Queen Mary
University of London

Main Examination period 2022 – January – Semester A

MTH786: Machine learning with Python

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: N. Perra

© Queen Mary University of London (2022)
Continue to next page

Question 1 [45 marks].

Consider the following matrix $\mathbf{A} = \begin{pmatrix} a & 1 \\ 1 & a \end{pmatrix}$ where $a \in \mathbb{R}$.

(a) Compute the eigenvalues and correspondent eigenvectors of $\mathbf{A}$. **[10]**

(b) Compute the inner product of the eigenvectors of $\mathbf{A}$. What do we learn from it? **[5]**

(c) Consider the following matrix $\mathbf{B} = \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix}$ compute its singular values. **[10]**

(d) Compute the left and right singular vectors of the matrix $\mathbf{B}$. **[10]**

(e) Consider a matrix $\mathbf{D} \in \mathbb{R}^{m,n}$, a vector $\mathbf{x} \in \mathbb{R}^{n,1}$, and the function $E(\mathbf{x}) = \frac{1}{2}\|\mathbf{D}\mathbf{x}\|^2$. Show that $\nabla E(\mathbf{x}) = \mathbf{D}^\top \mathbf{D}\mathbf{x}$. **[10]**

Solution:

(a) The eigenvectors and eigenvalues are the defined by $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ which is equivalent to $(\mathbf{A} - \lambda\mathbb{I})\mathbf{x} = 0$. This admits solutions, besides the trivial $\mathbf{x} = 0$, only if $\det[\mathbf{A} - \lambda\mathbb{I}] = 0$. Hence:

$$\mathbf{A} - \lambda\mathbb{I} = \begin{pmatrix} a-\lambda & 1 \\ 1 & a-\lambda \end{pmatrix} \tag{1}$$

Imposing the determinant to be equal zero, we obtain the characteristic equation $(a-\lambda)^2 - 1 = \lambda^2 - 2a\lambda + a^2 - 1 = 0$ which yields $\lambda_{1,2} = a \pm 1$. The eigenvectors can be derived imposing $\mathbf{A}\mathbf{x}_i = \lambda_i \mathbf{x}_i$:

$$\begin{pmatrix} a & 1 \\ 1 & a \end{pmatrix} \begin{pmatrix} x_1^{(1)} \\ x_1^{(2)} \end{pmatrix} = \begin{pmatrix} (a+1)x_1^{(1)} \\ (a+1)x_1^{(2)} \end{pmatrix} \tag{2}$$

and

$$\begin{pmatrix} a & 1 \\ 1 & a \end{pmatrix} \begin{pmatrix} x_2^{(1)} \\ x_2^{(2)} \end{pmatrix} = \begin{pmatrix} (a-1)x_2^{(1)} \\ (a-1)x_2^{(2)} \end{pmatrix} \tag{3}$$

From the first set of equations we obtain $x_1^{(1)} = x_1^{(2)}$ hence $\mathbf{x}_1 = (c, c) = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ where the last condition can be easily obtained imposing $\|\mathbf{x}_1\|^2 = 1$

From the second set of equations we obtain $x_2^{(1)} = -x_2^{(2)}$ hence $\mathbf{x}_2 = (c, -c) = (\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$ where the last condition can be easily obtained imposing $\|\mathbf{x}_2\|^2 = 1$. Note that since $x_2^{(1)} = -x_2^{(2)}$ also $\mathbf{x}_2 = (-c, c) = (-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ can be a solution. This is a vector pointing to the opposite direction with respect to the previous.

(b) The inner product $\mathbf{x}_2^\top \mathbf{x}_1 = \langle \mathbf{x}_2, \mathbf{x}_1 \rangle = 0$. The eigenvectors are orthogonal. This result is to be expected since the matrix is symmetric.

© Queen Mary University of London (2022) **Continue to next page**

(c) The singular values of the matrix $\mathbf{B}$ are obtained solving $\mathbf{B}^\top\mathbf{B}\mathbf{V} = \sigma^2\mathbf{V}$. Hence, we need to compute the eigenvalues of the matrix:

$$\mathbf{B}^\top\mathbf{B} = \begin{pmatrix} 1 & 1 & 0 \\ -2 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ -1 & 5 \end{pmatrix} \tag{4}$$

we note that the matrix $\mathbf{B}^\top\mathbf{B}$ is symmetric. Following the process outlined above, we obtain $\sigma^2_{1,2} = \frac{7\pm\sqrt{13}}{2}$

(d) The right singular vector is obtained solving $\mathbf{B}^\top\mathbf{B}\mathbf{V} = \sigma^2\mathbf{V}$. Following the process outlined above we obtain $\mathbf{V}_1 = \left(c, -c\frac{3+\sqrt{13}}{2}\right)$ and $\mathbf{V}_2 = \left(b, b\frac{\sqrt{13}-3}{2}\right)$ where the constants c and b can be defined imposing the vectors to be normalised.
Note how also the vectors $\mathbf{V}_1 = \left(c\frac{3-\sqrt{13}}{2}, c\right)$ (which is proportional to the previous with a factor $\frac{3-\sqrt{13}}{2}$) and $\mathbf{V}_2 = \left(b\frac{\sqrt{13}+3}{2}, b\right)$ (which is proportional to the previous with a factor $\frac{\sqrt{13}+3}{2}$) are solutions.
The left singular vectors $\mathbf{U}_i$ can be computed recalling that $\mathbf{U}_i = \sigma_i^{-1}\mathbf{B}\mathbf{V}_i$, hence by using the first expression for $\mathbf{V}_i$ we get:

$$\mathbf{U}_1 = \sqrt{\frac{2}{7+\sqrt{13}}} \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} c \\ -c\frac{3+\sqrt{13}}{2} \end{pmatrix} \tag{5}$$

$$= c\sqrt{\frac{2}{7+\sqrt{13}}} \begin{pmatrix} 4+\sqrt{13} \\ -\frac{1+\sqrt{13}}{2} \\ 0 \end{pmatrix} \tag{6}$$

Furthermore

$$\mathbf{U}_2 = \sqrt{\frac{2}{7-\sqrt{13}}} \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} b \\ b\frac{\sqrt{13}-3}{2} \end{pmatrix} \tag{7}$$

$$= b\sqrt{\frac{2}{7-\sqrt{13}}} \begin{pmatrix} 4-\sqrt{13} \\ \frac{\sqrt{13}-1}{2} \\ 0 \end{pmatrix} \tag{8}$$

Instead, by using the second expression for the $\mathbf{V}_i$:

$$\mathbf{U}_1 = \sqrt{\frac{2}{7+\sqrt{13}}} \begin{pmatrix} 1 & -2 \\ 1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} c\frac{3-\sqrt{13}}{2} \\ c \end{pmatrix} \tag{9}$$

$$= c\sqrt{\frac{2}{7+\sqrt{13}}} \begin{pmatrix} -\frac{1+\sqrt{13}}{2} \\ \frac{5-\sqrt{13}}{2} \\ 0 \end{pmatrix} \tag{10}$$

© Queen Mary University of London (2022) **Continue to next page**

Furthermore

$$\mathbf{U}_2 = \sqrt{\frac{2}{7-\sqrt{13}}}\begin{pmatrix}1 & -2\\ 1 & 1\\ 0 & 0\end{pmatrix}\begin{pmatrix}b\frac{\sqrt{13}+3}{2}\\ b\end{pmatrix} \tag{11}$$

$$= b\sqrt{\frac{2}{7-\sqrt{13}}}\begin{pmatrix}\frac{\sqrt{13}-1}{2}\\ \frac{5+\sqrt{13}}{2}\\ 0\end{pmatrix} \tag{12}$$

The matrix $BB^\top$ is a 3x3 matrix with another eigenvalue equal to zero. The associate vector needs to be orthogonal to the others, hence $\mathbf{U}_3 = (0, 0, 1^\top)$

(e) By definition of Euclidian norm we can write
$E(\mathbf{x}) = \frac{1}{2}\|\mathbf{D}\mathbf{x}\|^2 = \frac{1}{2}\sum_{i=1}^m(\mathbf{D}\mathbf{x})_i^2 = \frac{1}{2}\sum_{i=1}^m\left(\sum_{j=1}^n D_{i,j}x_j\right)^2$. The gradient of the norm as function of x can be then written as:

$$(\nabla E(\mathbf{x}))_p = \frac{\partial}{\partial x_p}E(\mathbf{x}) = \frac{1}{2}\sum_{i=1}^m\frac{\partial}{\partial x_p}\left(\sum_{j=1}^n D_{i,j}x_j\right)^2 = \tag{13}$$

$$\frac{1}{2}\sum_{i=1}^m 2\left(\sum_{j=1}^n D_{i,j}x_j\right)\frac{\partial}{\partial x_p}\left(\sum_{j=1}^n D_{i,j}x_j\right)$$

$$\sum_{i=1}^m\sum_{j=1}^n D_{i,j}x_j D_{i,p} = \sum_{i=1}^m\sum_{j=1}^n D_{p,i}^\top D_{i,j}x_j = (\mathbf{D}^\top\mathbf{D}\mathbf{x})_p$$

Alternatively, we can write $E(\mathbf{x}) = \frac{1}{2}\|\mathbf{D}\mathbf{x}\|^2 = \frac{1}{2}(\mathbf{D}\mathbf{x})^\top\mathbf{D}\mathbf{x} = \frac{1}{2}\mathbf{x}^\top\mathbf{D}^\top\mathbf{D}\mathbf{x}$. It is easy to show that $\nabla(\mathbf{x}^\top\mathbf{B}\mathbf{x}) = (\mathbf{B}+\mathbf{B}^\top)\mathbf{x}$. Indeed, we have

$$(\nabla\mathbf{x}^\top\mathbf{B}\mathbf{x})_p = \frac{\partial}{\partial x_p}\sum_i x_i\sum_j B_{ij}x_j \tag{14}$$

$$= \sum_j B_{pj}x_j + \sum_i B_{ip}x_i = \sum_j B_{pj}x_j + \sum_i B_{pi}^\top x_i$$

$$= (\mathbf{B}\mathbf{x})_p + (\mathbf{B}^\top\mathbf{x})_p$$

We can make use of this observation when calculating $\nabla E(\mathbf{x}) = \frac{1}{2}\nabla(\mathbf{x}^\top\mathbf{D}^\top\mathbf{D}\mathbf{x})$. Indeed, let us define $\mathbf{B} = \mathbf{D}^\top\mathbf{D}$, hence
$\nabla E(\mathbf{x}) = \frac{1}{2}\nabla(\mathbf{x}^\top\mathbf{B}\mathbf{x}) = \frac{1}{2}(\mathbf{B}+\mathbf{B}^\top)\mathbf{x} = \frac{1}{2}\left(\mathbf{D}^\top\mathbf{D} + (\mathbf{D}^\top\mathbf{D})^\top\right)\mathbf{x}$. Finally we observe that $(\mathbf{D}^\top\mathbf{D})^\top = \mathbf{D}^\top(\mathbf{D}^\top)^\top = \mathbf{D}^\top\mathbf{D}$. Hence,
$\nabla E(\mathbf{x}) = \frac{1}{2}\left(\mathbf{D}^\top\mathbf{D} + (\mathbf{D}^\top\mathbf{D})^\top\right)\mathbf{x} = \mathbf{D}^\top\mathbf{D}\mathbf{x}$

© Queen Mary University of London (2022) **Continue to next page**

Question 2 [25 marks].

Consider the following data samples $\left(x^{(1)}, y^{(1)}\right) = (1,0)$, $\left(x^{(2)}, y^{(2)}\right) = (2,1)$, $\left(x^{(3)}, y^{(3)}\right) = (3,2)$

(a) Write down, in explicit matricial form, the normal equation assuming a simple linear model. **[5]**

(b) Determine the solution of the normal equation. **[5]**

(c) Let us now assume that you made some errors measuring the output variables $y^{(i)}$ with $i \in \{1,2,3\}$. The perturbed measurements $\mathbf{y}_\delta$ read $y_\delta^{(1)} = \epsilon$, $y_\delta^{(2)} = 1+\epsilon$ and $y_\delta^{(3)} = 2-\epsilon$. Determine the solution of the normal equation considering these perturbed samples and considering the same initial data matrix. **[5]**

(d) Compute the error between $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_\delta$ in the Euclidean norm. **[5]**

(e) Compare the error computed in the previous question (i.e., question d) with the data error $\delta := \|\mathbf{y} - \mathbf{y}_\delta\|$. **[5]**

Solution:

(a) The normal equation reads $\mathbf{X}^\top \mathbf{X}\hat{\mathbf{w}} = \mathbf{X}^\top \mathbf{y}$ where

$$\mathbf{X} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix}, \quad \mathbf{X}^\top = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix}, \quad \hat{\mathbf{w}} = \begin{pmatrix} \hat{w}_0 \\ \hat{w}_1 \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \tag{15}$$

hence:

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} \hat{w}_0 \\ \hat{w}_1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix} \tag{16}$$

(b) Performing the multiplications on the left and right hand side we have

$$\begin{pmatrix} 3 & 6 \\ 6 & 14 \end{pmatrix} \begin{pmatrix} \hat{w}_0 \\ \hat{w}_1 \end{pmatrix} = \begin{pmatrix} 3 \\ 8 \end{pmatrix} \tag{17}$$

which leads to $\hat{w}_0 = -1$ and $\hat{w}_1 = 1$

(c) Considering the errors in the outputs our problem becomes

$$\begin{pmatrix} 3 & 6 \\ 6 & 14 \end{pmatrix} \begin{pmatrix} \hat{w}_0 \\ \hat{w}_1 \end{pmatrix} = \begin{pmatrix} 3+\epsilon \\ 8 \end{pmatrix} \tag{18}$$

which yields the solutions $\hat{w}_{0,\delta} = \frac{-3+7\epsilon}{3}$ and $\hat{w}_{1,\delta} = 1-\epsilon$. Note how for $\epsilon = 0$ we recover the previous values for $\hat{\mathbf{w}}$.

© Queen Mary University of London (2022) **Continue to next page**

(d) From what derived above we have

$$\hat{\mathbf{w}} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \quad \hat{\mathbf{w}}_\delta = \begin{pmatrix} \frac{-3+7\epsilon}{3} \\ 1-\epsilon \end{pmatrix} \tag{19}$$

We can write

$$\begin{aligned} \|\hat{\mathbf{w}} - \hat{\mathbf{w}}_\delta\| &= \sqrt{\left(-1 - \frac{-3+7\epsilon}{3}\right)^2 + (1-1+\epsilon)^2} \\ &= \sqrt{\left(-1 - \frac{-3+7\epsilon}{3}\right)^2 + (1-1+\epsilon)^2} \\ &= \frac{\epsilon}{3}\sqrt{58} = \Xi \end{aligned} \tag{20}$$

(e) By computing $\|\mathbf{y} - \mathbf{y}_\delta\| = \delta$ we obtain $\delta = \epsilon\sqrt{3}$. By taking the ratio

$$\frac{\delta}{\Xi} = \frac{3\sqrt{3}}{\sqrt{58}} < 1 \tag{21}$$

which implies $\delta < \Xi$. The error in the output is amplified in the regression.

© Queen Mary University of London (2022) **Continue to next page**

Question 3 [30 marks].

(a) Consider the following probability density function of a (continuous) random variable x: $p(x|\alpha) = Ax^{-\alpha}$ for $x \geqslant 1$ where $A \in \mathbb{R}$ and $\alpha \in \mathbb{R}$. Compute the value A as function of α and discuss where it is defined. **[10]**

(b) Compute the expectation value $\mathbb{E}[x]$ and the second moment of the distribution $\mathbb{E}[x^2]$ as function of α and discuss where they are defined. **[10]**

(c) Discuss the convex properties of the function $f(x) = ax^2$ where $a \in \mathbb{R}$. **[5]**

(d) Discuss the convex properties the function $f(x) = -a\log(x)$ where $a \in \mathbb{R}$. **[5]**

Solution:

(a) Since the PDF need to be normalized and the variable is continuous we can write

$$\int_1^\infty p(x|\alpha)dx = 1 \tag{22}$$

or

$$A\int_1^\infty x^{-\alpha}dx = 1 \rightarrow A = \frac{1}{\int_1^\infty x^{-\alpha}dx} \tag{23}$$

The integral can be easily computed as $\int_1^\infty x^{-\alpha}dx = \frac{1}{1-\alpha}x|_1^\infty$.
We note that only for $\alpha > 1$ that integral is finite. In this case, we obtain $A = \alpha - 1$

(b) By definition the expectation value is

$$\begin{aligned}\mathbb{E}[x] &= \int xp(x|\alpha)dx = (\alpha-1)\int_1^\infty x^{1-\alpha}dx \\ &= (\alpha-1)\frac{1}{2-\alpha}x^{2-\alpha}|_1^\infty\end{aligned} \tag{24}$$

It can be easily seen how for any $\alpha \leqslant 2$ the expectation value is divergent. For $\alpha > 2$ we get $\mathbb{E}[x] = \frac{\alpha-1}{\alpha-2}$

(c) By definition the second moment of the distribution reads

$$\begin{aligned}\mathbb{E}[x^2] &= \int x^2p(x|\alpha)dx = (\alpha-1)\int_1^\infty x^{2-\alpha}dx \\ &= (\alpha-1)\frac{1}{3-\alpha}x^{3-\alpha}|_1^\infty\end{aligned} \tag{25}$$

It can be easily seen how for any $\alpha \leqslant 3$ the second moment is divergent. For $\alpha > 3$ we get $\mathbb{E}[x^2] = \frac{\alpha-1}{\alpha-3}$

© Queen Mary University of London (2022) **Continue to next page**

(d) We can study the convexity of the function with two approaches. In the first we don't use any information about the other properties of the function and apply directly the definition of the convexity:

$$f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda)f(y) \tag{26}$$

Let us apply this condition to our function

$$\begin{aligned} a\left(\lambda x + (1-\lambda)y\right)^2 &\leqslant \lambda a x^2 + (1-\lambda)a y^2 && (27)\\ a\left(\lambda^2 x^2 + (1-\lambda)^2 y^2 + 2\lambda(1-\lambda)xy\right) &\leqslant \lambda a x^2 + (1-\lambda)a y^2 && (28)\\ -a\lambda(1-\lambda)x^2 - a\lambda(1-\lambda)y^2 + 2a\lambda(1-\lambda)xy &\leqslant 0 && (29) \end{aligned}$$

Since $\lambda(1-\lambda) > 0$ we can write:

$$\begin{aligned} -a\lambda(1-\lambda)x^2 - a\lambda(1-\lambda)y^2 + 2a\lambda(1-\lambda)xy &\leqslant 0 && (30)\\ -ax^2 - ay^2 + 2axy &\leqslant 0 && (31)\\ -a(x-y)^2 &\leqslant 0 && (32) \end{aligned}$$

which holds only if $a \geqslant 0$. Hence the function $f(x)$ is convex for any $a \geqslant 0$.
The second approach notes that the function is differentiable at least twice. These types of functions are convex if the second derivative is always equal or larger than zero. Hence $f(x) = ax^2$, $d_x f = 2ax$ and $d_x^2 f(x) = 2a$. Hence only if $a \geqslant 0$ the function is convex

(e) Following the same argument we can compute the first and second derivates of the function $f(x) = -a\log x$ as follows $d_x f(x) = -ax^{-1}$ and $d_x^2 f(x) = ax^{-2}$. The second derivative is negative only if $a < 0$, hence the function is convex for any $a \geqslant 0$.

End of Paper.

© Queen Mary University of London (2022)