Queen Mary
University of London

Late-Summer Examination period 2020

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, notes, and the internet. You may use calculators and computers, but you should show your working for any calculations you do. You must not seek or obtain help from anyone else.

At the start of your work, please **copy out and sign** the following declaration:

> I declare that my submission is entirely my own, and I have not sought or obtained help from anyone else.

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

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

- scan your work, convert it to a **single PDF file** and upload this using the upload tool on the QMplus page for the module;
- 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 not expected to spend a long time working on this assessment. We expect you to spend about **3 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 assessment period, 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 function $\mathrm{rank}(L)$ returns the rank of a matrix L. All computations should be done by hand where possible, with marks being awarded for intermediate steps in order to discourage computational aids.

Question 1 [35 marks].

(a) Explicitly derive the expectation for an absolutely continuous random variable X that follows an exponential distribution, i.e. the probability density functions reads

$$\rho(x) := \begin{cases} \lambda \exp(-\lambda x) & x \geq 0 \\ 0 & \text{otherwise} \end{cases},$$

where λ is one added to the seventh digit of your student ID. **[7]**

(b) Explicitly derive the variance for an absolutely continuous random variable X that follows an exponential distribution. **[7]**

(c) For a uniform (and absolutely continuous) random variable X on $[0,1]$ compute the expectation of $f(X)$ for

$$f(x) := \begin{cases} d_1\,(1 - \log(x)) & x \in [0, 1/d_2] \\ 0 & \text{otherwise} \end{cases},$$

where d_1 is one added to the eighth digit and d_2 is one added to the seventh digit of your student ID. Make use of the convention $0 \log(0) = 0$. **[7]**

(d) Compute the gradient of the function $J : [-1,1]^n \to [-1,1]^n$ defined as $J(x) := -\sum_{j=1}^n \sqrt{1 - x_j^2}$. **[7]**

(e) Explicitly derive the Bregman distance $D_J(x,y)$ with respect to the function J as defined in Question 1 (d). **[7]**

Question 2 [35 marks].

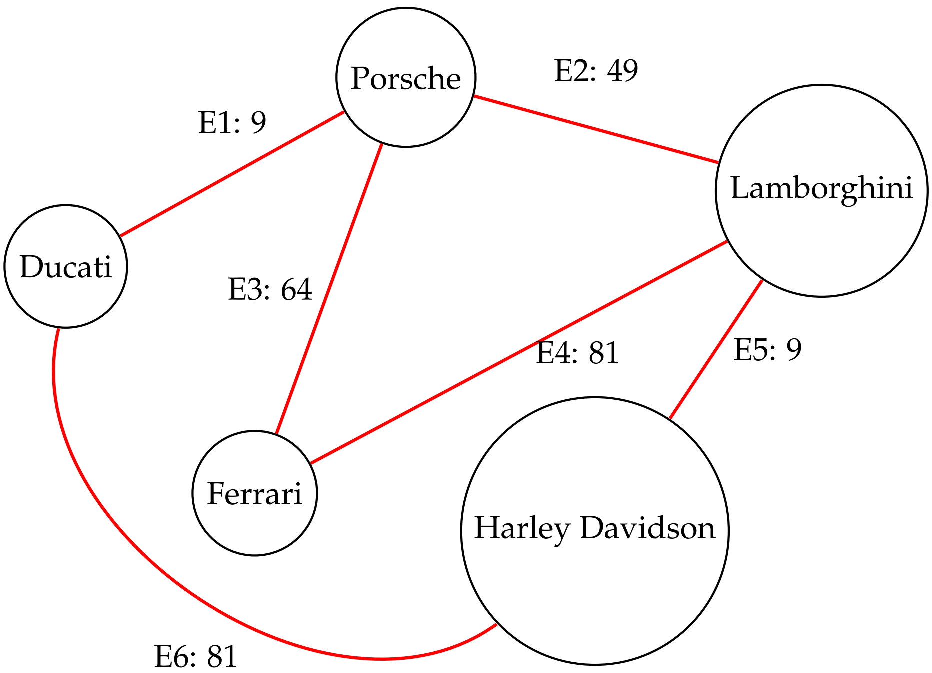
(a) You want to decide whether or not to ease lock-down restrictions. You base your decision on three factors:

- Are people required to wear face-masks when going outside?
- Have covid-19 case numbers dropped in previous weeks?
- Is comprehensive testing and tracing in place?

You do not ease lock-down measures unless case numbers have dropped. However, a drop in case numbers is only sufficient for easing lock-down measures if either people are required to wear face masks when going outside or if comprehensive testing and tracing is also in place.

Model this binary decision process with a perceptron and choose appropriate weights to mimic the decision process accurately. Justify your choice of weights. **[7]**

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



Order the columns of the incidence matrix alphabetically according to the vertex name and the rows according to the edge numbering (E1, E2, E3, ...). **[6]**

(c) Compute the corresponding graph Laplacian for the incidence matrix in Question 2(b). **[7]**

(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 "Lamborghini" is already labelled $v_{\text{Lamborghini}} = 1$ (class "cars") and the node "Ducati" is labelled as $v_{\text{Ducati}} = 0$ (class "motorbikes"). Determine the labels for all remaining nodes following the procedure explained in the lecture notes, and classify each node. **[8]**

(e) Manually determine parameters $w_1 \in \mathbb{R}$, $w_2 \in \mathbb{R}^2$, $b_1 \in \mathbb{R}$ and $b_2 \in \mathbb{R}$ of a neural network of the form

$$f(x_1, x_2) = w_1 \max\left(0, w_2^\top \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + b_2\right) + b_1$$

so that it mimics the logical NAND function, i.e.

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

.

[7]

Question 3 [30 marks].

(a) Perform two steps of k-means clustering by hand for the five data points $x_1 = -13$, $x_2 = -17$, $x_3 = 0$, $x_4 = 3$, and $x_5 = -2$. Assume $k = 2$ clusters and initialise your variables as

$$z_0 := \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}^\top ,$$

and

$$\mu_0 := \begin{pmatrix} d \\ 0 \end{pmatrix} ,$$

where d is one added to the eighth digit of your student ID number. For each iteration, update the variable z^l first, and then μ^l. Here $l \in \{1, 2\}$ denotes the iteration index. Does the iteration converge? **[8]**

(b) Find vectors $u \in \mathbb{R}^2$ and $v \in \mathbb{R}^3$ such that the following identity is satisfied for all known values:

$$uv^\top = \begin{pmatrix} -6 & 12 & ? \\ 2 & -4 & d \end{pmatrix} .$$

Here d is one added to the seventh digit of your student ID number. What value do you obtain at the missing entry denoted by a question mark? **[7]**

(c) Compute an approximation $\hat{L} \in \mathbb{R}^{2\times 3}$ with $\operatorname{rank}(\hat{L}) = 1$ of the matrix

$$X := d \begin{pmatrix} -2 & 3 & 2 \\ 2 & 2 & 3 \end{pmatrix}$$

by hand, where d is one added to the last digit of your student ID, that satisfies $\|\hat{L} - X\|_{\text{Fro}} \leq \|L - X\|_{\text{Fro}}$, for all $L \in \mathbb{R}^{2\times 3}$ with $\operatorname{rank}(L) = 1$. **[8]**

(d) Formulate a proximal gradient descent algorithm for a matrix completion problem of the form

$$(\hat{L}, \hat{S}) = \arg \min_{L,S \in \mathbb{R}^{s\times n}} \left\{ \frac{1}{2} \|P_\Omega(L + S) - y\|^2 + \alpha_1 \|L\|_* + \alpha_2 \|S\|_1 \right\} .$$

Here $P_\Omega \in \mathbb{R}^{s\times n} \to \mathbb{R}^r$ is a (known) projection operator that projects the known entries, specified by the index set Ω, of its argument to a vector of length r. The vector of known entries is denoted by $y \in \mathbb{R}^r$, α_1 and α_2 are positive constants, and $\|\cdot\|_*$ and $\|\cdot\|_1$ are the nuclear norm and the matrix one norm known from the lecture respectively. How does the algorithm compare to the robust PCA method introduced in the lecture? **[7]**

End of Paper.