Queen Mary
University of London

Main Examination period 2019

MTH786P: Machine Learning with Python: mock exam

Duration: 1 hours

Apart from this page, you are not permitted to read the contents of this question paper until instructed to do so by an invigilator.

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

Only non-programmable calculators that have been approved from the college list of non-programmable calculators are permitted in this examination. Please state on your answer book the name and type of machine used.

Complete all rough work in the answer book and cross through any work that is not to be assessed.

Possession of unauthorised material at any time when under examination conditions is an assessment offence and can lead to expulsion from QMUL. Check now to ensure you do not have any unauthorised notes, mobile phones, smartwatches or unauthorised electronic devices on your person. If you do, raise your hand and give them to an invigilator immediately.

It is also an offence to have any writing of any kind on your person, including on your body. If you are found to have hidden unauthorised material elsewhere, including toilets and cloakrooms, it will be treated as being found in your possession. Unauthorised material found on your mobile phone or other electronic device will be considered the same as being in possession of paper notes. A mobile phone that causes a disruption in the exam is also an assessment offence.

Exam papers must not be removed from the examination room.

Examiners: M. Benning

The term MSE is an abbreviation for Mean Squared Error. Note that for the multiple choice question there will only be **one** correct answer. The symbol $\mathbb{N}_0$ denotes all natural numbers and zero, and $a!$ is the factorial of the number a, i.e. $a! = 1 \cdot 2 \cdot \ldots \cdot a$.

Question 1. [36 marks]

(a) If a function L is convex, then gradient descent is guaranteed to converge...

A. ...more quickly than the Newton-Raphson method.

B. ...for any chosen step-size.

C. ...if the step-size τ is chosen such that the function $\frac{1}{2\tau}\|x\|^2 - L(x)$ is convex for all arguments x.

D. ...for any decreasing step-size.

[4]

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

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

Fill in the missing entries of the following table:

	$w_0 = -7$	$w_0 = -6$	$w_0 = -5$	$w_0 = -4$	$w_0 = -3$	$w_0 = -2$	$w_0 = -1$
y_1 = -4 y_2 = -3 y_3 = -2 y_4 = -1							
2 $\text{MSE}(w_0)s$							
y_5 = -20							
2 $\text{MSE}(w_0)s$							

Some help: $19^2 = 361, 18^2 = 324, 17^2 = 289, 16^2 = 256, 15^2 = 225, 14^2 = 196, 13^2 = 169$.

[5]

(c) Compute the gradient ∇L of the cost function $L : \mathbb{R} \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ with

$$L(w, v) = \frac{1}{2}\frac{|xw - y|^2}{v + \varepsilon},$$

for a constant $\varepsilon > 0$. **[5]**

(d) A function $L : \mathbb{R} \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ is convex if and only if its Hessian is positive semi-definite, i.e.

$$\begin{pmatrix} z_1 & z_2 \end{pmatrix} H_L(w, v) \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \geq 0$$

for all $z_1, w \in \mathbb{R}$ and $z_2, v \in \mathbb{R}_{\geq 0}$. Show that the function L defined in (c) is convex. **[5]**

(e) Recall the definition of Tikhonov regularisation / ridge regression. **[3]**

(f) Derive the maximum likelihood estimator of a linear model where the conditional probability for each data point is modelled as a Poisson distribution, i.e.

$$p(y_i|x_i, w) = \frac{\langle x_i, w\rangle^{y_i}}{y_i!} e^{-\langle x_i, w\rangle} .$$

- Derive the likelihood $p(y|X, w)$ for data $y = \begin{pmatrix} y_1 \\ \vdots \\ y_s \end{pmatrix} \in \mathbb{N}_0^s$, $x_i = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ and $X = \begin{pmatrix} x_1^T \\ \vdots \\ x_s^T \end{pmatrix} \in \mathbb{R}^{s\times n}$. Assume that the data samples are i.i.d.
- Compute the negative log-likelihood.

[6]

(g) Derive the proximal mapping prox_f for the (one-dimensional) function $f(x) = |x|$. **[3]**

(h) Verify the following matrix identity:

$$(X^T X + \alpha I)^{-1} X^T y = X^T (XX^T + \alpha I)^{-1} y ,$$

for $X \in \mathbb{R}^{s\times n}$, $y \in \mathbb{R}^s$ and $\alpha > 0$.
Hint: Use the Woodbury matrix identity
$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$. **[5]**

End of Paper.