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.

Degree Programmes	MSc, MSci
Module Code	7CCSMML1
Module Title	Machine Learning
Examination Period	May 2019 (Period 2) 2019
Time Allowed	Two hours
Rubric	ANSWER ALL QUESTIONS. ANSWER EACH QUESTION ON A NEW PAGE OF YOUR ANSWER BOOK AND WRITE ITS NUMBER IN THE SPACE PROVIDED.
Calculators	Calculators may be used. The following models are permitted: Casio fx83 / Casio fx85.
Notes	Books, notes or other written material may not be brought into this examination

PLEASE DO NOT REMOVE THIS PAPER FROM THE EXAMINATION ROOM

1. a. Write down the update rule for temporal difference (TD) learning, and explain each of the components of the update.

[6 marks]

b. How does temporal difference (TD) learning differ from adaptive dynamic programming (ADP).

[4 marks]

2. a. Using the example of recidivism risk, explain why a good learning algorithm is not enough to ensure that a machine learning system is not harmful.

[6 marks]

b. How could you ensure that a machine learning system is not harmful? Why do you think this approach would work?

[4 marks]

3. You go to the shop to buy tulip bulbs to plant in your garden. The shop stocks two kinds of bags of tulips. They have the "Spring Fire" collection and the "Dutch Gold" collection, each with different proportions of bulbs for red and orange tulips (the bulbs all look the same). However the only bags the shop has left have no label, so you don't know what kind you are buying. You buy a bag anyway, and plant 100 bulbs. When the tulips flower you find you have 29 red tulips and 71 orange tulips.

Use one round of the EM algorithm to estimate the probability that you bought a "Spring Fire" bag.

Write θ for the probability of you having a "Spring Fire" bag, θ_{SF} for the probability that a bulb from a "Spring Fire" bag is red, and θ_{DG} for the probability that a bulb from a "Dutch Gold" bag is red. Start your run of EM assuming that:

$$\theta = 0.5$$
$$\theta_{SF} = 0.6$$
$$\theta_{DG} = 0.3$$

[10 marks]

4. a. Explain the representation of individuals that is used in genetic programming, and give an example of this representation.

[4 marks]

b. Explain how crossover and mutation can be performed in a genetic program. How is fitness established?

[6 marks]

5. a. A quadratic kernel function can be used to map a data sample $\mathbf{x} \in \mathbb{R}^3$ to a higher dimensional space $\mathcal{H} \in \mathbb{R}^6$, as follows:

$$F(\mathbf{x}) = [x_1^2, x_2^2, x_3^2, \sqrt{2}x_1x_2, \sqrt{2}x_1x_3, \sqrt{2}x_3x_2]$$

Define the term "kernel trick". Explain how it can simplify the computation of the inner product seen in support vector machines.

[6 marks]

b. Explain how the soft-margin support vector machine attempts to mitigate the effects of errors in data during learning. How is the margin implemented in the optimisation problem for the soft-margin support vector machine?

[4 marks]

6. a. What is gradient descent in the context of machine learning? Your answer should explain what gradient is and how it is determined.

[8 marks]

b. Give two examples of the use of gradient descent in machine learning. Each example should be in the context of a different approach to machine learning.

[2 marks]

7. The following is the output of a classifier which is distinguishing between 3 classes, A, B and C:

Example	Classifier output	Correct class
E01	C	A
E02	B	A
E03	B	C
E04	B	B
E05	A	B
E06	A	A
E07	C	C
E08	B	B
E09	A	C
E10	C	C
E11	C	B
E12	A	A
E13	C	A
E14	B	B
E15	A	A

a. Draw up the confusion matrix for the classifier.

[6 marks]

b. What does the confusion matrix tell you about the classifier?

[4 marks]

8. a. Explain how a perceptron works. Give an example to illustrate your answer.

[4 marks]

b. Given the following set of training examples:

x_1	x_2	x_3	desired output
1	1	1	-1
1	1	0	1
0	1	1	1
1	0	0	-1

Starting with weights that are all 2, and taking α to be 0.5, use the error-correction method on these three examples to train a perceptron which uses the sign function to compute its output. You only have to use each training example once.

[6 marks]

9. a. Explain the process of backpropagation for training a multi-layer feed-forward neural network. Your answer should include details of how hidden units are trained.

[6 marks]

b. How is it possible to train recurrent neural networks?

[4 marks]

10. The Travelpedia website has collected the following data from reviews left about 10 hotels:

Example	Features			Result
	Bar	Shuttle	Rating	
E01	N	Y	★	N
E02	Y	N	★	Y
E03	Y	Y	★★	N
E04	N	N	★★	Y
E05	N	N	★★★	N
E06	Y	Y	★	Y
E07	N	Y	★★★	Y
E08	Y	N	★★	Y
E09	Y	N	★	N
E10	Y	N	★★★	Y

where *Bar* records whether or not the hotel has a bar, *Shuttle* records whether or not the hotel has a free shuttle to the airport, and *Rating* indicates whether the hotel has a one, two or three star rating. A Y in the Result column indicates that the hotel is full during peak season, and N indicates that the hotel is not full.

Show how Gini impurity can be used to select the root node of a decision tree that will help Travelpedia use this data to decide whether a new hotel will be full or not.

To get full marks you will need to compute the Gini impurity for one of the features.

[10 marks]