Queen Mary
University of London

Late-Summer Examination period 2021

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 determinant of a matrix is denoted by det. 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 $f(x) = \frac{3x^2-2x+4}{1+dx+2x^2}$ 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). **[6]**

(c) Consider the function $f(X) = \log(\det(X))$, acting on an invertible matrix $X \in \mathbb{R}^{2\times 2}$. Compute the partial derivatives with the help of dual number calculus, assuming each entry x_{ij} of X is a dual number of the form $a_{ij} + \varepsilon b_{ij}$ with $\varepsilon^2 = 1$, for $i, j \in \{1, 2\}$. Subsequently, show that the entire gradient equals $(X^\top)^{-1}$. **[8]**

(d) Verify that the subdifferential $\partial\|\cdot\|$ of the (non-squared!) Euclidean norm $f(x) := \|x\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$ for any $x \in \mathbb{R}^n$ is characterised via

$$\partial\|x\| = \begin{cases} \left\{ \frac{x}{\|x\|} \right\} & x \neq 0 \\ \{p \in \mathbb{R}^n \mid \|p\| \leq 1\} & x = 0 \end{cases} .$$

[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 different cases. **[8]**

Question 2 [31 marks].

(a) Determine parameters $W \in \mathbb{R}^{2\times 2}$, $w \in \mathbb{R}^2$, $b \in \mathbb{R}^2$ and $c \in \mathbb{R}$ of a neural network of the form

$$f(x_1, x_2) = w^\top \max\left(0, W^\top \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + b\right) + c$$

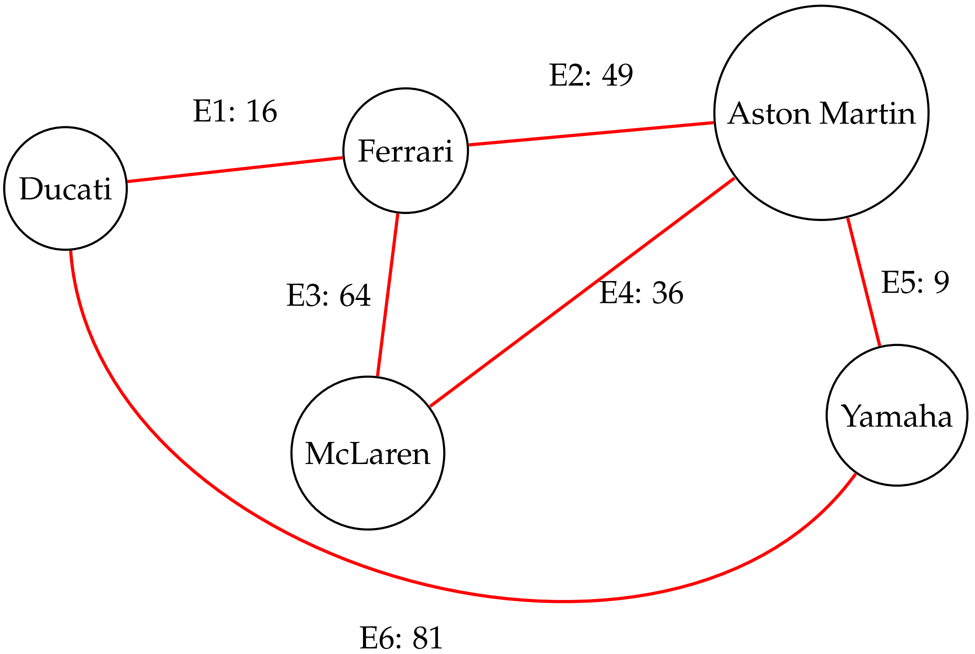
that is the logical XNOR-function, i.e.

x_1	x_2	$f(x_1, x_2)$
0	0	1
1	0	0
0	1	0
1	1	1

.

[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 ”cars” or the class ”motorbikes”. Suppose we are in a semi-supervised setting, where the node ”McLaren” is already labelled $v_{\text{McLaren}} = 1$ (class ”cars”) and the node ”Yamaha” is labelled as $v_{\text{Yamaha}} = 0$ (class ”motorbikes”). 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 $x_1 = \begin{pmatrix} -3 \\ 7 \end{pmatrix}$, $x_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, $x_3 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, $x_4 = \begin{pmatrix} -5 \\ 6 \end{pmatrix}$, and $x_5 = \begin{pmatrix} -4 \\ 5 \end{pmatrix}$. Assume $k = 2$ clusters and initialise your centroids as

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

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:

$$X = \begin{pmatrix} d & ? & 1 & 7 \\ ? & 6 & -3 & ? \\ ? & -12 & ? & 42 \end{pmatrix}.$$

Here d is the maximum of the seventh digit of your student ID number and 1. Depending on the rank, find a representation $UV^\top = X$ with suitable matrices U and V. **[6]**

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

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

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

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

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 $D_f(\tilde{z}, x) + \langle \tilde{z}, g(y) \rangle$ 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) Design a (Bregman) proximal gradient descent algorithm for the solution of the convex optimisation problem

$$\hat{x} = \arg\min_{x \in \mathbb{R}^k} \left\{ h(x) \quad \text{subject to} \quad x \in [0,1]^k, \text{ and } \sum_{j=1}^k x_j = 1 \right\}.$$

Here $h : \mathbb{R}^k \to \mathbb{R}$ is a convex and continuously differentiable function. **Hint**: make use of Question 3(c). **[9]**

End of Paper.