

University of St Andrews



MAY 2019 EXAMINATION DIET

SCHOOL OF COMPUTER SCIENCE

MODULE CODE: CS5014

MODULE TITLE: Machine Learning

EXAM DURATION: 2 hours

EXAM INSTRUCTIONS

- (a) Answer **three** questions.
- (b) Each question carries 20 marks.
- (c) Answer questions in the script book.

PERMITTED MATERIALS Non-programmable calculator

YOU MUST HAND IN THIS EXAM PAPER AT THE END OF THE EXAM.

**PLEASE DO NOT TURN OVER THIS EXAM PAPER UNTIL
YOU ARE INSTRUCTED TO DO SO.**

1. Machine Learning Concepts

- (a) In the context of machine learning, what is a *cost function* and why is it needed? Give an example of a common cost function and explain its parts.

[3 marks]

- (b) What is the gradient of the cost function and how many elements (dimensions) does it have? Explain how gradient is used to implement two common optimisation strategies: gradient descent and the normal equation.

[3 marks]

- (c) You are asked to predict the direction and magnitude of an underwater stream based on a set of sensor measurements. The measurements consist of input samples $X = [X_1, \dots, X_8]^T$, each of which corresponds to eight temperature readings from eight nearby sensors. The target is a three-dimensional vector $Y = [Y_1, Y_2, Y_3]^T$ which represents the velocity of water at a particular point.

- i. Give the equation for a linear regression model to solve this task. Make sure to explain any parameters of your model and state how many parameters there are.

[3 marks]

- ii. Is this model more likely to overfit or underfit and why? If it is likely to overfit or underfit, how would you address this problem?

[3 marks]

- iii. Discuss two ways to perform feature selection on the input data and explain how each of them works.

[4 marks]

- (d) A nearest neighbour classifier is used to classify music songs into three genres. It produces the following confusion matrix (correct class shown on top and predicted class shown on the left):

| | Grunge | Disco | Jazz |
|---------------|--------|-------|------|
| <i>Grunge</i> | 71 | 3 | 21 |
| <i>Disco</i> | 4 | 93 | 7 |
| <i>Jazz</i> | 25 | 4 | 72 |

- i. Calculate the mean classification rate. [1 mark]
- ii. Calculate Precision and Recall for the “*Disco*” class. [2 marks]
- iii. Are these results from the training or the test set? Your answer must include the reason why. [1 mark]

[Total 20 marks]

2. Deep Learning

- (a) What is a vanishing gradient in the context of backpropagation in neural networks? Explain why it can lead to problems in deep neural networks and discuss strategies for dealing with these problems.

[3 marks]

- (b) There are many regularisation strategies for deep neural networks. Name three of them. For each strategy you name, explain how it works, and discuss any strengths and weaknesses.

[6 marks]

- (c) Your task is to design a voice recognition system based on a deep neural network. The input for the network is 10 seconds of speech samples. The task of the network is to decide whether a 1-second keyword was uttered at any point during these 10 seconds. The deep network should learn the relevant features from samples directly.

- i. Sketch a suitable architecture for feature extraction stages of your network. The exact number of layers is not important, but make sure to identify different types of layers in your model and describe what they do.

[3 marks]

- ii. Sketch a suitable architecture for the classification stage of your network and explain the connectivity and the number of output neurons.

[2 marks]

- iii. Justify your choice of activation functions in each layer.

[2 marks]

- iv. Choose a suitable optimisation strategy for learning the parameters of your network and justify your decision.

[1 mark]

- (d) Why is *data augmentation* useful in this scenario? Explain some of the ways data augmentation could be applied to your input data to improve performance.

[3 marks]

[Total 20 marks]

3. Trees and Forests

- (a) Given a decision tree, describe the process for classifying a new input sample. [2 marks]
- (b) Overfitting can be a problem in decision trees. Describe one method to regularise a decision tree in order to reduce overfitting. Also include reasons why this would reduce overfitting. [4 marks]
- (c) What are the advantages of a random forest classifier over a single decision tree classifier? [2 marks]
- (d) Impurity measures for decision trees.
- Name one impurity measure and describe how it works. You can either include a sketch or use an equation to help with the explanation. [2 marks]
 - Explain how impurity measures are used to choose decision boundaries during decision tree learning. [2 marks]
- (e) Table 1 shown below contains data from the Iris Flower dataset.

Table 1: Iris Flower dataset

| Sepal Length | Sepal Width | Class |
|--------------|-------------|------------|
| 5.1 | 3.5 | Setosa |
| 4.9 | 3.0 | Setosa |
| 4.7 | 3.2 | Setosa |
| 4.6 | 3.1 | Setosa |
| 7.0 | 3.2 | Versicolor |
| 6.4 | 3.2 | Versicolor |
| 6.9 | 3.1 | Versicolor |
| 5.5 | 2.3 | Versicolor |
| 6.3 | 3.3 | Virginica |
| 5.8 | 2.7 | Virginica |
| 7.1 | 3.0 | Virginica |
| 6.3 | 2.9 | Virginica |

P.T.O

- i. Using a subset of this iris flower data create a decision tree to classify this data. Include all the steps you make in creating the tree, including the impurity measure used.

[6 marks]

- ii. Explain how your solution above could be extended to create a random forest classifier.

[2 marks]

[Total 20 marks]

4. Unsupervised Learning and Neural Networks

(a) What types of tasks are suitable for unsupervised learning? In your answer provide general class of tasks and also provide a specific example.

[2 marks]

(b) Describe the process of the K-means algorithm.

[5 marks]

(c) What does the acronym PCA stand for and what is it used for? Include a basic and brief explanation of how it works, you may include a sketch if it helps your explanation.

[3 marks]

(d) What is the purpose of forward propagation in neural networks?

[1 mark]

(e) You are given the task, using neural networks, to classify a powder as either cocaine, flour, or paracetamol using the inputs from an infrared optical sensor readings and two gas sensor readings.

i. Describe and explain any problems you foresee might happen when using the system to predict the class of an unknown powder.

[2 marks]

ii. Describe both forward propagation and backpropagation in neural networks for the above task. The network architecture you should use has the following properties: there is an input layer, 3 hidden layers and an output layer, you get to decide the number of neurons per layer. Your answer should use a sketch, any equations used and explanatory text of the process.

[7 marks]

[Total 20 marks]

***** END OF PAPER *****