Student Number | |

The University of Melbourne
School of Computing and Information Systems

PRACTICE EXAM, Semester 1, 2019
COMP30027 Machine Learning

Reading Time: 15 minutes. **Writing Time:** 2 hours.

This paper has 3 pages including this cover page.

Instructions to Invigilators:

> **Students should be provided with script books, and should answer all questions in the provided script book.** Students may not remove any part of the examination paper from the examination room.

Instructions to Students:

> There are 9 questions in the exam worth a total of 120 marks, making up 60% of the total assessment for the subject.

- Please answer all questions in the script book provided, starting each question on a new page. Please write your student ID in the space above and also on the front of each script book you use. When you are finished, place the exam paper inside the front cover of the script book.
- Your writing should be clear; illegible answers will not be marked.

Authorised Materials: No materials are authorised.

Calculators: Students are permitted to use calculators.

Library: This paper may be held by the Baillieu Library.

Examiners' use only									
1	2	3	4	5	6	7	8	9	Total

Section A: Short Answer Questions

Answer each of the questions in this section as briefly as possible. Expect to answer each sub-question in a couple of lines.

Question 1: Short Answer Questions

1. Classify the following attributes as discrete, ordinal or continuous:
   (a) number of patients in a hospital
   (b) ability to pass light, in terms of the following values: opaque, translucent and transparent
2. What is the difference between classification and regression?
3. Outline the nature of "hill climbing" and provide an example of a hill climbing algorithm. What basic approach does hill climbing contrast with?
4. Optionally with the use of a diagram, explain what it means for a dataset to be "linearly separable"?
5. One advantage of "bagging" is that is can be "parallelised": how and why?

Section B: Method Questions

In this section you are asked to demonstrate your conceptual understanding of a subset of the methods that we have studied in this subject.

Question 2: Logistic Regression

For a "logistic regression" model trained over n classes, based on k training instances and f features, and tested over t test instances, use big-O notation to describe the approximate number of parameters in the trained model.

Question 3: Evaluation

1. How does stratified cross validation contrast with (non-stratified) cross validation?
2. What is "sum of squared errors"? Name a learning task it is applied to.

Section C: Numeric Questions

Question 4: Naive Bayes

1. Estimate the conditional probabilities for $P(A|+)$, $P(B|+)$ and $P(C|-)$ based on the following training dataset *without* smoothing:

A	B	C	Class
0	0	0	$+$
0	0	1	$-$
0	1	1	$-$
0	1	1	$-$
0	0	1	$+$
1	0	1	$+$
1	0	1	$-$
1	0	1	$-$
1	1	1	$+$
1	0	1	$+$

2. Predict the class label for $(A = 0, B = 1, C = 0)$ using the given training dataset and non-smoothed probability estimates, based on the naive Bayes approach.

3. Estimate the conditional probabilities for $P(A|+)$, $P(B|+)$ and $P(C|-)$ based on the given training dataset with Laplacian ("+1") smoothing.

4. Predict the class label for $(A = 0, B = 1, C = 0)$ using the given training dataset and smoothed probability estimates using Laplacian smoothing, based on the naive Bayes approach.

5. With reference to the underlying assumptions made in Laplacian smoothing, explain the observed differences in the predictions from Parts 2 and 4.

— End of Exam —