Queen Mary University of London

**Late-Summer Examination period 2023**

# 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: 1st/2nd Examiners: N. Perra, N. Otter**

 **Continue to next page**

**Question 1 [45 marks].**

(a) Consider the following matrix $\mathbf{A} = \begin{pmatrix} a & 1 \\ 1 & -1 \end{pmatrix}$ where $a \in \mathbb{R}$. Compute the eigenvalues and correspondent eigenvectors of $\mathbf{A}$. Discuss for which values of $a$ the eigenvalues are real. **[15]**

(b) Consider the following matrix $\mathbf{M} = \begin{pmatrix} 2 & 2 \\ -1 & 1 \\ 0 & 0 \end{pmatrix}$ Compute the singular values, left and right singular vectors. **[15]**

(c) Consider a matrix $\mathbf{B} \in \mathbb{R}^{n,n}$, a matrix $\mathbf{C} \in \mathbb{R}^{n,n}$, a vector $\mathbf{x} \in \mathbb{R}^{n,1}$, a vector $\mathbf{v} \in \mathbb{R}^{n,1}$ and the function $E(\mathbf{x}) = \frac{1}{3}\|\mathbf{B}(\mathbf{x}+\mathbf{v}) + \mathbf{C}(\mathbf{x}-\mathbf{v})\|^2$. Compute the gradient of the energy function i.e., $\nabla E(\mathbf{x})$ **[15]**

---

**Total marks so far: 45**

***Please remove the showmarks command before passing the exam to the checker***

---

**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:

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

Imposing the determinant to be equal zero, we obtain the characteristic equation $\lambda^2 - \lambda(a-1) - a - 1 = 0$ which yields $\lambda_{1,2} = \frac{a-1\pm\sqrt{a^2+2a+5}}{2}$. Since $a^2 + 2a + 5 > 0 \ \ \forall a \in \mathbb{R}$ the solutions are always real. This can be seen also noting that $\mathbf{A} = \mathbf{A}^\top$. The eigenvectors can be derived imposing $\mathbf{A}\mathbf{x}_i = \lambda_i\mathbf{x}_i$:

$$\begin{pmatrix} a & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x_1^{(1)} \\ x_1^{(2)} \\ x_1^{(3)} \end{pmatrix} = \lambda_1 \begin{pmatrix} x_1^{(1)} \\ x_1^{(2)} \\ x_1^{(3)} \end{pmatrix} \tag{2}$$

and

$$\begin{pmatrix} a & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x_2^{(1)} \\ x_2^{(2)} \\ x_2^{(3)} \end{pmatrix} = \lambda_2 \begin{pmatrix} x_2^{(1)} \\ x_2^{(2)} \\ x_2^{(3)} \end{pmatrix} \tag{3}$$

From the first set of equations we obtain $\mathbf{x}_1 = (c(1+\lambda_1), c)$.

For the second set of equations we obtain $\mathbf{x}_2 = (c(1+\lambda_2), c)$.

(b) The singular values can be obtained by computing the eingevalues of $\mathbf{M}^\top\mathbf{M}$. This leads to:

$$\mathbf{M}^\top\mathbf{M} - \lambda\mathbb{I} = \begin{pmatrix} 5-\lambda & 3 \\ 3 & 5-\lambda \end{pmatrix} \tag{4}$$

Imposing the determinant to be equal zero, we obtain the characteristic equation $\lambda^2 - 10\lambda + 16 = 0$ which yields $\lambda_1 = 8$ and $\lambda_2 = 2$. We know that the singular values are then $\sigma_1 = 2\sqrt{2}$ and $\sigma_2 = \sqrt{2}$. The matrix of singular values can be then written as:

$$\mathbf{\Sigma} = \begin{pmatrix} 2\sqrt{2} & 0 \\ 0 & \sqrt{2} \\ 0 & 0 \end{pmatrix} \tag{5}$$

We can now obtain the right singular vectors by imposing $\mathbf{M}^\top\mathbf{M}\mathbf{V} = \sigma^2\mathbf{V}$. By considering $\sigma_1^2$ one gets $\mathbf{V_1} = (c, c)^\top$ which imposing the normalization leads to $\mathbf{V_1} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$. By considering $\sigma_2^2$ one gets $\mathbf{V_2} = (-c, c)^\top$ which imposing the normalization leads to $\mathbf{V_2} = \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$. Hence the matrix $\mathbf{V}$ reads:

$$\mathbf{V} = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \tag{6}$$

 

We can easily obtain the left singular vectors by noting how $\mathbf{U}_i = \sigma_i^{-1}\mathbf{M}\mathbf{V}_i$. Hence:

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

Also,

$$\mathbf{U}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 2 & 2 \\ -1 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \tag{8}$$

Finally, the third left singular vector is associated to the zero eigenvalue. Since it needs to be orthogonal to the other we have $\mathbf{U}_3 = (0, 0, 1)^\top$. By putting all together we obtain

$$\mathbf{U} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{9}$$

(c) By definition of Euclidian norm we can write

$$E(\mathbf{x}) = \frac{1}{3}\|\mathbf{B}(\mathbf{x}+\mathbf{v}) + \mathbf{C}(\mathbf{x}-\mathbf{v})\|^2 = \frac{1}{3}\sum_i\left(\sum_j B_{ij}(x_j+v_j) + \sum_j C_{ij}(x_j-v_j)\right)^2 \tag{10}$$

The component $p$ of the gradient is then

$$\begin{aligned}
\nabla E(\mathbf{x})_p &= \frac{1}{3}\frac{\partial}{\partial x_p}\sum_i\left(\sum_j B_{ij}(x_j+v_j) + \sum_j C_{ij}(x_j-v_j)\right)^2 \\
&= \frac{2}{3}\sum_i\left(\sum_j B_{ij}(x_j+v_j) + \sum_j C_{ij}(x_j-v_j)\right)(B_{ip}+C_{ip}) \\
&= \frac{2}{3}\sum_{ij}\left[B^\top_{pi}B_{ij}(x_j+v_j) + C^\top_{pi}B_{ij}(x_j+v_j) + B^\top_{pi}C_{ij}(x_j-v_j) + C^\top_{pi}C_{ij}(x_j-v_j)\right] \\
&= \frac{2}{3}\left[(\mathbf{B}^\top\mathbf{B} + \mathbf{C}^\top\mathbf{B})(\mathbf{x}+\mathbf{v})_p + (\mathbf{B}^\top\mathbf{C} + \mathbf{C}^\top\mathbf{C})(\mathbf{x}-\mathbf{v})_p\right] \qquad (11)
\end{aligned}$$

Hence, the gradient is

$$\nabla E(\mathbf{x}) = \frac{2}{3}\left[(\mathbf{B}^\top\mathbf{B} + \mathbf{C}^\top\mathbf{B})(\mathbf{x}+\mathbf{v}) + (\mathbf{B}^\top\mathbf{C} + \mathbf{C}^\top\mathbf{C})(\mathbf{x}-\mathbf{v})\right] \tag{12}$$

**Question 2 [25 marks].**

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

(a) Write down, in explicit matricial form, the normal equation assuming a ridge regression. **[5]**

(b) Determine the solution of the normal equation assuming a ridge regression. **[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 $y_\delta^{(1)} = 2 + \epsilon$, $y_\delta^{(2)} = 1$ and $y_\delta^{(3)} = 3 - \epsilon$. 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) Compute the data error $\delta := \|\mathbf{y} - \mathbf{y}_\delta\|$. **[5]**

---

**Total marks so far: 70**

***Please remove the showmarks command before passing the exam to the checker***

---

 

**Solution**:

(a) The normal equation for the ridge regression reads $(\mathbf{X}^\top\mathbf{X} + \alpha\mathbf{I})\hat{\mathbf{w}} = \mathbf{X}^\top\mathbf{y}$ where

$$\mathbf{X} = \begin{pmatrix}1 & 1\\ 1 & 0\\ 1 & 2\end{pmatrix},\ \mathbf{X}^\top = \begin{pmatrix}1 & 1 & 1\\ 1 & 0 & 2\end{pmatrix},\ \hat{\mathbf{w}} = \begin{pmatrix}\hat{w}_0\\ \hat{w}_1\end{pmatrix},\ \mathbf{Y} = \begin{pmatrix}2\\ 1\\ 3\end{pmatrix} \tag{13}$$

hence:

$$\left(\begin{pmatrix}1 & 1 & 1\\ 1 & 0 & 2\end{pmatrix}\begin{pmatrix}1 & 1\\ 1 & 0\\ 1 & 2\end{pmatrix} + \alpha\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right)\begin{pmatrix}\hat{w}_0\\ \hat{w}_1\end{pmatrix} = \begin{pmatrix}1 & 1 & 1\\ 1 & 0 & 2\end{pmatrix}\begin{pmatrix}2\\ 1\\ 3\end{pmatrix} \tag{14}$$

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

$$\begin{pmatrix}3+\alpha & 3\\ 3 & 5+\alpha\end{pmatrix}\begin{pmatrix}\hat{w}_0\\ \hat{w}_1\end{pmatrix} = \begin{pmatrix}6\\ 8\end{pmatrix} \tag{15}$$

which leads to $\hat{w}_0 = \frac{6(\alpha+1)}{\alpha^2+8\alpha+6}$ and $\hat{w}_1 = \frac{8\alpha+6}{\alpha^2+8\alpha+6}$ for $\alpha^2 + 8\alpha + 6 \neq 0$ i.e., $\alpha \neq \sqrt{10} - 4$ and $\alpha \neq -\sqrt{10} - 4$

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

$$\begin{pmatrix}3+\alpha & 3\\ 3 & 5+\alpha\end{pmatrix}\begin{pmatrix}\hat{w}_0\\ \hat{w}_1\end{pmatrix} = \begin{pmatrix}6\\ 8-\epsilon\end{pmatrix} \tag{16}$$

which yields the solutions $\hat{w}_{0,\delta} = \frac{3(2\alpha+\epsilon+2)}{\alpha^2+8\alpha+6}$ and $\hat{w}_{1,\delta} = \frac{8\alpha+6-\epsilon(\alpha+3)}{\alpha^2+8\alpha+6}$ for $\alpha^2 + 8\alpha + 6 \neq 0$ i.e., $\alpha \neq \sqrt{10} - 4$ and $\alpha \neq -\sqrt{10} - 4$. Note how for $\epsilon = 0$ we recover the previous values for $\hat{\mathbf{w}}$.

(d) From what derived above we have

$$\hat{\mathbf{w}} = \frac{1}{\alpha^2 + 8\alpha + 6}\begin{pmatrix}6\alpha + 6\\ 8\alpha + 6\end{pmatrix},\ \hat{\mathbf{w}}_\delta = \frac{1}{\alpha^2 + 8\alpha + 6}\begin{pmatrix}6\alpha + 6 + 3\epsilon\\ 8\alpha + 6 - \epsilon(\alpha + 3)\end{pmatrix} \tag{17}$$

We can write

$$\begin{aligned}\|\hat{\mathbf{w}} - \hat{\mathbf{w}}_\delta\| &= \frac{1}{\alpha^2 + 8\alpha + 6}\sqrt{(3\epsilon)^2 + (\epsilon(3+\alpha))^2}\\ &= \frac{\epsilon}{\alpha^2 + 8\alpha + 6}\sqrt{9 + (3+\alpha)^2}\end{aligned} \tag{18}$$

(e) By computing $\|\mathbf{y} - \mathbf{y}_\delta\| = \delta$ we obtain $\delta = \epsilon\sqrt{2}$.

 

**Question 3 [30 marks].**

Consider the following data samples $\left(x^{(1)}, y^{(1)}\right) = (-1, -1)$, $\left(x^{(2)}, y^{(2)}\right) = (0, 1)$, $\left(x^{(3)}, y^{(3)}\right) = (1, 2)$, $\left(x^{(4)}, y^{(4)}\right) = (2, 3)$

(a) Compute the MSE for a 1-parameter model by hand:

$$MSE(w^{(0)}) = \frac{1}{2s}\sum_{i=1}^{s} |y^{(i)} - w^{(0)}|^2$$

considering $w^{(0)} \in \{1, 2, 3\}$. Which of the three values minimizes the MSE? **[10]**

(b) a new data sample is added $\left(x^{(5)}, y^{(5)}\right) = (3, 1)$. Evaluate new error measure and corresponding minimiser considering the same three possible values for $w^{(0)}$ **[5]**

(c) Repeat the same exercise for what is known as the Mean Absolute Error (MAE)

$$MSE(w^{(0)}) = \frac{1}{2s}\sum_{i=1}^{s} |y^{(i)} - w^{(0)}|$$

What do you observe, in particular with regards to the outlier $y^{(5)}$? **[10]**

(d) Discuss the convex properties of the function $f(x) = a\sin(x)$ considering $0 \leqslant x \leqslant 2\pi$ and $a \in \mathbb{R}$. **[5]**

---

**Total marks so far: 100**

***Please remove the showmarks command before passing the exam to the checker***

---

**Solution**:

(a) for $w^{(0)} = 1$ the MSE reads:

$$MSE(w^{(0)} = 1) = \frac{1}{8}\left[(-1-1)^2 + (1-1)^2 + (2-1)^2 + (3-1)^2\right] = \frac{9}{8}$$

for $w^{(0)} = 2$ the MSE reads:

$$MSE(w^{(0)} = 2) = \frac{1}{8}\left[(-1-2)^2 + (1-2)^2 + (2-2)^2 + (3-2)^2\right] = \frac{11}{8}$$

for $w^{(0)} = 2$ the MSE reads:

$$MSE(w^{(0)} = 3) = \frac{1}{8}\left[(-1-3)^2 + (1-3)^2 + (2-3)^2 + (3-3)^2\right] = \frac{21}{8}$$

Hence the best model is obtained for $w^{(0)} = 1$.

(b) with the new data point for $w^{(0)} = 1$ the MSE reads:

$$MSE(w^{(0)} = 1) = \frac{1}{10}\left[(-1-1)^2 + (1-1)^2 + (2-1)^2 + (3-1)^2 + (1-1)^2\right] = \frac{9}{10}$$

for $w^{(0)} = 2$ the MSE reads:

$$MSE(w^{(0)} = 2) = \frac{1}{10}\left[(-1-2)^2 + (1-2)^2 + (2-2)^2 + (3-2)^2 + (1-2)^2\right] = \frac{12}{10}$$

for $w^{(0)} = 2$ the MSE reads:

$$MSE(w^{(0)} = 3) = \frac{1}{10}\left[(-1-3)^2 + (1-3)^2 + (2-3)^2 + (3-3)^2 + (1-3)^2\right] = \frac{25}{10}$$

the addition of a new data point leads to the same best model.

(c) for $w^{(0)} = 1$ the MAE reads:

$$MSE(w^{(0)} = 1) = \frac{1}{8}\left[|-1-1| + |1-1| + |2-1| + |3-1|\right] = \frac{5}{8}$$

for $w^{(0)} = 2$ the MSE reads:

$$MSE(w^{(0)} = 2) = \frac{1}{8}\left[|-1-2| + |1-2| + |2-2| + |3-2|\right] = \frac{5}{8}$$

for $w^{(0)} = 2$ the MSE reads:

$$MSE(w^{(0)} = 3) = \frac{1}{8}\left[|-1-3| + |1-3| + |2-3| + |3-3|\right] = \frac{7}{8}$$

Hence the best model is obtained for $w^{(0)} = 1$ and $w^{(0)} = 2$. With the new data point for $w^{(0)} = 1$ the MAE reads:

$$MSE(w^{(0)} = 1) = \frac{1}{10}\left[|-1-1| + |1-1| + |2-1| + |3-1| + |1-1|\right] = \frac{5}{10}$$

 **Continue to next page**

for $w^{(0)} = 2$ the MSE reads:

$$MSE(w^{(0)} = 2) = \frac{1}{10}\left[|-1-2| + |1-2| + |2-2| + |3-2| + |1-2|\right] = \frac{6}{10}$$

for $w^{(0)} = 2$ the MSE reads:

$$MSE(w^{(0)} = 3) = \frac{1}{10}\left[|-1-3| + |1-3| + |2-3| + |3-3| + |1-3|\right] = \frac{9}{10}$$

the addition of a new data point allows to differentiate between the first two models.

(d) We can study the convexity of the function noting that it is differentiable at least twice. These types of functions are convex if the second derivative is always equal or larger than zero. Since $f(x) = a\sin(x)$ we have $d_x f(x) = a\cos(x)$ and $d_x^2 f(x) = -a\sin(x)$. The second derivative is positive or equal to zero, hence the function if convex, if $a > 0$ and $\pi \leqslant x \leqslant 2\pi$ or if $a < 0$ and $0 \leqslant x \leqslant \pi$ or if $a = 0$ and $0 \leqslant x \leqslant 2\pi$.

---

**End of Paper.**