gene can be any digit between 0 and 9. Let the fitness of individual $x$ be calculated as:

$$f(x) = (a + b) - (c + d) + (e + f) - (g + h)$$

And let the initial population consist of four individuals with the following chromosomes:

$$x_1 = 65413532$$
$$x_2 = 87126601$$
$$x_3 = 23921285$$
$$x_4 = 41852094$$

a) Evaluate the fitness of each individual. Showing all your workings and arrange them in order with the fittest first and the least fit last. (3 marks)

b) Perform the following crossover operations:
   i) Cross the fittest two individual using one-point crossovers at the middle point. (2 marks)

   ii) Cross the second and third fittest individual using a two-point crossover (point $b$ and $f$). (2 marks)
iii) Cross the first and third fittest individual (ranked 1st and 3rd) using a uniform crossover.

(2 marks)

c) Suppose the new population of the six offspring individuals received by the crossover operations in the above question. Evaluate the fitness of the new population, showing all your workings. Has the overall fitness improved?

(6 marks)
d) By looking at the fitness function and considering that genes can only be digits between 0 and 9 find the chromosome representing the optimal solution (i.e. with the maximum fitness). Find the value of the maximum fitness.

(3 marks)

e) By looking at the initial population of the algorithm can you say whether it will be able to reach optimal solution without the mutation operator?

(2 marks)

11. As model complexity increases from low to high, what effect does this have on

a) bias?

(2 marks)
b) predictive accuracy on training data? (2 marks)

c) predictive accuracy on test data? (2 marks)
SECTION B: PROBLEM BASED LEARNING (40 MARKS)

CASE 1 (COMPARING LAZY AND EAGER LEARNING)

The following truth table gives an "m-of-n function" for three Boolean variables, where "1" denotes true and "0" denotes false. In this case the target function is: "exactly two out of three variables are true".

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>Z</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>false</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>false</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>false</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>true</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>false</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>true</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>true</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>false</td>
</tr>
</tbody>
</table>

a) Construct a decision tree which is complete and correct for the example in the table. [Hint: draw a diagram]. Given the formula;

\[
\text{Entropy}(S) = -p \cdot \log_2 p - q \cdot \log_2 q \\
\text{Gain}(S,A) = \text{E}(S) - \text{E}(A)
\]

(4 marks)

b) Construct a set of ordered classification rules which is complete and correct for the example in the table. [Hint: use an if-then-else representation].

(4 marks)

c) Based on your answer in Case 1a and Case 1b, assume that you would like to use an Instance Based Learning algorithm to classify the wrong example in the table. Find the example of data and;

i) show the classification using k=3 and use Euclidean Distance to calculate the distance between two such strings \( x_i \) and \( x_j \). The formula is:

\[
d(x_i, x_j) = \sqrt{\sum_{t=1}^{n} (a_t(x_i) - a_t(x_j))^2}
\]

(5 marks)

ii) by using distance weighted k-NN (wkNN), provide a classification using k=3. The formula is:

\[
f(x) = \frac{\sum_{i=1}^{k} w_i f(x_i)}{\sum_{i=1}^{k} w_i} \quad \text{where} \quad w_i = \frac{1}{d(x_i, x)^2}
\]

(5 marks)
d) Compare your three models. Which do you conclude provides a better representation for this particular problem? Give your reasoning in one sentence.  

(2 marks)

CASE 2 (BAYESIAN LEARNING)

In machine learning we are often interested in determining the best hypothesis from some space $H$, given the observed training data $D$. Bayes theorem provides a way to calculate the probability of a hypothesis based on its prior probability, the probabilities of observing various data given the hypothesis, and the observed data itself.

a) Explain the difference between the maximum a posteriori hypothesis $H_{\text{MAP}}$ and the maximum likelihood hypothesis $H_{\text{ML}}$.  

(4 marks)

b) Consider a two-class learning problem to “Play tennis”, with two Boolean attributes, “Cloudy” and “Windy”. Draw the Bayesian network corresponding to a Naive Bayes classifier for this problem.  

(2 marks)

c) Given the following examples, calculate all the probabilities required for your Naive Bayes classifier to be able to decide whether to play or not:

<table>
<thead>
<tr>
<th>Instance No.</th>
<th>Cloudy</th>
<th>Windy</th>
<th>Play tennis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>no</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>no</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>no</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>no</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>yes</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>yes</td>
</tr>
</tbody>
</table>

(10 marks)

d) To which class would your Naive Bayes classifier assign each of the following instances?

<table>
<thead>
<tr>
<th>Instance No.</th>
<th>Cloudy</th>
<th>Windy</th>
<th>Play tennis</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>?</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>1</td>
<td>?</td>
</tr>
</tbody>
</table>

(4 marks)

END OF QUESTION