Queen Mary
University of London

Late-Summer Examination period 2022

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:

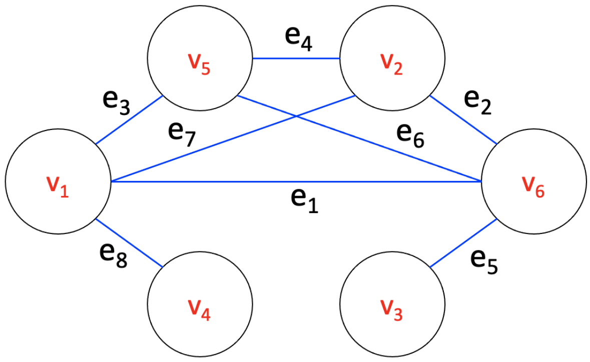
- scan your work, convert it to a **single PDF file**, and submit this file using the tool below the link to the exam;
- e-mail a copy to **maths@qmul.ac.uk** with your student number and the module code in the subject line;
- with your e-mail, include a photograph of the first page of your work together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

Question 1 [25 marks]. Graph clustering
Let G be the following graph:



We assign weights to the edges in the following way: $w(e_i) = 10 - i$.

(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 both v_2 and v_6 as '1'. Use the semi-supervised graph labelling method discussed in class to label all the other vertices. You are allowed to use a computer to solve a linear system, but should explain all the steps leading to this system, and how to interpret the output. **[10]**

(c) Suppose we modify the weights of e_5, e_8 to be $w(e_5) = 0.000005$ and $w(e_8) = 4,000,000$. Without running the algorithm again, explain what will be the effect of this change on the results. **[5]**

(d) Suppose that we repeat part (b), but now with v_1, v_2, v_6 all labelled as '1'. Without running the algorithm again, explain what will be the effect on the results. **[5]**

Solution:

(a) The weight and degree matrices are:

$$W = \begin{pmatrix} 0 & 3 & 0 & 2 & 7 & 9 \\ 3 & 0 & 0 & 0 & 6 & 8 \\ 0 & 0 & 0 & 0 & 0 & 5 \\ 2 & 0 & 0 & 0 & 0 & 0 \\ 7 & 6 & 0 & 0 & 0 & 4 \\ 9 & 8 & 5 & 0 & 4 & 0 \end{pmatrix}, \quad D = \begin{pmatrix} 21 & 0 & 0 & 0 & 0 & 0 \\ 0 & 17 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 17 & 0 \\ 0 & 0 & 0 & 0 & 0 & 26 \end{pmatrix}.$$

Therefore, the Laplacian is:

$$L = D - W = \begin{pmatrix} 21 & -3 & 0 & -2 & -7 & -9 \\ -3 & 17 & 0 & 0 & -6 & -8 \\ 0 & 0 & 5 & 0 & 0 & -5 \\ -2 & 0 & 0 & 2 & 0 & 0 \\ -7 & -6 & 0 & 0 & 17 & -4 \\ -9 & -8 & -5 & 0 & -4 & 26 \end{pmatrix},$$

(b) We first write the projection matrices:

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

Let $y = (0, 1, 1)^T$ 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

$$A = P_U \cdot L \cdot P_U^T = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 17 \end{pmatrix} \quad b = -P_U \cdot L \cdot P_L^T \cdot y = (5, 0, 10)^T.$$

Solving this equation yields,

$$g = (1, 0, 10/17)^T$$

The labelling vector we conclude will then be

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

(c) Note that v_4's only neighbor is v_1 whose label is '0'. Therefore, no matter what the weight of e_8 would be, we will always label it as '0'. Similarly, v_3's only neighbor is v_6 whose label is '1', and therefore v_3 will always be labeled as '1'.

(d) Since all given labels are '1' the algorithm aims to label their neighbours the same way. Therefore, all vertices will be labelled as '1'.

Question 2 [25 marks]. K-means clustering

Consider the set of data points:

$$x_1 = (1,1)^T, \quad x_2 = (2,3)^T, \quad x_3 = (-1,-2)^T, \quad x_4 = (4,4)^T, \quad x_5 = (-3,-3)^T.$$

(a) Perform the first two steps of the K-means algorithm for these points, with $k = 2$ and the initial centroids given by:

$$\mu_1 = (0,3)^T, \quad \mu_2 = (-2,-2)^T.$$

You should run the k-means algorithm **by hand**, and not use any software. You may use a calculator if you find it helpful. [**15**]

(b) Let C_1, C_2 be the two clusters produced by the K-means algorithm in part (a). Compute the Dunn-Index (DI) for these clusters, using the single-linkage inter-cluster distance, and the diameter intra-cluster distance. [**10**]

Solution:

(a) The results:

	assignments					centroids	
	x_1	x_2	x_3	x_4	x_5	μ_1	μ_2
Step 1:	1	1	2	1	2	$(\frac{7}{3}, \frac{8}{3})$	$(-2,-2.5)$
Step 2:	1	1	2	1	2	$(\frac{7}{3}, \frac{8}{3})$	$(-2,-2.5)$

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 $C_2 = \{x_3, x_5\}$. The centroids are $\mu_1 = (\frac{7}{3}, \frac{8}{3})$, $\mu_2 = (-2,-2.5)$.

(b) For the inter-cluster distance we compute:

$$\begin{aligned} \|x_1 - x_3\| &= 3.6056, & \|x_1 - x_5\| &= 5.6569, \\ \|x_2 - x_3\| &= 5.8310, & \|x_2 - x_5\| &= 7.8102, \\ \|x_4 - x_3\| &= 7.8102, & \|x_4 - x_5\| &= 9.8995. \end{aligned}$$

Therefore

$$\delta(C_1, C_2) = 3.6056.$$

For the intra-cluster distance we compute:

$$\begin{aligned} &\|x_1 - x_2\| = 2.2361, \quad \|x_1 - x_4\| = 4.2426, \quad \|x_2 - x_4\| = 2.2361 \\ &\|x_3 - x_5\| = 2.2361. \end{aligned}$$

Therefore

$$\Delta(C1) = 4.2426, \quad \Delta(C_2) = 2.2361.$$

and we conclude that

$$\mathrm{DI} = \frac{3.6065}{4.2426} = 0.8499.$$

Question 3 [35 marks]. **SVD and PCA**

In this problem you are **not** allowed to use a computer, except for part (c).

(a) Consider the matrix:

$$M = \begin{pmatrix} 3 & 4 & 0 & 0 \\ 5 & 0 & 0 & 0 \end{pmatrix}$$

Find the matrices U,Σ, and V of the SVD: $M = U\Sigma V^T$.
HINT: Use the matrices MM^T, M^TM. **[10]**

(b) You are given a set of points $\{x_1, x_2, x_3, x_4, x_5\} \subset \mathbb{R}^3$, where

$$x_1 = (3, 2, 3), x_2 = (2, 3, -2), x_3 = (-1, 3, 2), x_4 = (1, -2, 0), x_5 = (0, -1, 2).$$

We want to find the best fit for a line approximating these points using PCA. We will break it into a few steps:

(i) Find the empirical mean of the dataset, denoted $\bar{x}$.

(ii) Centre the data points $x_1, \ldots, x_5$ using $\bar{x}$, and stack the resulting vectors as columns in a matrix called X.

(iii) Using the SVD decomposition of X, find the principal components (directions) of X.

(iv) Find the projection of X onto its first principal component.

(v) Write down the resulting approximation, denoted $\hat{x}_1, \ldots, \hat{x}_5$ (don't forget to fix the mean).

Note: In this part you are allowed to use a computer to compute eigenvectors/singular vectors, but do **not** use any implemented PCA routine. **[15]**

(c) Recall the definition of the singular value thresholding operator:

$$X = U\Sigma V^T \longrightarrow D_\tau(X) = US_\tau(\Sigma)V^T,$$

where S_τ is the soft-thresholding function.

Considering the data matrix X from part (c) – what is rank(X)? What is rank$(D_4(X))$ **[10]**

Solution:

(a) Here

$$M = \begin{pmatrix} 3 & 4 & 0 & 0 \\ 5 & 0 & 0 & 0 \end{pmatrix}.$$

We start by finding MM^T:

$$MM^T = \begin{pmatrix} 25 & 15 \\ 15 & 25 \end{pmatrix}.$$

The characteristic polynomial is:

$$\det(MM^T - \lambda I) = \det \begin{pmatrix} 25-\lambda & 15 \\ 15 & 25-\lambda \end{pmatrix} = (25-\lambda)^2 - 225 = \lambda^2 - 50\lambda + 400 = (\lambda - 40)(\lambda - 10).$$

The eigenvalues of MM^T are therefore $\lambda_1 = 40, \lambda_2 = 10$.
We saw in class that the singular values in this case are

$$\sigma_1 = \sqrt{\lambda_1} = \sqrt{40}, \quad \sigma_2 = \sqrt{\lambda_2} = \sqrt{10} \quad \Rightarrow \quad \Sigma = \begin{pmatrix} \sqrt{40} & 0 & 0 & 0 & 0 \\ 0 & \sqrt{10} & 0 & 0 & 0 \end{pmatrix}.$$

We find the eigenvectors:

$$MM^T - \lambda_1 I = \begin{pmatrix} -15 & 15 \\ 15 & -15 \end{pmatrix}.$$

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

$$MM^T - \lambda_2 I = \begin{pmatrix} 15 & 15 \\ 15 & 15 \end{pmatrix}.$$

The kernel of this matrix is any vector of the form $\alpha \cdot (1,-1)^T$. The unit eigenvector is therefore $u_2 = \sqrt{1/2}(1,-1)^T$. The matrix U consists of u_1, u_2 found above. Therefore,

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

To find V we will consider $M^T M$:

$$M^T M = \begin{pmatrix} 34 & 12 & 0 & 0 \\ 12 & 16 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Taking $\tilde{v}_1 = (2,1,0,0)^T$ we have

$$(M^T M)\tilde{v}_1 = (80, 40, 0, 0)^T = 40\tilde{v}_1,$$

therefore $\tilde{v}_1$ is an eigenvector for λ_1. Nomalising it we get

$$v_1 = \frac{\tilde{v}_1}{\|\tilde{v}_1\|} = \frac{1}{\sqrt{5}}(2,1,0,0)^T.$$

Next, taking $\tilde{v}_2 = (1, -2, 0, 0)^T$ we have

$$(M^T M)\tilde{v}_2 = (10, -20, 0, 0)^T = 10\tilde{v}_2,$$

Taking $v_2 = \frac{1}{\sqrt{5}}(1, -2, 0, 0)^T$ we that $v_1^T v_2 = 0$. To complete the matrix V we need eigenvectors for $\lambda = 0$. For example we can take

$$v_3 = (0, 0, 1, 0)^T, \quad v_4 = (0, 0, 0, 1)^T.$$

Therefore, we have

$$V = \begin{pmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} & 0 & 0 \\ \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

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

$$v_1 \cdot v_1^T \cdot M = \frac{1}{2}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 3 & 4 & 0 & 0 \\ 5 & 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 4 & 2 & 0 & 0 \\ 4 & 2 & 0 & 0 \end{pmatrix}.$$

(c) The empirical mean is:

$$\bar{x} = (1, 1, 1)^T.$$

The data matrix is therefore

$$X = \begin{pmatrix} 2 & 1 & -2 & 0 & -1 \\ 1 & 2 & 2 & -3 & -2 \\ 2 & -3 & 1 & -1 & 1 \end{pmatrix}$$

The matrix U in the SVD of X is

$$U = \begin{pmatrix} 0.4019 & 0.0455 & -0.9146 \\ 0.8250 & 0.4155 & 0.3832 \\ 0.3974 & -0.9085 & 0.1294 \end{pmatrix}$$

Therefore, the principal components are

$$\begin{aligned} u_1 &= (0.4019, 0.8250, 0.3974)^T, \\ u_2 &= (0.0455, 0.4155, -0.9085)^T, \\ u_3 &= (-0.9145, 0.3832, 0.1294)^T. \end{aligned}$$

The projection on the first PC is given by

$$\hat{X} = u_1 u_1^T X = \begin{pmatrix} 0.9739 & 0.3454 & 0.4998 & -1.1543 & -0.6648 \\ 1.9993 & 0.7091 & 1.0259 & -2.3696 & -1.3648 \\ 0.9631 & 0.3416 & 0.4942 & -1.1415 & -0.6575 \end{pmatrix}$$

The resulting approximations are therefore

$$\begin{aligned} \hat{x}_1 &= (0.9739, 1.9993, 0.9631)^T + (1, 1, 1)^T = (1.9739, 2.9993, 1.9631)^T \\ \hat{x}_2 &= (0.3454, 0.7091, 0.3416)^T + (1, 1, 1)^T = (1.3454, 1.7091, 1.3416)^T \\ \hat{x}_3 &= (0.4998, 1.0259, 0.4942)^T + (1, 1, 1)^T = (1.4998, 2.0259, 1.4942)^T \\ \hat{x}_4 &= (-1.1543, -2.3696, -1.1415)^T + (1, 1, 1)^T = (-0.1543, -1.3696, -0.1415)^T \\ \hat{x}_5 &= (-0.6648, -1.3648, -0.6575)^T + (1, 1, 1)^T = (0.3352, -0.3648, 0.3425)^T. \end{aligned}$$

Question 4 [15 marks]. **Robust PCA & Matrix completion**

(a) Given a matrix

$$M = \begin{pmatrix} 2 & 1 & 4 & 3 \\ 4 & 21 & 8 & 6 \\ -2 & -1 & -4 & 13 \end{pmatrix}$$

find the decomposition $M = L + E$ where E is a sparse matrix (with at most 3 nonzero entries), and L is a low-rank matrix (lowest rank possible). [**10**]

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

$$M = \begin{pmatrix} 1 & ? & ? & ? \\ 0 & ? & 4 & 0 \\ 0 & ? & ? & 1 \end{pmatrix}$$

Can the above matrix be completed to be rank 2? Explain your answer. [**5**]

Solution:

(a) We can write

$$M = \begin{pmatrix} 2 & 1 & 4 & 3 \\ 4 & 21 & 8 & 6 \\ -2 & -1 & -4 & 13 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 4 & 3 \\ 4 & 2 & 8 & 6 \\ -2 & -1 & -4 & -3 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 19 & 0 & 0 \\ 0 & 0 & 0 & 16 \end{pmatrix},$$

where

$$L = \begin{pmatrix} 2 & 1 & 4 & 3 \\ 4 & 2 & 8 & 6 \\ -2 & -1 & -4 & -3 \end{pmatrix}$$

is rank-1 matrix, and

$$E = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 19 & 0 & 0 \\ 0 & 0 & 0 & 16 \end{pmatrix},$$

is indeed sparse.

(b) If the matrix is rank 2 (or less), then there exists a linear combination of the rows of the form: $aR_1 + bR_2 + cR_3 = 0$. For the first element we have: $a \cdot 1 + b \cdot 0 + c \cdot 0 = 0$, therefore $a = 0$. For the last element we have $b \cdot 0 + c \cdot 1 = 0$, therefore $c = 0$. Finally, taking the third element we have $b \cdot 4 = 0$, so also $b = 0$. In other words, the three rows are necessarily independent. Therefore we cannot complete the matrix to be rank 2.

End of Paper.