Late-Summer Examination period 2020

MTH786P: Machine Learning with Python

You should attempt ALL questions. Marks available are shown next to the questions.

In completing this assessment, you may use books, notes, and the internet. You may use calculators and computers, but you should show your working for any calculations you do. You must not seek or obtain help from anyone else.

At the start of your work, please **copy out and sign** the following declaration:

> I declare that my submission is entirely my own, and I have not sought or obtained help from anyone else.

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

You have **24 hours** in which to complete and submit this assessment. When you have finished your work:

- scan your work, convert it to a **single PDF file** and upload this using the upload tool on the QMplus page for the module;
- 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.

You are not expected to spend a long time working on this assessment. We expect you to spend about **3 hours** to complete the assessment, plus the time taken to scan and upload your work. Please try to upload your work well before the end of the assessment period, in case you experience computer problems. **Only one attempt is allowed – once you have submitted your work, it is final**.

Examiners: M. Benning

The terms MSE and MAE are abbreviations for Mean Squared Error and Mean Absolute Error, respectively. The notation $n!$ denotes the factorial of the argument n, i.e. $n! = 1 \times 2 \times 3 \times \cdots \times (n-1) \times n$. The notation $\overline{\mathbb{R}}$ is short-hand notation for $\mathbb{R} \cup \{\infty\}$.

Question 1 [50 marks].

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

$$\mathrm{MSE}(w_0) = \frac{1}{2s}\sum_{i=1}^{s} |y_i - w_0|^2 \,.$$

Fill in the missing entries of the following table:

	$w_0 = -5$	$w_0 = -3$	$w_0 = -1$	$w_0 = 0$	$w_0 = 1$	$w_0 = 3$	$w_0 = 5$
y_1 = -9 y_2 = -7 $y_3 = 0$ $y_4 = 5$ $y_5 = 7$							
$2s\,\mathrm{MSE}(w_0)$							
$y_6 = \min(7(d+1), 25)$							
$2s\,\mathrm{MSE}(w_0)$							

Here d denotes the last digit of your student ID. **[7]**

(b) Repeat the same exercise with the MAE, i.e.

$$\mathrm{MAE}(w_0) = \frac{1}{s}\sum_{i=1}^{s} |y_i - w_0| \,.$$

Fill in the missing entries of the following table:

	$w_0 = -5$	$w_0 = -3$	$w_0 = -1$	$w_0 = 0$	$w_0 = 1$	$w_0 = 3$	$w_0 = 5$
y_1 = -9 y_2 = -7 $y_3 = 0$ $y_4 = 5$ $y_5 = 7$							
$\mathrm{MAE}(w_0)s$							
$y_6 = \min(7(d+1), 25)$							
$\mathrm{MAE}(w_0)s$							

As in the previous exercise, d denotes the last digit of your student ID. What do you observe, in particular with regards to the outlier y_6? **[7]**

(c) Compute the gradient ∇L of the cost function $L : \mathbb{R}^3 \to \mathbb{R}$ defined as

$$L(x, y, z) = x^2 + 2y^2 + 3z^2 + 2x(y + z)\,.$$

[5]

(d) Derive the corresponding gradient descent update formula for L as defined in Question 1 (c). **[5]**

(e) Compute the Hessian $H_L(x, y, z)$ of L. **[5]**

(f) Verify that the Hessian is positive semi-definite, i.e.

$$\begin{pmatrix} u & v & w \end{pmatrix} H_L(x, y, z) \begin{pmatrix} u \\ v \\ w \end{pmatrix} \geq 0$$

holds true for all $u, v, w \in \mathbb{R}$ and all $x, y, z \in \mathbb{R}$. **[5]**

(g) What are the implications of Question 1 (e) for the function L? **[4]**

(h) Compute the gradient of the function $D(x) := \frac{1}{2}\langle Q(x - y), x - y\rangle$, where $Q \in \mathbb{R}^{n\times n}$ is a matrix, and $y \in \mathbb{R}^n$ a vector. How does the gradient simplify if Q is symmetric? **[4]**

(i) Compute the Hessian of D as defined in Question 1 (h). **[4]**

(j) Is the function D defined in Question 1 (h) convex if we choose $Q = H_L$, where H_L is the Hessian from Question 1 (e)? Justify your answer. **[4]**

Question 2 [30 marks].

(a) Verify that the function $f : \mathbb{R} \to \overline{\mathbb{R}}$, with

$$f(x) := \begin{cases} 0 & x \in [-d-1, d+1] \\ +\infty & x \notin [-d-1, d+1] \end{cases} \tag{1}$$

where d is the sixth digit of your student ID, is convex. **[5]**

(b) Derive a closed-form solution of the proximal mapping $\text{prox}_f : \mathbb{R} \to [-d-1, d+1]$ for the function defined in (1).
Hint: use a case-by-case analysis to derive the proximal mapping. **[5]**

(c) Compute the proximal map for a function $g : \mathbb{R}^n \to \mathbb{R}^n$ defined as $g(x) := f(x) + \alpha\|x\|^2 + \langle x, z\rangle$, for fixed $z \in \mathbb{R}^n$ and $\alpha > 0$. **[5]**

(d) Show that the function $f : \mathbb{R}^3 \to \mathbb{R}$ defined as

$$f(\underbrace{w_0, w_1, w_2}_{=:w}) := w_0 + w_1 x + w_2 x^2$$

is linear in its argument, for arbitrary but fixed $x \in \mathbb{R}$. **[5]**

(e) Verify that f as defined in Question 2 (d) is both convex and concave, i.e. $f(\lambda w + (1-\lambda)v) = \lambda f(w) + (1-\lambda) f(v)$, for all $w, v \in \mathbb{R}^3$ and $\lambda \in [0, 1]$. **[5]**

(f) Show that the MSE function

$$\text{MSE}(w_0, w_1, w_2) = \frac{1}{2}\left|w_0 + w_1 x + w_2 x^2 - y\right|^2$$

is convex. **Hint**: make use of the previous exercise and other properties of convex functions that you know from the lecture notes and coursework. **[5]**

Question 3 [20 marks].

(a) What is the maximum likelihood estimator (MLE)? Suppose we are given s samples $\{k_i\}_{i=1}^s$ of a Poisson random variable but do not know the parameter λ of the corresponding probability mass function

$$p(k_1, \ldots, k_s \mid \lambda) := \prod_{i=1}^{s} \frac{\lambda^{k_i} \exp(-\lambda)}{k_i!} \,. \tag{2}$$

Estimate λ by computing the MLE for (2). **[5]**

(b) Derive the MLE of a linear model assuming that the data $\{y_i\}_{i=1}^s$ are i.i.d. samples of a gamma distributed random variable with probability mass function

$$p(y_i|x_i, w) = \frac{y_i^d \langle x_i, w\rangle^{d-1}}{(d-1)!} \exp(-y_i\langle x_i, w\rangle) \,,$$

where d is the maximum between the seventh digit of your student ID and two.

- Derive $p(y|X, w)$ for $y = \begin{pmatrix} y_1 \\ \vdots \\ y_s \end{pmatrix} \in \mathbb{R}^s$, $x_i = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$, $X = \begin{pmatrix} x_1^T \\ \vdots \\ x_s^T \end{pmatrix} \in \mathbb{R}^{s\times n}$.
- Compute the negative log-likelihood.
- Compute the gradient of the negative log-likelihood.

[15]

End of Paper.