Queen Mary
University of London

Main Examination period 2022 – January – Semester A

MTH793: 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**.

The exam is available for a period of **24 hours**. Upon accessing the exam, you will have **4 hours** in which 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.

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: N. Otter, M. Poplavskyi

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.

Note: All the problems that are similar to coursework problems refer to coursework that was also assigned during the previous iteration of this module, in Semester B of the academic year 2020/21, therefore the level of difficulty should be the same for students who are taking this exam as a resit.

Question 1 [38 marks].

(a) Perform two steps of the K-means algorithm for the data points

$$x_1 = (0,1)^t\,, x_2 = (1,0)^t\,, x_3 = (1,4)^t\,, x_4 = (-1,-3)^t\,, x_5 = (-2,-2)^t$$

and for $K = 2$. Let the centroids be

$$\mu_1^0 = (0,d)^t \text{ and } \mu_2^0 = (-d,0)\,,$$

where d is the maximum between the fifth digit of your student ID and one. **[10]**

Solution: *This problem is similar to a coursework problem. Since the problem statement only asks to perform two iterations, the difficulty of the problem does not depend on d.*

We provide a solution for $d = 6$.

We thus have $\mu_1^0 = (0,6)^t$ and $\mu_2^0 = (-6,0)$.

First iteration. We first compute the Euclidean distances between the data points and the centroids, and assign each data point to the centroid closest to it. We obtain the matrix of assignments

$$z^1 = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

and the centroids

$$\mu_1^1 = \frac{x_1 + x_2 + x_3}{3} = \frac{1}{3}(2,5)^t$$
$$\mu_2^1 = \frac{x_4 + x_5}{2} = \frac{1}{2}(-3,-5)^t\,.$$

Second iteration. We compute the Euclidean distances between the data points and the centroids μ_1^1 and μ_2^1, and assign each data point to the centroid closest to it. We obtain the matrix of assignments

$$z^2 = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}\,.$$

Since the assignments are the same as the ones obtained during the first iteration, the centroids do not change, and we have

$$\mu_1^2 = \mu_1^1$$
$$\mu_2^2 = \mu_2^1\,.$$

(b) Compute the rank of the following matrix: **[2]**

$$A = \begin{bmatrix} 0 & 1 & 0 & 1 \\ \sqrt{2} & 2 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

Solution: *This problem is testing the students' understanding of basic concepts from linear algebra needed in the lecture, which have been covered in the previous module MTH786, and which were revised during a special revision session during the lectures. The problem itself has not been assigned as a coursework problem, and is therefore new.* We put A in RREF:

$$RREF(A) = \begin{bmatrix} 1 & 0 & 0 & -2/\sqrt{2} \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and therefore the rank of A is two.

(c) Find a matrix B with rank 1 that is the best approximation of matrix A with respect to the Frobenius norm. **[10]**

Solution: *This problem is similar to a coursework problem.* We know that we can obtain such a matrix by computing $u_1 u_1^t A$, where u_1 is the first column of U, and U comes from the SVD decomposition $U\Sigma V^t$ of A. For this, we compute the eigenvalues and eigenvectors of AA^t. We have

$$AA^t = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 6 & 2 \\ 2 & 2 & 2 \end{bmatrix}$$

The characteristic polynomial of this matrix is
$-\lambda^3 + 10\lambda^2 - 16\lambda = -\lambda(\lambda - 8)(\lambda - 2)$, and thus we get the eigenvalues $\lambda_1 = 8$, $\lambda_2 = 2$ and $\lambda_3 = 0$. The vector u_1 is the eigenvector corresponding to the largest eigenvalue, thus to λ_1.

We compute the kernel of $AA^t - \lambda_1 I_{3\times 3}$ and obtain the eigenvector $(1, 2, 1)^t$, which normalised gives $u_1 = (1/\sqrt{6}, 2/\sqrt{6}, 1/\sqrt{6})^t$. Thus we obtain

$$B = \begin{bmatrix} 1/6 & 2/6 & 1/6 \\ 2/6 & 4/6 & 2/6 \\ 1/6 & 2/6 & 1/6 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 1 \\ \sqrt{2} & 2 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 2\sqrt{2} & 6 & 0 & 2 \\ 4\sqrt{2} & 12 & 0 & 4 \\ 2\sqrt{2} & 6 & 0 & 2 \end{bmatrix} .$$

(d) Complete the following matrix by choosing non-zero values for the missing entries so that it has **minimal** rank. Here d is the maximum between the eighth digit of your student ID and 1. **[8]**

$$A = \begin{bmatrix} ? & -3 & 3 & d \\ ? & ? & 6 & ? \\ -2 & ? & ? & ? \end{bmatrix}$$

Solution: *This problem is similar to a coursework problem. The solution and difficulty do not depend on the value of d.* The minimal rank of this matrix is one. One can choose the missing entries so that the second row is 2 times the first row, and the third row is a non-zero multiple of the first row. A possible solution is, for instance:

$$A = \begin{bmatrix} 1 & -3 & 3 & d \\ 2 & -6 & 6 & 2d \\ -2 & 6 & -6 & -2d \end{bmatrix}.$$

(e) Complete the following matrix by choosing non-zero values for the missing entries so that it has **maximal** rank. Here d is the maximum between the eighth digit of your student ID and 1. [**8**]

$$A = \begin{bmatrix} 1 & 0 & 4 & d \\ 2 & 1 & 8 & ? \\ 0 & 0 & 1 & ? \end{bmatrix}$$

Solution: *This problem asks the students to transfer their knowledge acquired during the module to a new type of problem they have not seen before. The solution and difficulty do not depend on the value of d.*

We put the matrix in RREF and obtain:

$$\begin{bmatrix} 1 & 0 & 0 & d-4f \\ 0 & 1 & 0 & e-2d \\ 0 & 0 & 1 & f \end{bmatrix}$$

with d being the maximum between 1 and the eigth digit of the student's ID and e, f being two real numbers to be determined. The reduced matrix contains three pivots, independently from the values of d, e, f, and thus A will have rank 3 for any choice of values. Thus, one possible solution, assuming that $d = 6$, and choosing $e = 10$ and $f = 1$ is:

$$\begin{bmatrix} 1 & 0 & 4 & 6 \\ 2 & 1 & 8 & 10 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

Question 2 [21 marks].

(a) Rewrite the function $f(x) = \frac{x^2-1}{3x+2}$ as a function of the form $g(a) + \epsilon h(a,b)$ where $x = a + \epsilon b$ with $\epsilon^2 = 0$. [**8**]

Solution: *This problem is similar to a coursework problem.* We compute

$$\begin{aligned} f(a+\varepsilon b) &= \frac{(a+\epsilon b)^2 - 1}{3(a+\epsilon b)+2} \\ &= \frac{a^2 - 1 + \epsilon 2ab}{3a + 2 + \epsilon 3b} \\ &= \frac{u+\epsilon v}{w+\epsilon z}\,, \end{aligned}$$

for $u = a^2 - 1$, $v = 2ab$, $w = 3a + 2$ and $z = 3b$. We further compute

$$\begin{aligned} f(a+\varepsilon b) &= \frac{u+\varepsilon\, v}{w+\varepsilon\, z} \\ &= \frac{(u+\varepsilon\, v)(w-\varepsilon\, z)}{(w+\varepsilon\, z)(w-\varepsilon\, z)} = \frac{uw+\varepsilon(vw-uz)}{w^2} \\ &= \frac{u}{w} + \varepsilon\frac{vw-uz}{w^2} \\ &= \frac{a^2-1}{3a+2} + \varepsilon\frac{2ab(3a+2)-(a^2-1)3b}{9a^2+4+12a} \\ &= \frac{a^2-1}{3a+2} + \varepsilon\frac{3a^2b+4ab+3b}{9a^2+12a+4} \\ &= g(a) + \varepsilon h(a,b)\,, \end{aligned}$$

for $g(a) = (a^2 - 1)/(3a + 2)$ and $h(a,b) = (3a^2b + 4ab + 3b)/(9a^2 + 12a + 4)$.

(b) Compute the derivative $f'(a)$ by using your result from Question 2(a). [**6**]

Solution: *This problem is similar to a coursework problem.* From the lecture notes we know that $h(a, 1) = f'(a)$. Hence, the derivative of f w.r.t. the argument a is

$$f'(a) = h(a,1) = \frac{3a^2+4a+3}{9a^2+12a+4}\,.$$

(c) Consider the function $f\colon \mathbb{R}^n \to \mathbb{R}^n$, with $f(x) = (x^t x)x$. Rewrite $f(x)$ as a function of the form $g(a) + \epsilon h(a,b)$, where $x = a + \epsilon b$ is a vector of dual numbers. Specify both g and h, and compute the partial derivates $\frac{\partial}{\partial_i} f(a_1, \ldots, a_n)$ for all i. [**7**]

Solution: *This problem is similar to a coursework problem.* We have

$$f(a+\epsilon b) = \sum_{i=1}^n (a_i+\epsilon b_i)^2(a+\epsilon b) = \sum_{i=1}^n (a_i^2 + \epsilon 2a_i b_i)(a+\epsilon b)\,,$$

and thus the jth component of this vector is

$$f(a+\epsilon b)_j = a_j \sum_{i=1}^n a_i^2 + \epsilon\left(b_j \sum_{i=1}^n a_i^2 + 2a_j \sum_{i=1}^n a_i b_i\right) =: g(a)_j + \epsilon h(a,b)_j$$

If we specify $e_l = (0\ \dots\ 0\ \underbrace{1}_{l-\text{th position}}\ 0\ \dots\ 0)$, we can compute the l-th partial derivative by evaluating $h(a, e_l)$, which reads

$$h(a, e_l) = \sum_{i=1}^{n} a_i^2 + 2a_j \sum_{i=1}^{n} a_i b_i \,.$$

Question 3 [8 marks]. Determine the parameters of a one-layer perceptron with two inputs that models the IMPLY function f, namely

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

.

[8]

Solution: *This problem asks the students to apply methods learned during the module to a problem that is slightly different from a problem seen in the coursework.* We can model the IMPLY function f as the following perceptron:

$$f(x_1, x_2) = \begin{cases} 1 & \text{if } \sum_{i=1}^{2} w_i x_i \geq b \\ 0 & \text{if } \sum_{i=1}^{2} w_i x_i < b \end{cases}$$

where the bias is $b = 0$, and the weights are given by $w_1 = -1$ and $w_2 = 1$.
I will accept any choice of bias and weights that models the function correctly.

Question 4 [33 marks].

(a) Write down the incidence matrix of the weighted undirected graph in Figure 1. 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 $(E_1, E_2, E_3, \dots)$. [7]

Solution: *This problem is similar to a coursework problem.* The incidence matrix for the displayed graph is

$$I = \left(\begin{array}{c|ccccc} \text{E1} & 0 & 0 & 0 & -2 & 2 \\ \text{E2} & -9 & 0 & 0 & 9 & 0 \\ \text{E3} & -1 & 0 & 0 & 0 & 1 \\ \text{E4} & -2 & 0 & 2 & 0 & 0 \\ \text{E5} & 0 & 0 & -2 & 2 & 0 \\ \text{E6} & 0 & -1 & 0 & 1 & 0 \\ \text{E7} & 0 & 0 & -8 & 0 & 8 \\ \text{E8} & 0 & -8 & 8 & 0 & 0 \\ \text{E9} & 0 & -9 & 0 & 0 & 9 \\ \hline & \text{beluga} & \text{black dolphin} & \text{melon-headed whale} & \text{narwhal} & \text{panda doplhin} \end{array} \right)$$

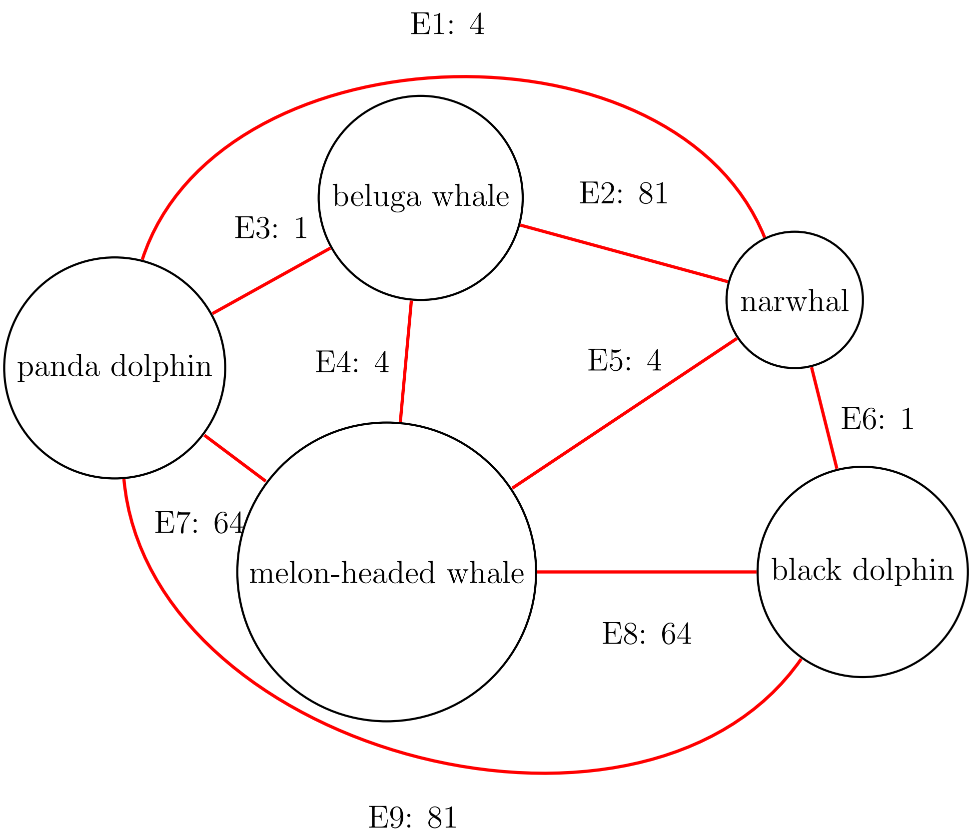

Figure 1: A weighted undirected graph.

(b) Compute the graph Laplacian for the incidence matrix in Question 4(a). **[8]**

Solution: *This problem is similar to a coursework problem.* The graph Laplacian is

$$L = I^t I = \begin{bmatrix} 86 & 0 & -4 & -81 & -1 \\ 0 & 146 & -64 & -1 & -81 \\ -4 & -64 & 136 & -4 & -64 \\ -81 & -1 & -4 & 90 & -4 \\ -1 & -81 & -64 & -4 & 150 \end{bmatrix} .$$

(c) We want to use the graph from Question 4(a) to determine whether a node in the graph belongs to the class "dolphin" or the class "whale". Suppose we are in a semi-supervised setting, where the node "narwhal" is already labelled $v_{\text{narwhal}} = 1$ (class "whale") and the node "black dolphin" is labelled as $v_{\text{black dolphin}} = 0$ (class "dolphin"). Determine the remaining labels with the same procedure as discussed in the lecture notes and classify each node. **[8]**

Solution: *This problem is similar to a coursework problem.* From the lecture notes we know that the label vector $v \in \mathbb{R}^5$ can be decomposed as

$$v = P_{R^\perp}^\top \tilde{v} + P_R^\top y ,$$

where P_R denotes the projection of v onto the known indices, and $P_{R^\perp}$ onto the unknown indices. The known indices are denoted by y, the unknown by $\tilde{v}$. For

$$v = \begin{pmatrix} v_{\text{beluga}} \\ v_{\text{black dolphin}} \\ v_{\text{melon-headed whale}} \\ v_{\text{narwhal}} \\ v_{\text{panda dolphin}} \end{pmatrix}$$

we know the second and fourth entry; the second belongs to the class "dolphin" and therefore takes on the value $v_{\text{black dolphin}} = 0$, whereas the fourth entry belongs to the class "whales", hence $v_{\text{narwhal}} = 1$. Thus, for $y = \begin{pmatrix} 0 & 1 \end{pmatrix}^\top$ we have

$$v = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tilde{v} + \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} y \,.$$

From the lecture notes we also know that we can estimate $\tilde{v}$ via

$$\begin{aligned} \tilde{v} &= \arg\min_{\overline{v}} \left\| M_w \left(P_{R^\perp}^\top \overline{v} + P_R^\top y \right) \right\|^2 , \\ &= -\left(P_{R^\perp} L P_{R^\perp}^\top \right)^{-1} \left(P_{R^\perp} L P_R^\top y \right) , \end{aligned}$$

which for our matrices reads

$$\begin{pmatrix} -86 & 4 & 1 \\ 4 & -136 & 64 \\ 1 & 64 & -150 \end{pmatrix} \tilde{v} = \begin{pmatrix} -81 \\ -4 \\ -4 \end{pmatrix} ,$$

Solving this linear system leads to the (approximate) solution

$$\tilde{v} \approx \begin{pmatrix} 0.9469 \\ 0.0910 \\ 0.0718 \end{pmatrix} .$$

Rounding all values above 1/2 to one and below 1/2 to zero then yields the classification

$$v = \begin{pmatrix} v_{\text{beluga}} \\ v_{\text{black dolphin}} \\ v_{\text{melon-headed whale}} \\ v_{\text{narwhal}} \\ v_{\text{panda dolphin}} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} .$$

(d) Now suppose that we are in an unsupervised setting, and we do not know any node label. Use the second eigenvector of the graph Laplacian to divide the nodes of the graph into two classes. **[10]**

Solution: *This problem is similar to a coursework problem.* The eigenvector corresponding to the second smallest eigenvalue of the Laplacian is, approximately:

$$v = \begin{bmatrix} -1.53 \\ 1.04 \\ 0.9 \\ -1.4 \\ 1 \end{bmatrix}$$

If we set all entries that are larger than zero to 1 and the remaining entries to 0 we obtain

$$v = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 1 \end{bmatrix},$$

and we thus obtain the same classes as in Question 4(c).

End of Paper.