University College Dublin
An Coláiste Ollscoile Baile Átha Cliath

Spring, 22/23 TRIMESTER EXAMINATIONS

COMP47750

Machine Learning with Python

Module Coordinator: Professor Pádraig Cunningham

Student Number

Seat Number

Time Allowed: 60 minutes

Materials Permitted in the Exam Venue:
Foreign language dictionary (hard copy)
Non-programmable or scientific calculator

Materials to be Supplied to Students:
8 Page Answer Booklets

Instructions to Students:
Answer Question 1 and any two other questions. Question 1 is worth 40 marks and all other questions are worth 30 marks each. The value of each part of each question is shown in brackets next to it.

Question 1

a. In *k*-Nearest Neighbour retrieval how can distances be calculated for Ordinal features?

(5 marks)

b. With scikit-learn, when StandardScalar normalisation (also known as N(0,1) normalisation) is applied to data, what is the distribution of the data after normalisation?

(5 marks)

c. If you have a collection of 3 green and 4 red balls and you add 2 green balls to the collection what happens to the entropy of the collection?

(5 marks)

d. If we have two decision trees, one simple and one more complex that both perfectly explain the training data, which should we prefer? Briefly explain why.

(5 marks)

e. When scoring the performance of classifiers what is the motivation for using balanced measures such as Balanced Accuracy Rate or Balanced Error Rate?

(5 marks)

f. Name two options for dealing with numeric (real valued) features in a Naive Bayes classifier.

(5 marks)

g. Why is it not possible to use *k*-Means clustering with categorical data.

(5 marks)

h. Briefly describe one method to achieve diversity in an ensemble.

(5 marks)

Question 2

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

(8 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?

(7 marks)

c. Explain the operation of the following components in the training of a neural network using gradient descent;

i. Cost function
ii. Weight update
iii. Stopping condition

(10 marks)

d. What is the difference between stochastic gradient descent and batch gradient descent?

(5 marks)

Question 3

a. When grid search is being used as part of a model selection process, explain the difference between random and exhaustive grid search. Mention one advantage of each strategy.

(8 marks)

b. Explain the difference between hyperparameters and ordinary model parameters using neural networks as an example.

(7 marks)

c. An important principle in evaluating machine learning models is that the test data should not be accessed in the model training process because it can result in unrealistic estimates of generalisation accuracy. This applies to the data preprocessing pipeline as well as the model fitting itself. In practice some scenarios are more serious than others; comment on the seriousness of each of the following scenarios:

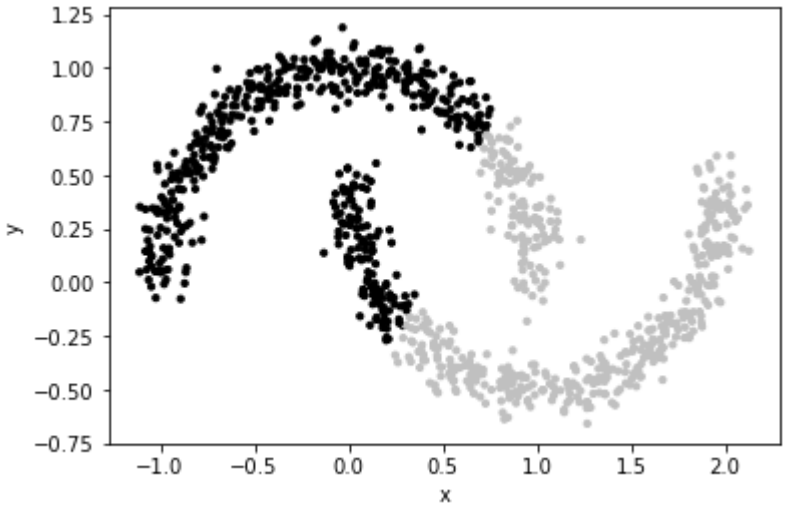
i. The training data is used during the feature selection process.
ii. The training data is used to fit a One-Hot Encoder.
iii. Missing values are replaced using the mean values for features across the training and test sets, i.e. the test data is used when the means are being calculated.

(15 marks)

Question 4

a. Sample data from the synthetic half-moons dataset from scikit-learn is shown in the plot below. This is a tabular dataset with each point represented by two features, the *x* and *y* coordinates. The clusters found by *k*-Means clustering are shown (black and grey), *k*-Means has not been effective for finding clusters in this dataset, why is that?

(10 marks)



b. In contrast, spectral clustering is able to uncover the correct clusters in this data.
   i. Explain in outline how spectral clustering can be applied to tabular data.
   ii. Explain why spectral clustering would work well on this data.

(10 marks)

c. The half-moons data has a simple feature vector format where the features are the *x* and *y* coordinates. Explain how this could be converted into the affinity matrix format required for spectral clustering.

(10 marks)

oOo