Queen Mary
University of London

Main Examination period 2020 – 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, 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 set of all natural numbers (starting from one) is denoted by $\mathbb{N}$. The function $\text{rank}(L)$ 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 [36 marks].

(a) Compute the expected value $\mathbb{E}_x$ of a (discrete) Poisson-distributed random variable X with probability

$$p_x := \exp(-\lambda)\frac{\lambda^x}{x!}, \quad x = 1, 2, \ldots, s$$

for a constant $\lambda > 0$. What is the solution for $s \to \infty$? **Hint**: Make use of the identity $\exp(\lambda) = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}$. **[6]**

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

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

where d is the maximum of the last digit of your student ID and 1. Make use of the convention $0 \log(0) = 0$. **[6]**

(c) Let X be a random variable with expectation μ and variance σ^2. Show that the variance of $aX + b$, where $a, b \in \mathbb{R}$, is

$$\text{Var}_x[ax + b] = a^2\sigma^2 .$$

[6]

(d) Verify that the gradient of the function $J(x) := \frac{1}{2}\langle Qx, x\rangle$, where $Q \in \mathbb{R}^{n\times n}$ is a (square) matrix, is $\nabla J(x) = \frac{1}{2}(Q + Q^\top)x$. What does the gradient simplify to if Q is also symmetric? **[6]**

(e) Compute the Bregman distance with respect to the function $J(x) = \frac{1}{2}\langle Qx, x\rangle$, where $Q \in \mathbb{R}^{n\times n}$ is a (square) matrix. **[6]**

(f) If Q in Question 1(e) is a symmetric, positive semi-definite matrix, the function J is guaranteed to be convex for all arguments. What does this imply for the corresponding Bregman distance? **[6]**

Question 2 [34 marks].

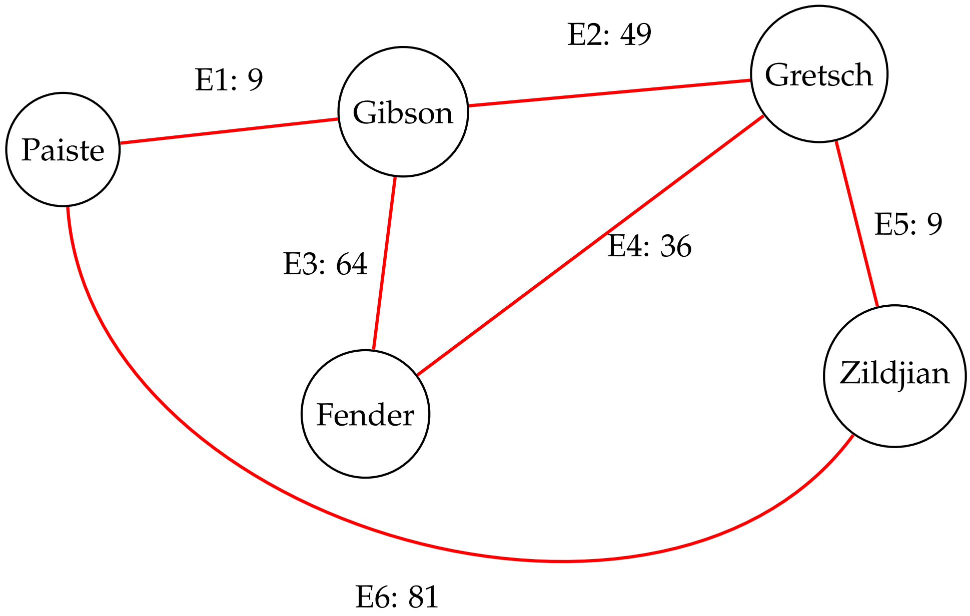
(a) You want to decide whether or not to become a data scientist. You base your decision on three factors:

- Am I excited about machine learning?
- Does becoming a data scientist involve understanding complicated mathematics?
- Do companies likely hire data scientists?

Suppose you do not mind if becoming a data scientist involves understanding complicated mathematics as long as many companies likely hire data scientists. However, you really would not want to become a data scientist if machine learning does not excite you.

Model this binary decision process with a perceptron and choose some appropriate weights to mimic the decision process accurately. **[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, ...). **[7]**

(c) Compute the corresponding graph Laplacian for the incidence matrix in Question 2(b). **[6]**

(d) We want to use the graph from Question 2(b) to determine whether a node in the graph belongs to the class "guitars" or the class "cymbals". Suppose we are in a semi-supervised setting, where the node "Fender" is already labelled $v_{\text{Fender}} = 1$ (class "guitars") and the node "Zildjian" is labelled as $v_{\text{Zildjian}} = 0$ (class "cymbals"). Determine the labels for all remaining nodes, and classify each node. **[8]**

(e) Determine manually some parameters $w \in \mathbb{R}^2$ and $b \in \mathbb{R}$ of a neural network of the form

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

that is supposed to mimic the logical AND function, i.e.

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

.

[6]

Question 3 [30 marks].

(a) Perform two steps of k-means clustering by hand for the six data points $x_1 = 1$, $x_2 = 3$, $x_3 = 0$, $x_4 = 15$, and $x_5 = 17$. 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 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. Did the iteration converge? **[8]**

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

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

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

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

$$X := d \begin{pmatrix} 6 & 4 & 4 \\ 4 & 6 & -4 \end{pmatrix} ,$$

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

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

$$\hat{L} = \arg \min_{L \in \mathbb{R}^{s\times n}} \left\{ \frac{1}{2} \|P_\Omega L - y\|_{\mathrm{Fro}}^2 \quad \text{subject to} \quad \mathrm{rank}(L) \leq k \right\} .$$

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 $y \in \mathbb{R}^r$ and $k \in \mathbb{N}$ is a fixed constant that determines the rank of $\hat{L}$. What does the proximal mapping (in the proximal gradient descent) look like? What is its closed-form solution? **[8]**

End of Paper.