Name of Candidate: ....

Student number: ....

Signature: ....

COMP9417 Machine Learning and Data Mining

Final Examination: SAMPLE QUESTIONS

Here are seven questions which are representative of the type that will be in the final exam. Each question is of equal value, and should not take longer than 30 minutes to complete. In the actual exam there will be some degree of choice as to which questions you can answer. Candidates may bring authorised calculators to the examination, but no other materials will be permitted.

This page intentionally left blank.

Please see over

Question 1 [20 marks]

Association Rule Learning & and ILP

A) **[7 marks]**
An association rule has the form $A \rightarrow C$ where A is the antecedent and C is the consequent. Suppose you have a database of one million transactions.

1) If an association rule has 90% support, how many transactions contain all the items in A ?
2) For the same rule, how many transactions contain all the items in C ?
3) For the same rule, how many transactions contain all the items in both A and C ?
4) If another rule has 40% support and 100% confidence, how many transactions contain all the items in A ?
5) If yet another rule has 40% support and 50% confidence, how many transactions contain all the items in A ?

B) **[3 marks]**
You are mining a data warehouse. In three sentences or less propose a method to derive all candidate item sets of size $k+1$ from the k-item sets with at least minimum support.

C) **[4 marks]**
Consider the following two clauses:
$C = P(A, x) \vee Q(A, B)$ and $C_1 = Q(y, B) \vee T(y)$
Using inverse resolution, provide at least one solution for C_2. [Show all substitutions].

D **[6 marks]**
Consider the following two clauses:
$C = Q(A, x, B) \vee S(y, B)$ and $C_1 = S(w, B) \vee \neg R(z)$
Using inverse resolution, provide at least two solutions for C_2. [Show all substitutions].

Question 2 [20 marks]

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".

X	Y	Z	Class
0	0	0	false
0	0	1	false
0	1	0	false
0	1	1	true
1	0	0	false
1	0	1	true
1	1	0	true
1	1	1	false

A) **[4 marks]**
Construct a decision tree which is complete and correct for the examples in the table. [Hint: draw a diagram.]

B) **[4 marks]**
Construct a set of **ordered** classification rules which is complete and correct for the examples in the table. [Hint: use an *if–then–else* representation.]

C) **[10 marks]**
Suppose we define a simple measure of distance between two equal length strings of Boolean values, as follows. The distance between two such strings B_1 and B_2 is:

$$\text{distance}(B_1, B_2) = |(\sum B_1) - (\sum B_2)|$$

where $\sum B_i$ is simply the number of variables with value 1 in string B_i. For example:

$$\text{distance}(\langle 0, 0, 0\rangle, \langle 1, 1, 1\rangle) = |0 - 3| = 3$$

and

$$\text{distance}(\langle 1, 0, 0\rangle, \langle 0, 1, 0\rangle) = |1 - 1| = 0$$

What is the LOOCV ("Leave–one–out cross-validation") error of 2–Nearest Neighbour using our distance function on the examples in the table ? [Show your working.]

D) **[2 marks]**
Compare your three models. Which do you conclude provides a better representation for this particular problem ? Give your reasoning (one sentence).

Question 3 [20 Marks]

Mistake Bounds

Consider the following learning problem in which each instance is an integer in the range $[1, 5]$, and in which each hypothesis is an interval over the integers. More precisely, each hypothesis h in the hypothesis space H is an interval of the form $a \leq x \leq b$, where a and b are integer constants and x refers to the instance. For example, the hypothesis $3 \leq x \leq 5$ classifies the integers 3, 4 and 5 as positive and all others as negative.

Instance	**Class**
1	Negative
2	Positive
3	Positive
4	Positive
5	Negative

A) **[15 marks]**
Apply the HALVING ALGORITHM to the five examples in the order in which they appear in the table above. Show each class prediction and whether or not it is a mistake, plus the initial G and S sets and those at the end of each iteration.

B) **[5 marks]**
What is the worst-case mistake bound for the HALVING ALGORITHM given the hypothesis space described above ? Give an informal derivation of your bound.

Question 4 [20 marks]

Evaluation of Learning

A) **[3 marks]**
The AUC (area under the ROC curve) measure originated in signal detection theory. For the evaluation of classifier learning on a two-class prediction problem, explain the *probabilistic* interpretation of this measure.

B) **[3 marks]**
You have implemented a new two-class classifier learning algorithm that outputs a "score" to indicate how strongly it believes an input instance to be in the positive class. Explain how to you would use a *margin*-based approach to penalise incorrect classifications.

C) **[14 marks]**
Suppose the following decision tree has been learned on a training set of 10 positive and 10 negative examples for a "play sport" task.

```
Outlook = sunny:
    Temperature = warm:  [ play (5); don't play (2) ]
    Temperature = cold:  [ play (4); don't play (3) ]
Outlook = rainy:  [ play (1); don't play (5) ]
```

The three leaves of the tree each show the number of positive (play) and negative (don't play) examples in each leaf. Using a scoring function based on the ratio of positive to negative examples in each leaf node, generate the *ranking* of the training set produced by this scoring function. Then compute the *ranking error* for the tree on the training data.

Question 5 [20 Marks]

Computational Learning Theory

A) **[8 marks]**
An instance space X is defined using m Boolean attributes. Let the hypothesis space H be the set of decision trees defined on X (you can assume two classes). What is the largest set of instances in this setting which is shattered by H ? [Show your reasoning.]

B) **[10 marks]**
Suppose we have a consistent learner with a hypothesis space restricted to conjunctions of exactly 8 attributes, each with values $\{\text{true}, \text{false}, \text{don't care}\}$. What is the size of this learner's hypothesis space ? Give the formula 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%. [Note: you are *not* required to compute the solution.]

C) **[2 marks]**
Informally, which of the following are consequences of the No Free Lunch theorem:

a) averaged over all possible training sets, the variance of a learning algorithm dominates its bias

b) averaged over all possible training sets, no learning algorithm has a better off-training set error than any other

c) averaged over all possible target concepts, the bias of a learning algorithm dominates its variance

d) averaged over all possible target concepts, no learning algorithm has a better off-training set error than any other

e) averaged over all possible target concepts and training sets, no learning algorithm is independent of the choice of representation in terms of its classification error

Question 6 [20 marks]

Ensemble Learning

A) **[8 marks]**
As model complexity increases from low to high, what effect does this have on:

1) Bias ?
2) Variance ?
3) Predictive accuracy on training data ?
4) Predictive accuracy on test data ?

B) **[3 marks]**
Is decision tree learning relatively stable ? Describe decision tree learning in terms of bias and variance in no more than two sentences.

C) **[3 marks]**
Is nearest neighbour relatively stable ? Describe nearest neighbour in terms of bias and variance in no more than two sentences.

D) **[3 marks]**
Bagging reduces bias. True or false ? Give a one sentence explanation of your answer.

E) **[3 marks]**
Bagging reduces variance. True or false ? Give a one sentence explanation of your answer.

Question 7 [20 marks]

Bayesian Learning

A) **[4 marks]**
Explain the difference between the *maximum a posteriori* hypothesis H_{MAP} and the *maximum likelihood* hypothesis H_{ML}.

B) **[2 marks]**
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.

C) **[10 marks]**
Given the following examples, calculate *all* the probabilities required for your Naive Bayes classifier to be able to decide whether to play or not:

Instance No.	Cloudy	Windy	Play tennis
1	0	0	no
2	0	1	no
3	1	1	no
4	0	0	no
5	0	1	yes
6	1	0	yes

D) **[4 marks]**
To which class would your Naive Bayes classifier assign each of the following instances ?

Instance No.	Cloudy	Windy	Play tennis
7	0	0	?
8	0	1	?

END OF PAPER