A36902

Calculators may be used in this examination provided they are not capable of being used to store alphabetical information other than hexadecimal numbers

UNIVERSITY OF BIRMINGHAM

School of Computer Science

Fourth Year Undergraduate/Postgraduate

06-32250

LM Mathematical Foundations of Artificial Intelligence (AI) and Machine Learning

Main Summer Examinations 2025

Time allowed: 2 hours

[Answer all questions]

Note

Answer ALL questions.

Question 1

Consider the set $\mathcal{L}$ of regression models represented as parametrised functions over real numbers of the form:

$$f(x) = a \cdot \sin(x) + b \cdot \cos(2x) + c \cdot x^2, \quad x \in \mathbb{R},$$

where $a, b, c \in \mathbb{R}$ are real learnable parameters.

The regression models can be combined by addition and scaled with a scalar as follows: On $\mathcal{L}$ we define a binary (point wise) addition operation $\oplus$: for any $f, g \in \mathcal{L}$, the function $h = f \oplus g$ reads,

$$h(x) = (f \oplus g)(x) = f(x) + g(x).$$

We also define a scalar multiplication operation $\odot$: For any $f \in \mathcal{L}$ and $\alpha \in \mathbb{R}$, the function $q = \alpha \odot f$ reads,

$$q(x) = (\alpha \odot f)(x) = \alpha \cdot f(x).$$

Here $+$ and $\cdot$ denote the usual addition and multiplication on real numbers, respectively.

(a) We would like to show that $(\mathcal{L}, \oplus)$ is a group. As part of the proof we need to verify the existence of the neutral element of $(\mathcal{L}, \oplus)$. Recall that the neutral element $e \in \mathcal{L}$ satisfies $e \oplus f = f \oplus e = f$, for any $f \in \mathcal{L}$. Write down the neutral element of $(\mathcal{L}, \oplus)$ and argue that it indeed satisfies the required properties. **[5 marks]**

(b) Assume we have shown that $(\mathcal{L}, \oplus)$ is a group. Is it an Abelian group, i.e. is the operation $\oplus$ on $\mathcal{L}$ commutative? Justify your answer. **[5 marks]**

(c) Consider a function $f \in \mathcal{L}$. What is its inverse element in the group $(\mathcal{L}, \oplus)$? Recall that the inverse element of f is $f^{-1} \in \mathcal{L}$ such that $f \oplus f^{-1} = f^{-1} \oplus f = e$, where $e \in \mathcal{L}$ is the neutral element. Justify your answer. **[5 marks]**

(d) It can be shown that $(\mathcal{L}, \oplus, \odot)$ is a vector space.

(i) Suggest a basis for $(\mathcal{L}, \oplus, \odot)$. Justify your answer. **[10 marks]**

(ii) What is the dimensionality of the vector space $(\mathcal{L}, \oplus, \odot)$? **[5 marks]**

(e) We now endow $\mathcal{L}$ with an inner product: For any two functions

$$f_1(x) = a_1 \cdot \sin(x) + b_1 \cdot \cos(2x) + c_1 \cdot x^2$$

and

$$f_2(x) = a_2 \cdot \sin(x) + b_2 \cdot \cos(2x) + c_2 \cdot x^2,$$

we have

$$\langle f_1, f_2 \rangle = a_1 \cdot a_2 + 2 \cdot b_1 \cdot b_2 + 5 \cdot c_1 \cdot c_2$$

Consider two subspaces of $\mathcal{L}$:

$$\mathcal{L}_1 = \{f \mid f(x) = a \cdot \sin(x), a \in \mathbb{R}\}$$

and

$$\mathcal{L}_2 = \{g \mid g(x) = b \cdot \cos(2x), b \in \mathbb{R}\}$$

Show that, under this inner product, $\mathcal{L}_1$ and $\mathcal{L}_2$ are orthogonal. In other words, show that any pair of non-zero functions $f \in \mathcal{L}_1$ and $g \in \mathcal{L}_2$ is orthogonal, i.e. $\langle f, g \rangle = 0$.

[6 marks]

Question 2

(a) Let $Z = (Z^1, Z^2) \in \mathbb{R}^2$ be a 2-dimensional random vector distributed according to the distribution

$$q_\phi(Z) = q_{\phi_1}(Z^1) q_{\phi_2}(Z^2),$$

where $\phi = (\phi_1, \phi_2)$ denotes the variational parameters and $q_{\phi_i}(Z^i) = \mathcal{N}(0, \phi_i^2)$, for $i = 1, 2$, is the zero-mean Gaussian distribution with standard deviation $\phi_i > 0$. Prove that the entropy of Z evaluates as

$$H(Z) = \log(2\pi e \phi_1 \phi_2).$$

Hint: The entropy of a zero-mean, σ^2-variance Gaussian random variable X equals $H(X) = \frac{1}{2}\log(2\pi e \sigma^2)$.

[8 marks]

(b) Consider the following posterior distribution $p(Z|X)$ of the latent vector Z conditioned on the input data $X \in \mathbb{R}^2$:

$$p(Z|X) = p_1(Z^1|X) p_2(Z^2|X), \text{ where } p_i(Z^i|X) = \mathcal{N}(w^\top X, \sigma^2), \text{for } i = 1, 2,$$

with $w \in \mathbb{R}^2$ and $\sigma > 0$. Assume the variational distribution $q_\phi(Z)$ from part (a). Provide an equation for the KL divergence,

$$J(\phi) = D(q_\phi(Z) \| p(Z|X)),$$

between $q_\phi(Z)$ and $p(Z|X)$ explicitly in terms of the parameters $\phi = (\phi_1, \phi_2)$. You may ignore additive terms that do not depend on ϕ.
Hint: The KL divergence $D(p(Y)\|q(Y))$ between distributions $p(Y)$ and $q(Y)$ is defined as

$$D(p(Y)\|q(Y)) = \mathbb{E}_{p(Y)}\Big[\log \frac{p(Y)}{q(Y)}\Big].$$

[16 marks]

(c) Consider a classifier which for a given input $X = x$, outputs the class label $Y = 0$ with probability p, for some $p \in [0, 1]$, and outputs label $Y = 1$ with probability $1 - p$. For what value of p is the entropy $H(Y|X = x)$ maximized? Explain. Also report the maximum entropy attained. **[8 marks]**

Question 3

(a) Assume that the input data (X, Y), $X \in \mathbb{R}^2$ and $Y \in \mathbb{R}$, is generated according to the distribution $P_{X,Y} = P_X P_{Y|X}$ where $P_{Y|X} = \mathcal{N}(b^\top X, \nu^2)$ is the Gaussian distribution with mean $b^\top X$ and variance ν^2. Considering mean-squared error as the loss function, i.e., $\ell(Y, f(X)) = (Y - f(X))^2$, state the Bayes optimal regressor.
Hint: A regressor $f : \mathbb{R}^2 \to \mathbb{R}$ is Bayes-optimal if it minimizes the population/test loss. **[8 marks]**

(b) Consider the hypothesis class of binary classifiers

$$\mathcal{F} = \{f_t : t \in [1, 100], t \text{ is an integer}\},$$

where each function $f_t(x)$ labels the input $x \in \mathbb{R}$ as

$$f_t(x) = \begin{cases} 0 & \text{if } x < t \\ 1 & \text{else} \end{cases}. \tag{1}$$

Assuming realisable setting and $0 - 1$ loss function, state an upper bound on the sample complexity required to (ϵ, δ)-probably approximately correct (PAC) learn this hypothesis class. **[8 marks]**

(c) Let us now modify the hypothesis class in part(b) as

$$\mathcal{F} = \{f_t : t \in \mathbb{R}\},$$

where $f_t(x)$ for $x \in \mathbb{R}$ is defined as in (1). Assume realisability, i.e., there exists a true classifier $f_{t^*} \in \mathcal{F}$ defined by $t^* \in \mathbb{R}$. Consider a learning algorithm that takes as input an n-sample training data and outputs a classifier $f_{\hat{t}} \in \mathcal{F}$, where $\hat{t}$ corresponds to the smallest x in the dataset with label 1. Show that for $n \geq \frac{1}{\epsilon} \log \frac{1}{\delta}$, this learning algorithm is (ϵ, δ)-probably approximately correct.
Hint: Recall that a learning algorithm is said to be probably approximately correct if when trained on $n \geq n(\epsilon, \delta)$ samples, it returns a hypothesis whose population loss/test loss does not exceed ϵ, with probability at least $1 - \delta$ (with respect to random draws of datasets) for some $\epsilon, \delta > 0$. **[16 marks]**

This page intentionally left blank.

A36902 LM Mathematical Foundations of Artificial Intelligence (AI) and Machine Learning Non-alpha only

Do not complete the attendance slip, fill in the front of the answer book or turn over the question paper until you are told to do so

Important Reminders

- Coats/outwear should be placed in the designated area.
- Unauthorised materials (e.g. notes or Tippex) must be placed in the designated area.
- Check that you do not have any unauthorised materials with you (e.g. in your pockets, pencil case).
- Mobile phones and smart watches **must** be switched off and placed in the designated area or under your desk. They must not be left on your person or in your pockets.
- You are not permitted to use a mobile phone as a clock. If you have difficulty seeing a clock, please alert an Invigilator.
- You are not permitted to have writing on your hand, arm or other body part.
- Check that you do not have writing on your hand, arm or other body part – if you do, you must inform an Invigilator immediately
- Alert an Invigilator immediately if you find any unauthorised item upon you during the examination.

Any students found with non-permitted items upon their person during the examination, or who fail to comply with Examination rules may be subject to Student Conduct procedures.