Queen Mary
University of London

Main Examination period 2022 – May/June – Semester B

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**.

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:

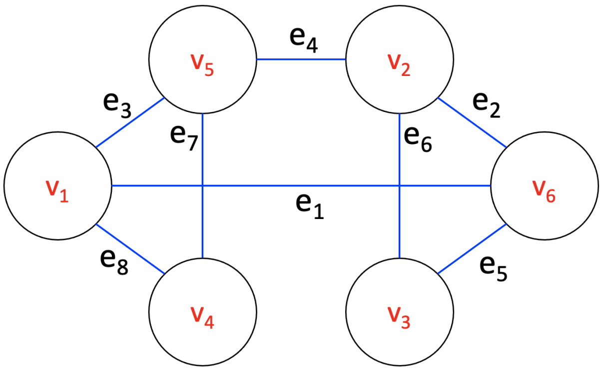
- scan your work, convert it to a **single PDF file**, and submit this file using the tool below the link to the exam;
- e-mail a copy to **maths@qmul.ac.uk** with your student number and the module code in the subject line;
- with your e-mail, include a photograph of the first page of your work together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

(1) For involved mathematical computations (e.g., inverting a matrix, computing eigenvectors, etc.) you are encouraged to use a calculator or a computer, unless stated otherwise in the problem. You should make it clear where you used a computer and how.
(2) When asked to perform a certain machine-learning task (e.g., K-means, PCA), you should present all the steps of execution for the algorithm, and **not** to run the algorithm on a computer.

Question 1 [35 marks]. Graph clustering
Let G be the following graph:



We assign weights to the edges in the following way: $w(e_i) = i$.

(a) Write down the graph Laplacian L for this weighted graph. Make sure the rows of the Laplacian have the same ordering as the vertices. **[5]**

(b) Suppose that we are given that v_1 is labelled as '0', and v_6 as '1'. Use the semi-supervised graph labelling method discussed in class to label all the other vertices. You may use a computer to solve linear systems, but should explain all the steps of how you arrive at the system, and how to interpret the output. **[10]**

(c) Suppose we modify the weights of e_2, e_4 to be $w(e_2) = 20$ and $w(e_4) = 40$. Repeat the labelling procedure above with the new weights. What is the new graph labelling that you get? Provide an intuitive explanation of your results. **[5]**

(d) In this part we assume **no label information** is given, with the original weights on the edges ($w(e_i) = i$). Compute the eigenvector of L corresponding to the second smallest eigenvalue, and use it to divide the vertices into $k = 2$ clusters. Do your result agree with part (b)? **[10]**

(e) Suppose that the true clusters for this graph are given by:

$$C_1^* = \{v_1, v_4, v_5\}, \quad C_2^* = \{v_2, v_3\}, \quad C_3^* = \{v_6\}.$$

Compute the Rand Index, to rank the quality of the clustering result you received in (d). Show all the steps in the calculation. **[5]**

Solution:

(a) The weight and degree matrices are:

$$W = \begin{pmatrix} 0 & 0 & 0 & 8 & 3 & 1 \\ 0 & 0 & 6 & 0 & 4 & 2 \\ 0 & 6 & 0 & 0 & 0 & 5 \\ 8 & 0 & 0 & 0 & 7 & 0 \\ 3 & 4 & 0 & 7 & 0 & 0 \\ 1 & 2 & 5 & 0 & 0 & 0 \end{pmatrix}, \quad D = \begin{pmatrix} 12 & 0 & 0 & 0 & 0 & 0 \\ 0 & 12 & 0 & 0 & 0 & 0 \\ 0 & 0 & 11 & 0 & 0 & 0 \\ 0 & 0 & 0 & 15 & 0 & 0 \\ 0 & 0 & 0 & 0 & 14 & 0 \\ 0 & 0 & 0 & 0 & 0 & 8 \end{pmatrix}.$$

Therefore, the Laplacian is:

$$L = D - W = \begin{pmatrix} 12 & 0 & 0 & -8 & -3 & -1 \\ 0 & 12 & -6 & 0 & -4 & -2 \\ 0 & -6 & 11 & 0 & 0 & -5 \\ -8 & 0 & 0 & 15 & -7 & 0 \\ -3 & -4 & 0 & -7 & 14 & 0 \\ -1 & -2 & -5 & 0 & 0 & 8 \end{pmatrix},$$

(b) We first write the projection matrices:

$$P_L = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad P_U = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Let $y = (0, 1)^T$ be the known label, and let g be the values we assign the unlabelled vertices. Then g is the solution for

$$A \cdot g = b,$$

where

$$A = P_U \cdot L \cdot P_U^T = \begin{pmatrix} 12 & -6 & 0 & -4 \\ -6 & 11 & 0 & 0 \\ 0 & 0 & 15 & -7 \\ -4 & 0 & -7 & 14 \end{pmatrix} \quad b = -P_U \cdot L \cdot P_L^T \cdot y = (2, 5, 0, 0)^T.$$

Solving this equation yields,

$$g = (0.6532, 0.8109, 0.1136, 0.2434)^T$$

The labelling vector we conclude will then be

$$(0, 1, 1, 0, 0, 1)^T.$$

(c) Repeating the same procedure as before, we now hav

$$g = (0.7977, 0.8897, 0.3186, 0.6828)^T,$$

implying the following labels

$$(0, 1, 1, 0, 1, 0)^T.$$

In other words, for the new weights we have that v_5 changed its label from 0 to 1. The reason is that the increase in the weights, implies that the connection between v_5 and v_6, much stronger than the connection between v_5 and v_1. Therefore, v_5 takes the label from v_6 now.

(d) The second eigenvector of L is

$$u_2 = (0.4399, -0.2707, -0.4703, 0.4572, 0.3032, -0.4594)^T.$$

Using the signs of the values we have the following labels

$$(1, 0, 0, 1, 1, 0)^T.$$

These labels concur with part (b).

(e) The true clusters are:

$$C_1^* = \{v_1, v_4, v_5\}, \quad C_2^* = \{v_2, v_3\}, \quad C_3^* = \{v_6\}.$$

The following edges are true-positives:

$$(1, 4), (1, 5), (2, 3), (4, 5).$$

The following edges are true-negatives:

$$(1, 2), (1, 3), (1, 6), (2, 4), (2, 5), (3, 4), (3, 5), (4, 6), (5, 6).$$

Therefore the Rand index is:

$$\mathrm{RI} = \frac{\mathrm{TP} + \mathrm{TN}}{15} = \frac{4 + 9}{15} = 0.8667.$$

Since this index is close to 1, the results are considered good.

Question 2 [25 marks]. K-means clustering

In this problem, you are asked to run the k-means algorithm **by hand**, showing all the steps and calculations. You may use a calculator if you find it helpful. For your convenience, you can use the tables on the last page of this exam paper, where you can document the results of each step in the iteration of the algorithm. If you do, make sure to attach them to your submitted pdf. Note that the tables may be longer than the number of steps required.

(a) Suppose that we are given the following dataset in $\mathbb{R}^2$:

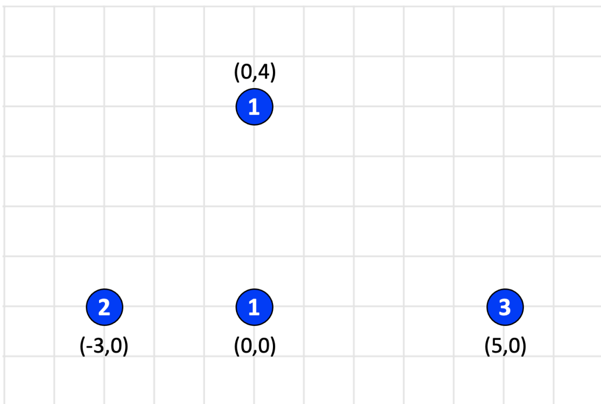
$$x_1 = (0,0), \quad x_2 = (0,4), \quad x_3 = (-3,0), \quad x_4 = (5,0).$$

Run the k-means algorithm (until it converges) with $k = 2$ and with initial centroids $\mu_1^{(0)} = (0,0)$, $\mu_2^{(0)} = (-3,0)$. Explain the steps in each iteration. What are the resulting clusters and centroids? **[5]**

(b) Next, we assume we have data points at the same locations as in (a), but with some points appearing multiple times. Our new dataset is:

$$x_1 = (0,0), \quad x_2 = (0,4), \quad x_3 = x_4 = (-3,0), \quad x_5 = x_6 = x_7 = (5,0).$$

This is summarised in the following figure (the numbers in the circles are the number of points placed at each location).



Repeat part (a) for the new dataset, with the same initial centroids as in (a). **[10]**

(c) Repeat part (b) but for initial centroids $\mu_1^{(0)} = (0,0)$ and $\mu_2^{(0)} = (0,4)$. **[5]**

(d) Are your clustering results from parts (b) and (c) the same?
If yes – is it true that running k-means always yields the same results? (regardless of initial centroids). **If no** – explain how is that possible. **[5]**

Solution:

(a) The results:

	assignments				centroids	
	x_1	x_2	x_3	x_4	μ_1	μ_2
Step 1:	1	1	2	1	$(\frac{5}{3}, \frac{4}{3})$	$(-3, 0)$
Step 2:	1	1	2	1	$(\frac{5}{3}, \frac{4}{3})$	$(-3, 0)$

Since the assignments haven't changed between step 1 and step 2, the algorithm stops. The output clusters are: $C_1 = \{x_1, x_2, x_4\}$, and $C_2 = \{x_3\}$. The centroids are $\mu_1 = (\frac{5}{3}, \frac{4}{3})$, $\mu_2 = (-3, 0)$.

(b) The iterations for the current input are given by:

	assignments							centroids	
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	μ_1	μ_2
Step 1:	1	1	2	2	1	1	1	$(3, \frac{4}{5})$	$(-3, 0)$
Step 2:	2	1	2	2	1	1	1	$(\frac{15}{4}, 1)$	$(-2, 0)$
Step 3:	2	2	2	2	1	1	1	$(5, 0)$	$(-\frac{3}{2}, 1)$
Step 4:	2	2	2	2	1	1	1	$(5, 0)$	$(-\frac{3}{2}, 1)$

Since the assignments haven't changed between step 1 and step 2, the algorithm stops. The output clusters are: $C_1 = \{x_5, x_6, x_7\}$, and $C_2 = \{x_1, x_2, x_3, x_4\}$. The centroids are $\mu_1 = (5, 0)$, $\mu_2 = (-\frac{3}{2}, 1)$.

(c) The iterations for the current input are given by:

	assignments							centroids	
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	μ_1	μ_2
Step 1:	1	2	1	1	1	1	1	$(\frac{3}{2}, 0)$	$(0, 4)$
Step 2:	1	2	1	1	1	1	1	$(\frac{3}{2}, 0)$	$(0, 4)$

Since the assignments haven't changed between step 1 and step 2, the algorithm stops. The output clusters are: $C_1 = \{x_1, x_3, x_4, x_5, x_6, x_7\}$, and $C_2 = \{x_2\}$. The centroids are $\mu_1 = (\frac{3}{2}, 0)$, $\mu_2 = (0, 4)$.

(d) The k-means algorithm is known to converge to a local minimum, which might be different depending on the initial centroids.

Question 3 [25 marks]. SVD and PCA

In this problem you must show all your calculations.

(a) Consider the matrix:

$$M = \begin{pmatrix} 1 & 2 & -1 & 0 & -2 \\ 2 & 1 & 0 & -2 & -1 \end{pmatrix}$$

Find the matrices U and Σ of the SVD: $M = U\Sigma V^T$ (no need to find V).
HINT: Use the matrix $M \cdot M^T$. **[10]**

(b) Find the solution to the following problem

$$\tilde{M} = \underset{M' \in \mathbb{R}^{2\times 5}}{\arg\min} \|M - M'\|_F, \quad \text{subject to} \quad \text{rank}(M') = 1.$$

where $\|\cdot\|_F$ stands for the Frobenius norm. **[5]**

(c) You are given a set of points $\{x_1, x_2, x_3, x_4, x_5\} \subset \mathbb{R}^2$, where

$$x_1 = (1,2), x_2 = (2,1), x_3 = (-1,0), x_4 = (0,-2), x_5 = (-2,-1).$$

Denote by $\hat{x}_1, \ldots, \hat{x}_5$ the lower dimensional approximation of the dataset using the first principal component. Find the coordinates of the points $\hat{x}_1, \ldots, \hat{x}_5$, and in a single axis system provide a sketch for the following:

- The original points x_1, x_2, x_3, x_4, x_5.
- Their lower dimensional approximation $\hat{x}_1, \hat{x}_2, \hat{x}_3, \hat{x}_4, \hat{x}_5$.
- The line that represents the direction of the principal component.

[10]

Solution:

(a) We start by finding MM^T:

$$MM^T = \begin{pmatrix} 1 & 2 & -1 & 0 & -2 \\ 2 & 1 & 0 & -2 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 \\ 2 & 1 \\ -1 & 0 \\ 0 & -2 \\ -2 & -1 \end{pmatrix} = \begin{pmatrix} 10 & 6 \\ 6 & 10 \end{pmatrix}.$$

The characteristic polynomial is:

$$\det(MM^T - \lambda I) = \det \begin{pmatrix} 10-\lambda & 6 \\ 6 & 10-\lambda \end{pmatrix} = (10-\lambda)^2 - 36 = \lambda^2 - 20\lambda + 64 = (\lambda-16)(\lambda-4).$$

The eigenvalues of MM^T are therefore $\lambda_1 = 16, \lambda_2 = 4$. We find the eigenvectors:

$$MM^T - \lambda_1 I = \begin{pmatrix} -6 & 6 \\ 6 & -6 \end{pmatrix}.$$

The kernel of this matrix is any vector of the form $\alpha \cdot (1,1)^T$. The unit eigenvector is therefore $u_1 = \sqrt{1/2}(1,1)^T$.

$$MM^T - \lambda_2 I = \begin{pmatrix} 6 & 6 \\ 6 & 6 \end{pmatrix}.$$

The kernel of this matrix is any vector of the form $\alpha \cdot (1,-1)^T$. The unit eigenvector is therefore $u_2 = \sqrt{1/2}(1,-1)^T$.
We saw in class that the singular values in this case are

$$\sigma_1 = \sqrt{\lambda_1} = 4, \quad \sigma_2 = \sqrt{\lambda_2} = 2 \quad \Rightarrow \quad \Sigma = \begin{pmatrix} 4 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \end{pmatrix}.$$

The matrix U consists of u_1, u_2 found above. Therefore,

$$U = \sqrt{1/2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

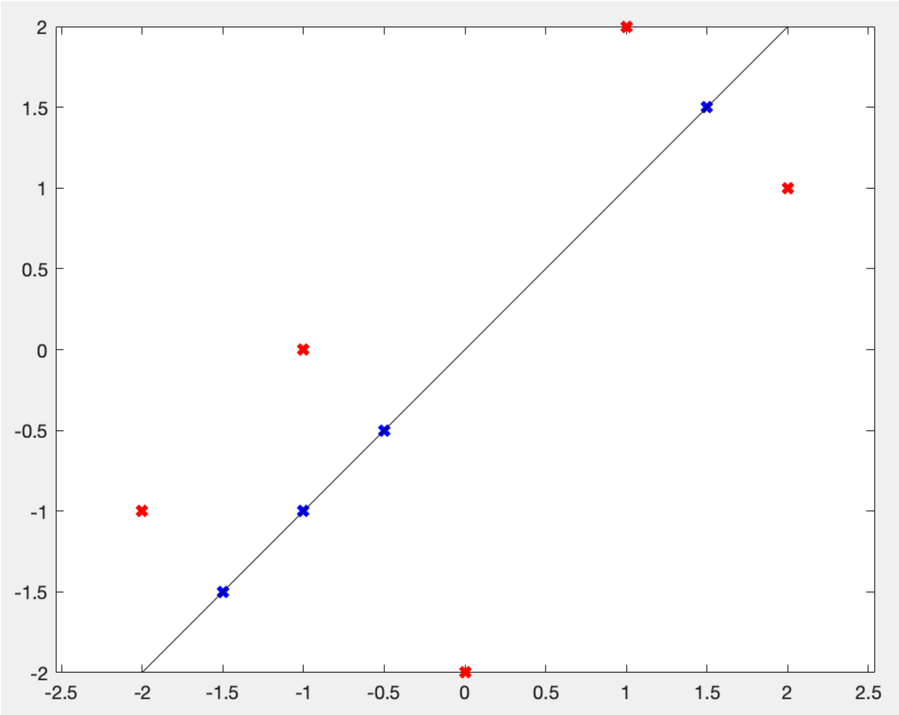
(b) The best rank 1 approximation for M is:

$$v_1 \cdot v_1^T \cdot M = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 2 & -1 & 0 & -2 \\ 2 & 1 & 0 & -2 & -1 \end{pmatrix} = \begin{pmatrix} \frac{3}{2} & \frac{3}{2} & -\frac{1}{2} & -1 & -\frac{3}{2} \\ \frac{3}{2} & \frac{3}{2} & -\frac{1}{2} & -1 & -\frac{3}{2} \end{pmatrix}.$$

(c) Note that the stacked data matrix for the given points is exactly M we had before. The approximation of the points via the first principal component is given by the rank 1 approximation of M, which we computed in part (b). Therefore, we have

$$\hat{x}_1 = \left(\frac{3}{2}, \frac{3}{2}\right), \hat{x}_2 = \left(\frac{3}{2}, \frac{3}{2}\right), \hat{x}_3 = \left(-\frac{1}{2}, -\frac{1}{2}\right), \hat{x}_4 = \left(-1, -1\right), \hat{x}_5 = \left(-\frac{3}{2}, -\frac{3}{2}\right),$$

and the line that fits the data is $y = x$. This is drawn in the following figure.


2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5
2
2.5

Question 4 [15 marks]. Matrix completion

(a) Suppose you are given the following matrix with missing entries:

$$M = \begin{pmatrix} 1 & ? & ? & 2 & 1 \\ 0 & 1 & 2 & -1 & 0 \\ ? & 3 & ? & -3 & 0 \\ 1 & 4 & 2 & 1 & ? \end{pmatrix}.$$

Can the above matrix be completed to be rank 1? Explain your answer. [**5**]

(b) Find the missing values in order for the resulting matrix to have minimum rank. [**10**]

Solution:

(a) Notice that the first two lines of M will be independent, no matter how we complete the missing values. Therefore, the rank of M will be at least 2.

(b) The minimum rank possible for M is 2. We will denote the rows of M by R_i $(i = 1, 2, 3, 4)$. As the R_1, R_2 are independent, to get rank 2, we need R_3, R_4 to be a liner combination R_1,R_2. First, we will complete R_3 to have $R_3 = 3R_1 = (0, 3, 6, -3, 0)$. For R_4 it cannot be a multiple of either R_1 or R_2 but we can try complete it so $R_4 = R_1 + R_2$. In other words, we want to have

$$(1, M_{11} + 1, M_{12} + 2, 1, 1) = (1, 4, 2, 1, M_{44}).$$

The solution is $R_1 = (1, 3, 0, 2, 1)$ and $R_4 = (1, 4, 2, 1, 1)$. The complete matrix is:

$$M = \begin{pmatrix} 1 & 3 & 0 & 2 & 1 \\ 0 & 1 & 2 & -1 & 0 \\ 0 & 3 & 6 & -3 & 0 \\ 1 & 4 & 2 & 1 & 1 \end{pmatrix}.$$

End of Paper.

Table for Question 2(a):

	assignments				centroids	
	x_1	x_2	x_3	x_4	μ_1	μ_2
Step 1:						
Step 2:						
Step 3:						
Step 4:						
Step 5:						

Table for Question 2(b):

	assignments							centroids	
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	μ_1	μ_2
Step 1:									
Step 2:									
Step 3:									
Step 4:									
Step 5:									

Table for Question 2(c):

	assignments							centroids	
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	μ_1	μ_2
Step 1:									
Step 2:									
Step 3:									
Step 4:									
Step 5:									