

# King's College London

This paper is part of an examination of the College counting towards the award of a degree. Examinations are governed by the College Regulations under the authority of the Academic Board.

**Examination Period** May 2022 (Period 2)  
**Module Code** 6CCS3ML1 and 6CCS3PRE  
**Module Title** Machine Learning and Pattern Recognition

**Format of Examination** Written questions  
**Start time** 30 May 10.00am  
**Time Allowed** 1.5 hours  
**Instructions** You are permitted to access any materials you wish, but this is not mandated and is not expected. You may use a calculator if you find this helpful.  
**Rubric** ANSWER ALL QUESTIONS

The rubric for this paper must be followed and extra answers should not be submitted. For answers that are handwritten, write with blue/black ink on light coloured paper. Include the Module code, question number and student number on every page to be submitted. For answers that are typed, use the template provided.

**Submission Deadline** 11.30am  
**Submission Process** Work must be submitted to the **level 6** Informatics Assessments KEATS page.

Your work must be submitted as a PDF file. If you have prepared some answers on computer, and some on paper (which have then been digitised), you may upload at most two PDF files – one for computer-prepared answers, one for digitised answers. Do not duplicate answers across the two PDFs – if you do this, the computer-prepared answer will be taken. You should check that your work displays correctly after it has been uploaded.

## ACADEMIC HONESTY AND INTEGRITY

Students at King's are part of an academic community that values trust, fairness and respect and actively encourages students to act with honesty and integrity. It is a College policy that students take responsibility for their work and comply with the university's standards and requirements. Online proctoring / invigilation will not be used for our online assessments. By submitting their answers students will be confirming that the work submitted is completely their own. Misconduct regulations remain in place during this period and students can familiarise themselves with the procedures on the College website

**Important: Students should copy out the following statement and include it with their submission for each examination:**

I agree to abide by the expectations as to my conduct, as described in the academic honesty and integrity statement.

1. You are a journalist for a media organisation and your task is to identify fake news online. You are given a large collection of articles that have been annotated as “fake news” or “OK”.

a. How would you use the available data to apply a machine learning classifier to the task at hand? Justify your answer.

[3 marks]

b. How can you apply a Naive Bayes classifier to the task at hand? Describe your approach and how you would do parameter estimation. Provide equations.

[12 marks]

c. You decide to evaluate your model using precision and recall. Explain in which way these metrics are suitable here.

[6 marks]

d. How would you calculate precision and recall for the task at hand?

[4 marks]

2. You are a climatologist studying the weather, and your task is to estimate the temperature every day as either Hot (H) or Cold (C). However, the *only* observations you have about the weather are the number of ice creams your daughter eats on a given day.

a. Describe the components of a first-order Hidden Markov Model (HMM) for the task at hand. Provide equations where appropriate.

[6 marks]

b. What are the assumptions we make in a first order HMM and how are they interpreted for the task at hand?

[5 marks]

c. Given the training set below, which temperature is going to result in the highest number of ice creams consumed? Show your calculations.

1	3	3	1	3	2	3	2	(number of ice creams)
C	C	H	C	H	H	H	C	(temperature)

[6 marks]

d. Given a first-order HMM trained on the training set given in c. above, is eating 3 ice creams more likely to be followed by eating 1 or 2 ice creams?

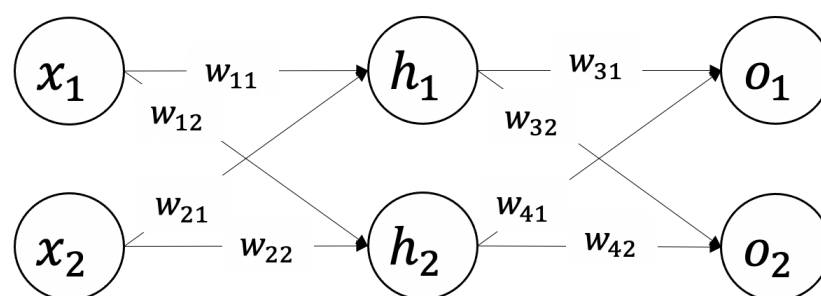
[8 marks]

3.

- a. You are given the 2 scenarios below. For each of these, determine the type of learning you would use to solve them – supervised learning, unsupervised learning, reinforcement learning – and explain your answer. Also write down an appropriate machine learning algorithm you would choose for each scenario.
- A film distributor has a collection of film posters, where each poster is annotated as belonging to one of the following genres: "Horror", "Comedy", "Drama", "Animated". The distributor wants to use this collection to build a system that identifies the genre of upcoming films based on their posters.
  - A supermarket has a database of its customers. It wants to automatically discover information about its customers and group them into different market segments to improve on their ad targeting.

[14 marks]

- b. The following diagram represents a feed-forward neural network with one hidden layer:



The input nodes to this network are nodes  $x_1$  and  $x_2$ , and the output nodes are nodes  $o_1$  and  $o_2$ . The weights in the network are as follows:

$w_{11} = 1$	$w_{12} = -0.5$	$w_{21} = 0.5$	$w_{22} = -1$
$w_{31} = -0.5$	$w_{32} = -1$	$w_{41} = -1$	$w_{42} = 0.5$

Each hidden and output node uses the following activation function:

$$g(s) = \begin{cases} 1, & \text{if } s \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

Calculate the network output for the following input. Show your calculations.

$$\begin{array}{cc} x_1 & x_2 \\ \hline 2 & 1 \end{array}$$

[4 marks]

- c. You want to train the above network using backpropagation. What might you change to do this? Explain your answer.

[4 marks]

- d. You want to build a model to predict whether a film will become a box-office success (label=1) or not (label=0). You have a dataset of films with 100 input features, such as number of actors, number of actors who have been in a successful film before, film budget, etc. You decide to train a neural network with a single hidden layer of 32 hidden units. How many weights does the neural network have? Do not include biases. Show your calculations.

[3 marks]

4.

You have a budget of £150 that you can spend on books. Each book has a price and a value which represents the level of importance, as defined below:

Value	Price	Book
30	£35	Machine learning
23	£25	Operating systems
22	£30	Medical imaging
20	£23	Compilers
19	£29	Natural language processing
15	£20	Python programming
10	£21	Java programming

If you buy a book, you can buy it only once. In this exercise, you need to find the ideal combination of books to buy that maximises the value without exceeding your allocated budget. Formally, each book  $b_i$  has price  $p_i$  and value  $v_i$ , and we want to:

$$\text{maximise } \sum_{i=1}^n v_i \text{ given the constraint: } \sum_{i=1}^n p_i \leq B \text{ where } B \text{ is the budget}$$

To solve this task, you will be using a genetic algorithm.

- a.** Explain how you will represent the individuals to encode the problem.  
Give three examples of individuals using your representation.

[8 marks]

- b. Explain how we can obtain an individual which is not valid given the task definition and write the genotypes of two invalid individuals.

[6 marks]

- c. Now consider a different task, where you have the following two parents:

1	1	0	1	0	0	1	1
---	---	---	---	---	---	---	---

0	1	1	1	0	0	1	0
---	---	---	---	---	---	---	---

Given the following *crossover mask*:

1	0	0	1	1	0	0	0
---	---	---	---	---	---	---	---

Write down the resulting offspring.

[5 marks]

- d. Assuming the following two individuals represent the entire population, explain the importance of the mutation operator. Also apply the mutation operator to one of the individuals and write down the resulting individual.

1	0	1	0	0	0	1	0
---	---	---	---	---	---	---	---

0	1	0	1	0	1	0	1
---	---	---	---	---	---	---	---

[6 marks]