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

Mathematical Foundations of Artificial Intelligence and Machine Learning

Main Summer Examinations 2024

Time allowed: 2 hours

[Answer all questions]

Turn Over

Note

Answer ALL questions. Each question will be marked out of 20. The paper will be marked out of 60, which will be rescaled to a mark out of 100.

Question 1

Consider a discriminative binary classifier operating in the four-dimensional vector space $\mathbb{R}^4$. During the training it learnt the following metric on $\mathbb{R}^4$:

$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^4; \quad d(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^\top \mathbf{A} (\mathbf{x} - \mathbf{y})},$$

where $\mathbf{A} = diag(1, 2, 3, 1)$ is a diagonal 4×4 matrix with diagonal terms $A_{1,1} = 1$, $A_{2,2} = 2$, $A_{3,3} = 3$ and $A_{4,4} = 1$; all the off-diagonal terms are zero.

(a) Explain how this metric differs from the usual isotropic distance formulation on $\mathbb{R}^4$:

$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^4; \quad d_{iso}(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{j=1}^{4} (x_j - y_j)^2}.$$

Provide examples of two points $\mathbf{x}, \mathbf{y} \in \mathbb{R}^4$ that appear more distant under the metric d than under d_{iso}. **[5 marks]**

(b) There is a linear subspace $\mathcal{L}$ of $\mathbb{R}^4$ where the two distances coincide, i.e.

$$\forall \mathbf{x}, \mathbf{y} \in \mathcal{L}; \quad d(\mathbf{x}, \mathbf{y}) = d_{iso}(\mathbf{x}, \mathbf{y}).$$

What is the dimensionality of $\mathcal{L}$? Specify a basis for $\mathcal{L}$. **[8 marks]**

(c) Paul suggested a different metric:

$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^4; \quad d_{paul}(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^\top \mathbf{B} (\mathbf{x} - \mathbf{y})},$$

with $\mathbf{B} = diag(1, 2, 3, 0)$. John objected that this is not a metric on $\mathbb{R}^4$. Paul insists that it is. Who is right? Justify your answer. **[7 marks]**

Question 2

Let $\mathcal{X} = \mathbb{R}^2$ denote the 2-D space of data points with orthonormal basis $(\mathbf{e}_1, \mathbf{e}_2)$. Consider the hypothesis class H defined by the set of right triangle classifiers ABC with two sides parallel to the axes, with $\vec{AB}/\|\vec{AB}\| = \mathbf{e}_1$ and $\vec{AC}/\|\vec{AC}\| = \mathbf{e}_2$, and $\|\vec{AB}\|/\|\vec{AC}\| = \alpha$ for some positive real $\alpha \in \mathbb{R}$. See Figure 1 for an illustration of the definition of right triangle classifiers. The classifier assigns label '$+$' to a point if it is inside the right triangle and '$-$' otherwise. Under **realisability** assumption and using $0-1$ loss function:

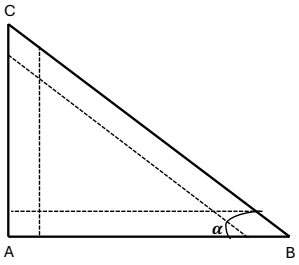


Figure 1: Illustration of the right triangle classifiers

(a) Design an ERM-H algorithm using the observed training set. Explain why the proposed algorithm is ERM-H. **[5 marks]**

(b) State and explain the generalization error of the classifier returned by the ERM-H algorithm from part (a). **[5 marks]**

(c) Derive the sample complexity required to PAC learn the hypothesis class H. (Hint: Use the strips illustrated in Figure 1). **[10 marks]**

Question 3

Let $\mathcal{X} = \mathbb{R}^2$ denote the 2-D space of data points and $\mathcal{Y} = \mathbb{R}$ denote the space of output labels with $h : \mathcal{X} \to \mathcal{Y}$. Consider the following hypothesis class of parameterized linear regressors,

$$H = \{h_{\boldsymbol{\theta}} : h_{\boldsymbol{\theta}}(\mathbf{x}) = \boldsymbol{\theta}^T \mathbf{x} = \theta_1 x_1 + \theta_2 x_2, \theta_2 \in \{0.1, 0.5\}, \theta_1 \in \mathbb{R}, \|\boldsymbol{\theta}\|_2^2 = 1\}, \qquad (1)$$

where the parameter $\boldsymbol{\theta}^T = (\theta_1, \theta_2)$ satisfies that the squared 2-norm of $\boldsymbol{\theta}$ is 1 and that $\theta_2 \in \{0.1, 0.5\}$.

(a) Suggest a loss function to evaluate the performance of a hypothesis $h \in H$ for regression problem. **[5 marks]**

(b) State the (general) condition under which a hypothesis h returned by an algorithm A is termed agnostically probably approximately correct. **[5 marks]**

(c) Is the loss function from part (a) bounded? If not, suggest a modification of the loss function to ensure that it is bounded. State the sample complexity required to agnostically PAC learn the hypothesis class H in (1). **[10 marks]**

This page intentionally left blank.

Mathematical Foundations of Artificial Intelligence 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.