Queen Mary University of London

Main Examination period 2021 – May/June – Semester B
Online Alternative Assessments

MTH793P: Advanced Machine Learning

You should attempt ALL questions. Marks available are shown next to the questions.

In completing this assessment:

- **You may use books and notes.**
- **You may use calculators and computers, but you must show your working for any calculations you do.**
- **You may use the Internet as a resource, but not to ask for the solution to an exam question or to copy any solution you find.**
- **You must not seek or obtain help from anyone else.**

All work should be **handwritten** and should **include your student number**.

You have **24 hours** to complete and submit this assessment. When you have finished:

- scan your work, convert it to a **single PDF file**, and submit this file using the tool below the link to the exam;
- e-mail a copy to **maths@qmul.ac.uk** with your student number and the module code in the subject line;
- with your e-mail, include a photograph of the first page of your work together with either yourself or your student ID card.

You are expected to spend about 2 hours to complete the assessment, plus the time taken to scan and upload your work. Please try to upload your work well before the end of the submission window, in case you experience computer problems. **Only one attempt is allowed – once you have submitted your work, it is final**.

Examiners: M. Benning

The notation log refers to the natural logarithm. The set of real numbers is denoted by $\mathbb{R}$. All computations should be done by hand where possible, with marks being awarded for intermediate steps in order to discourage computational aids.

Question 1 [38 marks].

(a) Rewrite the function $f(x) = \frac{dx-2x^2}{1+4x^2}$ to $g(a) + \varepsilon h(a, b)$, where the argument x is a dual number of the form $x = a + \varepsilon b$ with $\varepsilon^2 = 0$, and specify both g and h. The number d is one added to the last digit of your student ID number. [8]

(b) Compute the derivative $f'(a)$ of f as defined in Question 1(a) at argument a by making use of your result of Question 1(a). [8]

(c) Rewrite the function $f(x) = \frac{1}{2}\langle Qx, x\rangle$, acting on a vector $x = a + \varepsilon b$ of dual numbers, to $g(a) + \varepsilon h(a, b)$. Here, $Q \in \mathbb{R}^{n\times n}$ is a square matrix. Specify both g and h, and compute the partial derivative $\partial/\partial a_l f(a_1, \ldots, a_n)$ for some $l \in \{1, \ldots, n\}$. [6]

(d) Characterise all cases of the subdifferential $\partial\chi_{\geq 0}$ of the characteristic function

$$\chi_{\geq 0}(x) := \begin{cases} 0 & x \geq 0 \\ \infty & \text{otherwise} \end{cases} .$$

[8]

(e) Compute the generalised Bregman distance $D_f^p(x, y)$ with respect to the function $f(x) = |x|$ for a specific subgradient $p \in \partial f(y)$. Make sure to characterise all four different cases. [8]

Question 2 [31 marks].

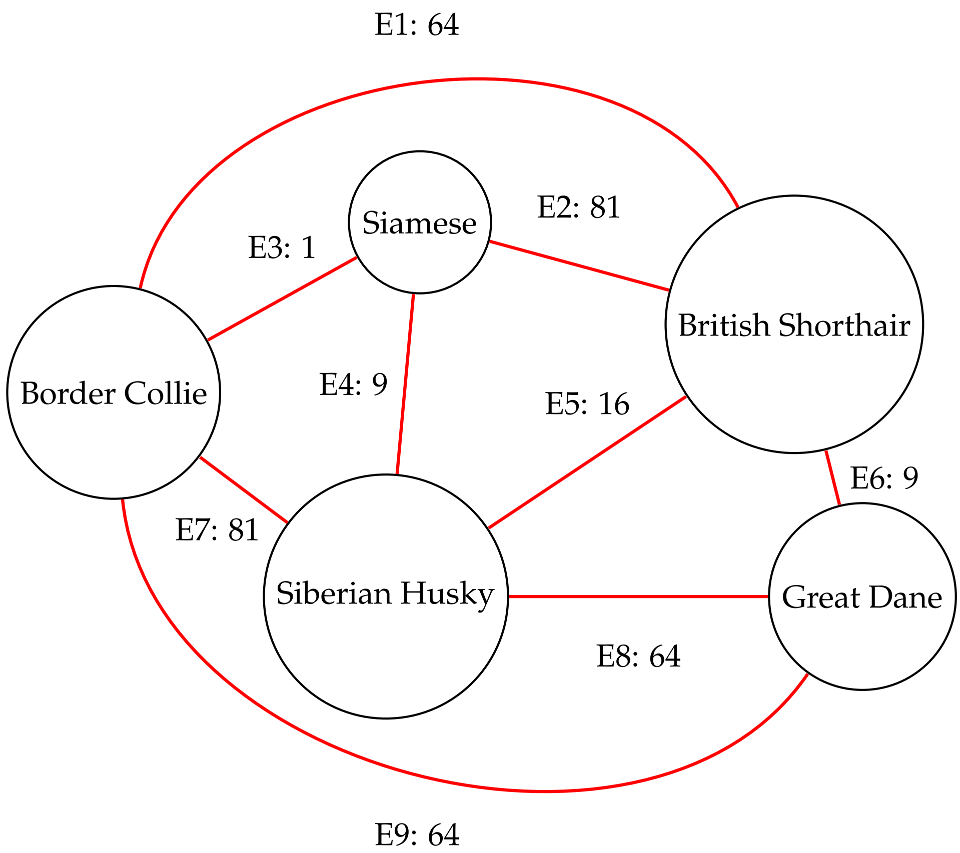
(a) It is the start of the Covid-19 vaccine role-out program and you want to decide whether or not to get vaccinated. You base your decision on three factors:

- Am I in any of the groups (age above threshold, living in care home, front-line worker etc.) eligible for the vaccine?
- Is the scientific advice in favour of getting vaccinated?
- Do people in my WhatsApp group chat think that I should get vaccinated?

Suppose you only get vaccinated if you are in any of the eligible vaccination groups and only if the scientific advice is in favour of you getting vaccinated. However, you really do not care about what people in your WhatsApp group chat think about you getting vaccinated or not.

Model this binary decision process with a perceptron and choose some appropriate weights to mimic the decision process accurately. **[8]**

(b) Write down the incidence matrix for the following weighted, undirected graph:



Use the definition of an incidence matrix from the lecture notes and order the columns of the incidence matrix alphabetically according to the vertex name and the rows according to the edge numbering (E1, E2, E3, ...). **[7]**

(c) Compute the corresponding graph Laplacian for the incidence matrix in Question 2(b). **[8]**

(d) We want to use the graph from Question 2(b) to determine whether a node in the graph belongs to the class ″dogs″ or the class ″cats″. Suppose we are in a semi-supervised setting, where the node ″Great Dane″ is already labelled $v_{\text{Great Dane}} = 1$ (class ″dogs″) and the node ″Siamese″ is labelled as $v_{\text{Siamese}} = 0$ (class ″cats″). Determine the remaining labels with the same procedure as described in the lecture notes and classify each node. **[8]**

Question 3 [31 marks].

(a) Perform k-means clustering by hand for the five data points $x_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $x_2 = \begin{pmatrix} -7 \\ 3 \end{pmatrix}$, $x_3 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$, $x_4 = \begin{pmatrix} 13 \\ 15 \end{pmatrix}$, and $x_5 = \begin{pmatrix} -11 \\ 17 \end{pmatrix}$. Assume $k = 2$ clusters and initialise your centroids as

$$\mu_1^0 := \begin{pmatrix} d \\ 0 \end{pmatrix} \qquad \text{and} \qquad \mu_2^0 := \begin{pmatrix} 0 \\ -d \end{pmatrix} ,$$

where d is one added to the eighth digit of your student ID number. For each iteration, update the assignments first, and then the centroids. Perform as many iterations as are required for the k-means clustering algorithm to converge. **[8]**

(b) Complete the following matrix such that it has minimal rank:

$$\begin{pmatrix} 1 & -2 & d & -7 \\ -3 & 6 & ? & 21 \\ ? & -10 & ? & -35 \end{pmatrix} .$$

Here d is the maximum of the seventh digit of your student ID number and 1. Justify your choice. **[6]**

(c) Show that for vectors $x, y \in \mathbb{R}^k$, the vector $z \in \mathbb{R}^k$ defined as

$$z_i = \frac{x_i \exp(y_i)}{\sum_{j=1}^k x_j \exp(y_j)} = \operatorname{softmax}(\log(x) y)_i ,$$

for all $i \in \{1, \ldots, k\}$, is the solution of the optimisation problem

$$z = \operatorname*{arg\,min}_{\tilde{z} \in \mathbb{R}^k} \left\{ D_f(\tilde{z}, x) - \langle \tilde{z}, y \rangle \quad \text{subject to} \quad \tilde{z} \in [0,1]^k , \text{ and } \sum_{j=1}^k \tilde{z}_j = 1 \right\} ,$$

where D_f denotes the Bregman distance with respect to the convex function $f(z) := \sum_{j=1}^k z_j \log(z_j)$.
Hint: reformulate the unconstrained objective $D_f(\tilde{z}, x) - \langle \tilde{z}, y \rangle$ to $D_f(\tilde{z}, z) + c \left(1 - \sum_{j=1}^k \tilde{z}_j\right) + d$ for constants c and d independent of $\tilde{z}$. **[8]**

(d) Consider the following modification of k-means clustering with uncertainty:

$$(\hat{z}, \hat{\mu}) = \operatorname*{arg\,min}_{z \in \mathbb{R}^{s \times k}, \mu \in \mathbb{R}^{n \times k}} \left\{ \sum_{i=1}^s \sum_{j=1}^k z_{ij} \|x_i - \mu_j\|^2 \text{ s.t. } z_{ij} \in [0,1] , \sum_{j=1}^k v_{ij} = 1, \forall i, j \right\} ,$$

for s data points $\{x_i\}_{i=1}^s$, where every $x_i \in \mathbb{R}^n$ is n-dimensional, centroids $\mu \in \mathbb{R}^{n \times k}$ and assignments $z \in \mathbb{R}^{s \times k}$. What is the difference to traditional k-means clustering? Formulate an algorithm to computationally solve this modified k-means clustering problem. **[9]**

End of Paper.