Queen Mary
University of London

Main Examination period 2023 – May/June – Semester B

MTH793P: Advanced machine learning

Duration: 4 hours

The exam is available for a period of **4 hours**, within which you must complete the assessment and submit your work. **Only one attempt is allowed – once you have submitted your work, it is final**.

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

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.**

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.

Examiners: 1^{st} **Dr. N. Perra,** 2^{nd} **Dr. N. Otter**

Question 1 [20 marks].

You are given the three data points coloured in blue in Figure 1. You are tasked to add two points, the red points in Figure 1, by interpolation. Considering the problem as a semi-supervised prediction task, complete the vector $\mathbf{y} = (1, ?, 3, ?, 2)$ which represents the y-coordinates of blue (known) and red (unknown) points.
In particular:

(a) By ordering nodes from left to right, write down the incidence matrix $\mathbf{M}$, and using the incidence matrix write down the Laplacian matrix $\mathbf{L}$. **[5]**

(b) We know that the Laplacian matrix can be expressed as $\mathbf{L} = \mathbf{D} - \mathbf{A}$ where $\mathbf{D}$ is a diagonal matrix whose elements are the degree of each node (i.e., data point) and $\mathbf{A}$ is the adjacency matrix. Write down the expressions for these two matrices and verify that $\mathbf{L} = \mathbf{D} - \mathbf{A}$. In doing so, keep the order of nodes as before: from left to right. **[5]**

(c) We know that this problem leads to a normal equation of the form

$$P_{I_1/I_2}^{\top} \mathbf{L} P_{I_1/I_2} \hat{\mathbf{w}} = -P_{I_1/I_2}^{\top} \mathbf{L} P_{I_2} \mathbf{v}$$

where $\mathbf{v}$ is the vector of known y-values. Write down the expressions for P_{I_2} and P_{I_1/I_2}. **[5]**

(d) Find $\hat{\mathbf{w}}$ by solving the normal equation. Show all the steps and do not just infer/read the values from the figure **[5]**

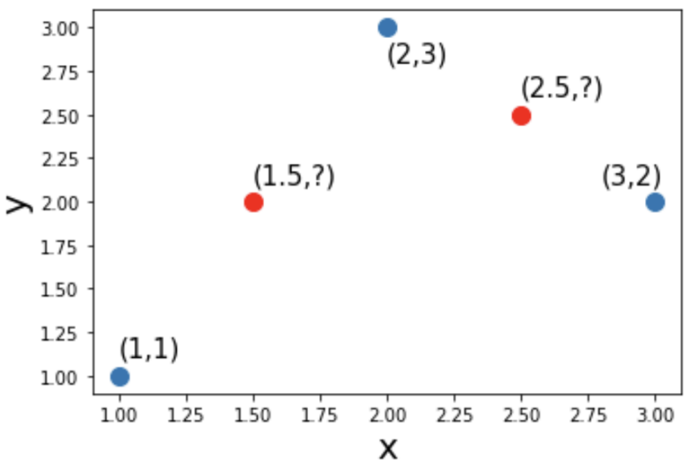

Figure 1: Known points are shown in blue. The red points describe new points we need to add by interpolation.

Solution:

(a) *This is a variation of a problem discussed in the coursework.* The incidence matrix $\mathbf{M}$ reads

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

The Laplacian matrix $\mathbf{L}$ is defined as $\mathbf{L} = \mathbf{M}^\top \mathbf{M}$ hence

$$\mathbf{L} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & 1 \end{pmatrix} \tag{2}$$

(b) *This is a variation of a problem discussed in the coursework.* The matrix $\mathbf{D}$ is equal to $\mathbf{D} = diag(1, 2, 2, 2, 1)$. The adjacency matrix is instead

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \tag{3}$$

hence $\mathbf{D} - \mathbf{A}$ is indeed equal to the Laplacian.

(c) *This is a variation of a problem discussed in the coursework.* P_{I_2} is a 5×3 matrix equal to

$$P_{I_2} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{4}$$

as it **selects** the known values of y which are the data points with ids equal to $\{1, 3, 5\}$. P_{I_1/I_2} is a 5×2 matrix equal to

$$P_{I_1/I_2} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{5}$$

since the unknown values of y are correspondent to the data points with ids equal to $\{2, 4\}$. Finally we have that $\mathbf{v} = (1, 3, 2)^\top$

(d) *This is a variation of a problem discussed in the coursework.* The normal equation reads

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \hat{\mathbf{w}} = \begin{pmatrix} 4 \\ 5 \end{pmatrix} \tag{6}$$

which implies $\hat{\mathbf{w}} = \left(2, \frac{5}{2}\right)$

Question 2 [25 marks].

Consider the five data points plotted in Figure 2. Assume that they are ordered clockwise such that $\mathbf{p_1} = (1,1)^\top$, $\mathbf{p_2} = (2,1)^\top$, $\mathbf{p_3} = (-1,-1)^\top$, $\mathbf{p_4} = (-2,-1)^\top$, $\mathbf{p_5} = (-1,1)^\top$.

(a) Write the pairwise distance matrix $\mathbf{d}$ whose elements d_{ij} describe the Euclidian distance between point $\mathbf{p_i}$ and $\mathbf{p_j}$. **[10]**

(b) Starting from the distance matrix $\mathbf{d}$ let us build a graph $\mathbf{G}$ such that:

- data points are connected only if $d_{ij} \leqslant 2$,
- the weight of each pair of connected nodes is defined as $w_{ij} = d_{ij}^{-1}$,

draw the graph, write down the adjacency matrix $\mathbf{A}$, the diagonal matrix $\mathbf{D}$, and the Laplacian $\mathbf{L}$ of the correspondent graph. **[5]**

(c) just by looking at the Laplacian matrix what can we say about i) the value of the its second smallest eigenvalue λ_2, ii) the graph $\mathbf{G}$. **[5]**

(d) Consider a connected graph $\mathbf{G}(N, E)$ characterised by the adjacency matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$, the diagonal degree matrix $\mathbf{D} \in \mathbb{R}^{N\times N}$ and the Laplacian matrix $\mathbf{L} \in \mathbb{R}^{N\times N}$. Consider a vector $\mathbf{p} \in \mathbb{R}^{N\times 1}$ that satisfies the following equation $\mathbf{p} = \mathbf{A}\mathbf{D}^{-1}\mathbf{p}$. Prove that $\mathbf{p} = c\mathbf{D}\mathbb{1}$, where c is a constant and $\mathbb{1} \in \mathbb{R}^{N\times 1}$ is a vector whose components are all equal to one. **Hint**: Use the properties of the Laplacian matrix. **[5]**

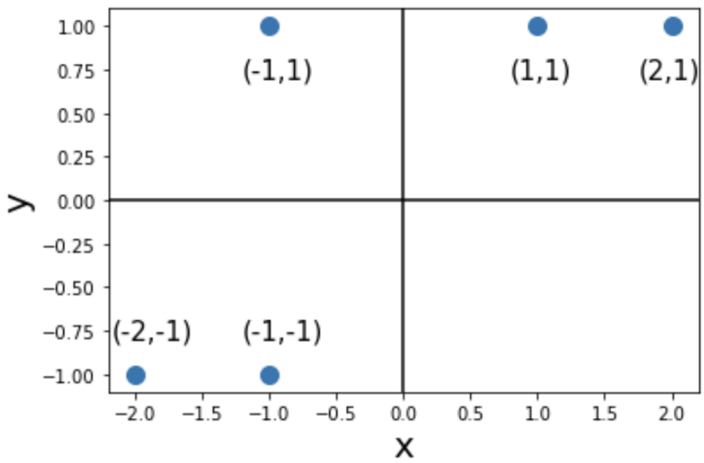

Figure 2: Data points

Solution:

(a) *This is a variation of a problem discussed in the coursework.* The pairwise distance matrix reads as

$$\mathbf{d} = \begin{pmatrix} 0 & 1 & 2\sqrt{2} & \sqrt{13} & 2 \\ 1 & 0 & \sqrt{13} & 2\sqrt{5} & 3 \\ 2\sqrt{2} & \sqrt{13} & 0 & 1 & 2 \\ \sqrt{13} & 2\sqrt{5} & 1 & 0 & \sqrt{5} \\ 2 & 3 & 2 & \sqrt{5} & 2 \end{pmatrix} \tag{7}$$

indeed for example the element $d_{1,3}$ is the distance between point $\mathbf{p}_1 = (1,1)^\top$ and $\mathbf{p}_3 = (-1,-1)^\top$ which is equal to

$$d_{1,3} = \|\mathbf{p}_1 - \mathbf{p}_3\| = \sqrt{(1+1)^2 + (1+1)^2} = \sqrt{8} = 2\sqrt{2} \tag{8}$$

analogously $d_{2,3}$ is the distance between point $\mathbf{p}_2 = (2,1)^\top$ and $\mathbf{p}_3 = (-1,-1)^\top$ which is equal to

$$d_{2,3} = \|\mathbf{p}_2 - \mathbf{p}_3\| = \sqrt{(2+1)^2 + (1+1)^2} = \sqrt{13} \tag{9}$$

(b) *This is a variation of a problem discussed in the coursework.* The graph is a line network that, by keeping the position of the nodes according their 2D coordinates, is shown in Figure 3 , The adjacency matrix reads

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & \frac{1}{2} \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \frac{1}{2} \\ 0 & 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \end{pmatrix} \tag{10}$$

The matrix $\mathbf{D}$ instead is $\mathbf{D} = diag\left(\frac{3}{2}, 1, \frac{3}{2}, 1, 1\right)$. Finally the Laplacian matrix reads

$$\mathbf{L} = \mathbf{D} - \mathbf{A} = \begin{pmatrix} \frac{3}{2} & -1 & 0 & 0 & -\frac{1}{2} \\ -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{3}{2} & -1 & -\frac{1}{2} \\ 0 & 0 & -1 & 1 & 0 \\ -\frac{1}{2} & 0 & -\frac{1}{2} & 0 & 1 \end{pmatrix} \tag{11}$$

(c) *We have discussed this point in several lectures.* The Laplacian matrix is not block diagonal, hence we know that the graph is not disconnected and $\lambda_2 > 0$. Furthermore, this does not depend on the choice of ids for the nodes. Since, the graph is connected, the kernel of the Laplacian matrix contains only the span of the constant vector.

(d) *We did not discuss problems like this one in the coursework. However, we discussed at length the properties of the Laplacian matrix with examples similar to this problem.*We can write $\mathbf{p} = \mathbf{A}\mathbf{D}^{-1}\mathbf{p}$ as:

$$
\begin{aligned}
\mathbf{p} &= \mathbf{A}\mathbf{D}^{-1}\mathbf{p} && (12)\\
\mathbf{p} &- \mathbf{A}\mathbf{D}^{-1}\mathbf{p} = 0 && (13)\\
(\mathbb{1} &- \mathbf{A}\mathbf{D}^{-1})\mathbf{p} = 0 && (14)\\
(\mathbf{D} &- \mathbf{A})\mathbf{D}^{-1}\mathbf{p} = 0 && (15)\\
& \mathbf{L}\mathbf{D}^{-1}\mathbf{p} = 0 && (16)
\end{aligned}
$$

from the last expression we can see that $\mathbf{D}^{-1}\mathbf{p}$ is an eigenvector of the Laplacian correspondent to eigenvalue zero. Since the graph is connected, the kernel of the Laplacian has size one and is constituted by the span of $\mathbb{1}$. Hence, we can write $\mathbf{D}^{-1}\mathbf{p} = c\mathbb{1}$. Multiplying on the left both sides of last expression for $\mathbf{D}$ we prove that $\mathbf{p} = c\mathbf{D}\mathbb{1}$

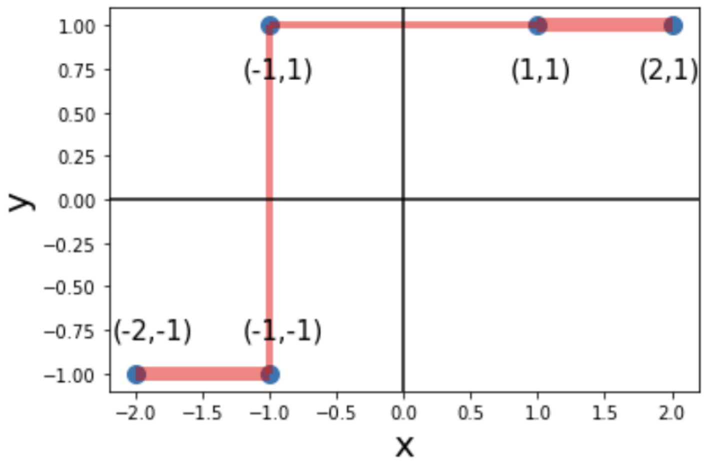

Figure 3: Plot of the graph $\mathbf{G}$

 Continue to next page

Question 3 [35 marks].

Consider the following six data points: $\mathbf{p_1} = (1,1)^\top$, $\mathbf{p_2} = (2,2)^\top$, $\mathbf{p_3} = (1,-1)^\top$, $\mathbf{p_4} = (-1,-1)^\top$, $\mathbf{p_5} = (-2,-2)^\top$, $\mathbf{p_6} = (-1,1)^\top$.

(a) Write down the correspondent $\mathbf{X} \in \mathbb{R}^{2\times 5}$ matrix and show that the data is already centred. **[5]**

(b) Find the principal components of $\mathbf{X}$. **[10]**

(c) Compute the projections $y_1, \ldots, y_6 \in \mathbb{R}^1$ of $\mathbf{p_1}, \ldots, \mathbf{p_6} \in \mathbb{R}^2$ on the first principal component. **[5]**

(d) Compute the reconstructions of $\mathbf{p_1}, \ldots, \mathbf{p_6}$ using the first principal components, denoted $\hat{\mathbf{p}}_1, \ldots, \hat{\mathbf{p}}_6 \in \mathbb{R}^2$. **[5]**

(e) In a 2D axis system, plot the following:

- The original points $x_1, \ldots, x_6$,
- The reconstructed points $\hat{x}_1, \ldots, \hat{x}_6 \in \mathbb{R}^2$,
- The principal components (directions).

[5]

(f) Consider the following matrix:

$$\mathbf{M} = \begin{pmatrix} 3 & 6 & 1 & 9 \\ 1 & 2 & \frac{1}{3} & 10+d \\ 2 & 14+d & \frac{2}{3} & 6 \end{pmatrix}$$

where d is the last digit of your student ID number. Find the decomposition $\mathbf{M} = \mathbf{L} + \mathbf{E}$ where $\mathbf{E}$ is a sparse matrix (with at most 3 nonzero entries), and $\mathbf{L}$ is a low-rank matrix (lowest rank possible). **[5]**

Solution:

(a) *This is a variation of a problem discussed in the coursework.* The data matrix $\mathbf{X}$ reads

$$\mathbf{X} = \begin{pmatrix} 1 & 2 & 1 & -1 & -2 & -1 \\ 1 & 2 & -1 & -1 & -2 & 1 \end{pmatrix} \tag{17}$$

(b) *This is a variation of a problem discussed in the coursework.* It is easy to see that $\bar{x} = \frac{1}{6}\sum_j X_{1j} = 0$ as well as $\bar{y} = \frac{1}{6}\sum_j X_{2j} = 0$ hence the data points are centred.

(c) *This is a variation of a problem discussed in the coursework.* In order to find the principal components we need to consider the singular value decomposition of the matrix $\mathbf{X}$ that is $\mathbf{X} = \mathbf{U\Sigma V}^\top$. To answer the question we can just compute the matrix $\mathbf{U}$ which is correspondent to the eigenvectors of the matrix $\mathbf{XX}^\top$. The matrix is

$$\mathbf{XX}^\top = \begin{pmatrix} 1 & 2 & 1 & -1 & -2 & -1 \\ 1 & 2 & -1 & -1 & -2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & 2 \\ 1 & -1 \\ -1 & -1 \\ -2 & -2 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 12 & 8 \\ 8 & 12 \end{pmatrix} \tag{18}$$

The eigenvalues of the matrix are $\sigma_1^2 = 20$ and $\sigma_2^2 = 4$. The correspondent eigenvectors are $\mathbf{u}_1 = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^\top$ and $\mathbf{u}_2 = \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^\top$. Hence, the matrix $\mathbf{U}$ is

$$\mathbf{U} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \tag{19}$$

The principal components are the two vectors $\mathbf{u}_1$ and $\mathbf{u}_2$.

(d) *This is a variation of a problem discussed in the coursework.* The projections can be easily obtained computing

$$\mathbf{Y} = \mathbf{u}_1^\top \mathbf{X} = \left(\sqrt{2}, 2\sqrt{2}, 0, -\sqrt{2}, -2\sqrt{2}, 0\right) \tag{20}$$

(e) *This is a variation of a problem discussed in the coursework.* The reconstruction of the points using the first principal component $\hat{\mathbf{p}}_1, \ldots, \hat{\mathbf{p}}_6$ is

$$\hat{\mathbf{X}} = \mathbf{u}_1 \mathbf{Y} = \begin{pmatrix} 1 & 2 & 0 & -1 & -2 & 0 \\ 1 & 2 & 0 & -1 & -2 & 0 \end{pmatrix} \tag{21}$$

In Figure 4 we show in blue the original points, in red the reconstructed points along the first principal component

(f) *This is a variation of a problem discussed in the coursework. The complexity of the problem does not change as function of d.*The matrix $\mathbf{M}$ can be written as:

$$\mathbf{M} = \mathbf{L} + \mathbf{E} = \begin{pmatrix} 3 & 6 & 1 & 9 \\ 1 & 2 & \frac{1}{3} & 3 \\ 2 & 4 & \frac{2}{3} & 6 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 7+d \\ 0 & 10+d & 0 & 0 \end{pmatrix} \tag{22}$$

where indeed the first matrix has rank 1 (as the second row can be obtained by multiplying the first by 1/3 and the third row by multiplying the first by 2/3).

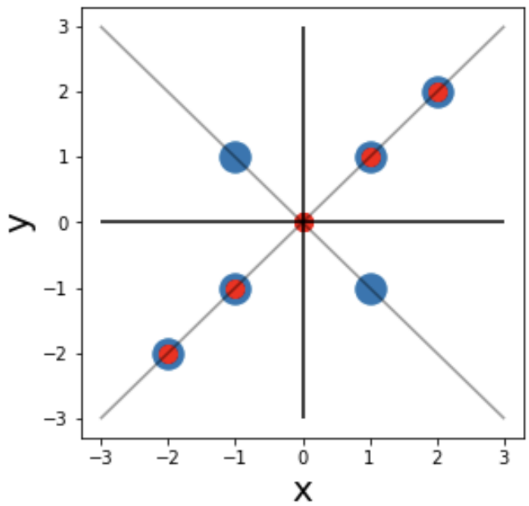

Figure 4: Original points in blue, reconstructed points in red and principal components shown in light grey

Question 4 [20 marks].

(a) Compute the nuclear norm and the rank of the matrix:

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

[10]

(b) Consider a matrix $\mathbf{M}$, show that $\|M\|_F^2 = \sum_{i=1}^{r} \sigma_i^2$ where the σ_i are the singular values. **[5]**

(c) Consider a matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$ and its singular value decomposition $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$. Prove that $c\mathbf{U}^\top\mathbf{X}\mathbf{V}\mathbf{1} = (c\sigma_1, \ldots, c\sigma_r)^\top$ where $\mathbf{1} \in \mathbb{R}^{n \times 1}$ is a constant vector with all components equal to one, and $c \in \mathbb{R}$. **[5]**

Solution:

(a) *We have discussed about nuclear norms and matrix rank in several lectures.* The rank is the number of non-zero singular values. The nuclear norm is the sum of the singular values. Hence, the solution implies computing the singular values of the matrix. To this end, it is convenient to consider $\mathbf{M}\mathbf{M}^\top$ (which is a 2×2 matrix):

$$\mathbf{M}\mathbf{M}^\top = \begin{pmatrix} 6 & 5 \\ 5 & 5 \end{pmatrix}$$

from which we can obtain the characteristic equation that leads us to $\sigma_1^2 = \frac{11+\sqrt{101}}{2}$ and $\sigma_2^2 = \frac{11-\sqrt{101}}{2}$. Hence, the rank of the matrix is two, and the nuclear norm reads:

$$\|M\|_* = \sum_{i=1}^{r} \sigma_i = \sqrt{\frac{11+\sqrt{101}}{2}} + \sqrt{\frac{11-\sqrt{101}}{2}} \tag{23}$$

(b) *We did not discuss problems like this one in the coursework. However, we discussed at length the properties of the Frobenius norm.* The Frobenius norm of a matrix can be written as

$$\|\mathbf{M}\|_F^2 = Trace(\mathbf{M}\mathbf{M}^\top) \tag{24}$$

considering now the SVD of the matrix $\mathbf{M} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$ we can write

$$Trace(\mathbf{M}\mathbf{M}^\top) = Trace(\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top\mathbf{V}\boldsymbol{\Sigma}^\top\mathbf{U}^\top) \tag{25}$$

which, due to the orthonormality of the singular vectors, leads us to

$$Trace(\mathbf{M}\mathbf{M}^\top) = Trace(\mathbf{U}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top\mathbf{U}^\top) \tag{26}$$

Considering the $Trace(\mathbf{AB}) = Trace(\mathbf{BA})$ we can write the right hand side us

$$Trace(\mathbf{M}\mathbf{M}^\top) = Trace(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top\mathbf{U}^\top\mathbf{U}) \tag{27}$$

hence

$$Trace(\mathbf{M}\mathbf{M}^\top) = Trace(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top) \tag{28}$$

Since the matrix $\boldsymbol{\Sigma}$ is diagonal (in the sense that the only elements different than zeros are the elements Σ_{ii} and that these are the singular values) the matrix $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top$ is a squared diagonal matrix containing the squares of the singular values. Hence the trace of such matrix is

$$Trace(\mathbf{M}\mathbf{M}^\top) = Trace(\boldsymbol{\Sigma}\boldsymbol{\Sigma}^\top) = \sum_{i=1}^{r} \sigma_i^2 \tag{29}$$

which proves the initial statement

(c) *We discussed a similar problem in one of the lectures*. Starting from the SVD of the matrix $\mathbf{X} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top$ we can multiply both sides by $\mathbf{U}^\top$ on the left and by $\mathbf{V}$ on the right obtaining

$$\mathbf{U}^\top\mathbf{X}\mathbf{V} = \mathbf{U}^\top\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top\mathbf{V} = \boldsymbol{\Sigma} \tag{30}$$

multiplying both sides, on the right, by a constant vector $c\mathbf{1}$ we get

$$c\mathbf{U}^\top\mathbf{X}\mathbf{V}\mathbf{1} = c\boldsymbol{\Sigma}\mathbf{1} \tag{31}$$

which leads to a vector $\mathbb{R}^{n\times 1}$ equal to $c(\sigma_1, \ldots, \sigma_r)^\top$

End of Paper.