University College Dublin
An Coláiste Ollscoile, Baile Átha Cliath

AUTUMN TRIMESTER EXAMINATIONS

ACADEMIC YEAR 2019/2020

COMP 47750
Machine Learning with Python

Prof. Jeremy Pitt
Dr. Chris Bleakley
Prof. P. Cunningham *

Time Allowed: 2 Hours

Instructions for Candidates

Answer any four questions. Each question is worth 25 marks.

Total marks available are 100.

Use of non-programmable calculators is allowed.

Student Number

Seat Number

Instructions for Invigilators

Use of non-programmable calculators is allowed.

Question 1

a. In supervised Machine Learning, the Naive Bayes classifier can be considered a *ranking* classifier. What does this mean?

(10 marks)

b. Describe an application of supervised Machine Learning where the *ranking* ability of Naive Bayes might be useful.

(5 marks)

c. What functions does ***scikit learn*** provide to enable ranking with Naive Bayes? Describe how the ranking ability of Naive Bayes can be used in ***scikit learn***.

(10 marks)

Question 2

a. The potential for an ensemble of predictors to achieve an accuracy higher than that of a single predictor depends on there being diversity in the ensemble members. Describe two mechanisms by which diversity can be achieved in ensembles.

(5 marks)

b. In boosting, the ensemble members are trained in sequence (one after the other) with the distribution of the training data changed at each stage in the training process. Explain how this works and explain the underlying principle.

(10 marks)

c. In machine learning, error can be decomposed into **bias** and **variance** components. Explain these concepts with the aid of examples. What impact would you expect an ensemble to have on these error components?

(10 marks)

Question 3

a. Explain how cross validation maximizes the use of the available data in both training and testing.

(5 marks)

b. Explain with the aid of an example dataset of five to ten instances how to construct an ROC curve.

(10 marks)

c. How can ROC curves be used to compare two different classifiers.

(10 marks)

Question 4

a. The table below shows a training set with 7 examples represented by 4 categorical features, describing a person's preferences for booking hotels. Each example has a binary class label: Book? = *{yes, no}.*

i. Calculate the overall entropy for this data.
ii. Using Information Gain, identify the best feature to split the root node of a DecisionTree classifier built on the training set. Show your calculations.

(15 marks)

Example	Stars	Pool	Beach	Gym	Book?
Hotel 1	2	N	N	Y	no
Hotel 2	2	Y	N	N	yes
Hotel 3	3	N	Y	N	no
Hotel 4	3	Y	N	Y	no
Hotel 5	3	N	N	N	no
Hotel 6	3	Y	Y	Y	yes
Hotel 7	4	Y	Y	Y	yes

b.

i. In the context of supervised learning, what is the difference between overfitting and underfitting?
ii. Briefly outline one real-world application of classification, where the practical implications of a False Positive error and a False Negative error might differ.

(10 marks)

Question 5

a. k-Means clustering sets out to minimize the following sum of squared errors (SSE) function. Explain the principle behind this objective and describe the steps in the k-Means algorithm.

$$SSE(C) = \sum_{c=1}^{k} \sum_{x_i \in C_c} D(x_i, \mu_c)^2$$

where

$$\mu_c = \frac{\sum_{x_i \in C_c} x_i}{|C_c|}$$

and

$$D(x_i, \mu_c) = \sqrt{\sum_{l=1}^{m} (x_l - \mu_l)^2}$$

(8 marks)

b. Discuss the stability problems associated with the *k*-Means algorithm. What strategy is used in ***scikit learn*** to address this instability?

(8 marks)

c. The Silhouette method is a technique for quantifying the validity of a clustering. Explain how it works and explain the principle underlying the technique.

(9 marks)

Question 6

a. Explain why training a multi-layer feedforward neural network is considerably more difficult than training a single layer network.

(7 marks)

b. Even in a simple single layer Feedforward Neural Network the units (neurons) will have a fixed bias input. What is the reason for this bias input?

(6 marks)

c. "*One of the major advantages of neural nets is their ability to generalize.*" What does 'generalize' mean in this context?

(6 marks)

d. Explain in terms of the dynamics of a single neuron why it is that single layer perceptrons are only able to learn patterns that are linearly separable.

(6 marks)