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

Machine Learning

Mock Exam

Time allowed: 2 hours

[Answer all questions]

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.

The end of the paper has an appendix with some formulas and definitions that you may find useful.

Question 1 Core Concepts

(a) Explain what is a supervised learning algorithm, including its core goal. Please give your answer formally, making use of appropriate mathematical symbols and terminology whenever relevant. **[5 marks]**

(b) Answer the following questions regarding feature transformations in the context of machine learning:

- What is a non-linear feature transformation? Please provide a detailed definition.
- When could it be useful to adopt a non-linear feature transformation?

[5 marks]

(c) Question covering Jian's part of the module to be added here, but the format / style of these questions will be similar to the kind of questions asked above.

[5 marks]

(d) Question covering Jian's part of the module to be added here, but the format / style of these questions will be similar to the kind of questions asked above.

[5 marks]

Question 2 Classification

(a) Consider a machine learning problem with two parameters to be learned (w_1 and w_2) and the following loss function:

$$E(\mathbf{w}) = w_1^2 + 0.001 w_2^2,$$

where $w_1 \in R$ and $w_2 \in R$.

- Explain in detail why Gradient Descent could be inefficient to minimise this function. **[5 marks]**
- Explain in detail why standardisation of the input variables (e.g., by deducting the mean from each input variable and then dividing each input variable by the standard deviation) could be potentially helpful to improve the efficiency of Gradient Descent for this problem. **[5 marks]**

(b) In the dual representation of the Support Vector Machines, it could happen that a training example is on the margin, but is associated to a Lagrange multiplier of zero. Despite being on the margin, this training example is not considered as a support vector, as it would not contribute towards the predictions made by the model. Explain in detail why an example that is on the margin could possibly have a Lagrange multiplier of zero.

Hint: you can explain that by reflecting about the steps to go from the primal to the dual representation. **[10 marks]**

Question 3

(a) Question to be added based on Jian's content. **[10 marks]**

(b) Question to be added based on Jian's content. **[10 marks]**

Appendix

Primal representation of hard margin support vector machines

$$\underset{\mathrm{w},b}{\operatorname{argmin}} \left\{\tfrac{1}{2}\|\mathbf{w}\|^2\right\}$$

$$\text{subject to} \quad y^{(n)} h(\mathbf{x}^{(n)}) \geq 1, \quad \forall(\mathbf{x}^{(n)}, y^{(n)}) \in \mathcal{T},$$

where $\mathbf{w}$ and b are the parameters to be learned, $\mathbf{x}^{(i)} \in R^d$ are the input variables of example i, d is the number of input variables, $y^{(i)} \in \{-1, 1\}$ is the output label of example i, $\mathcal{T}$ is the training set, $h(\mathbf{x}) = \mathbf{w}^T\phi(\mathbf{x}) + b$, and $\phi(\mathbf{x})$ is a feature embedding. When $\phi(\mathbf{x}) = \mathbf{x}$, no embedding is being used.

Intermediate step between primal and dual hard margin support vector machines

$$\operatorname{argmin}_{\mathrm{w},b} \left\{\tfrac{1}{2}\|\mathbf{w}\|^2 + g(\mathbf{w}, b)\right\}$$

$$\text{where} \quad g(\mathbf{w}, b) = \max_{\mathrm{a}} \textstyle\sum_{n=1}^{N} a^{(n)}(1 - y^{(n)}(\mathbf{w}^T\phi(\mathbf{x}^{(n)}) + b))$$

$$\text{subject to} \quad a^{(n)} \geq 0, \ \forall n \in \{1, \cdots, N\}$$

where $a^{(i)}$ is the Lagrange multiplier associated to training example i, N is the number of training examples, and $\mathbf{x}^{(i)} \in R^d$ are the input variables of example i.

Dual representation of hard margin support vector machines

$$\underset{\mathrm{a}}{\operatorname{argmax}} \tilde{L}(\mathbf{a}) = \textstyle\sum_{n=1}^{N} a^{(n)} - \tfrac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} a^{(n)}a^{(m)}y^{(n)}y^{(m)}k(\mathbf{x}^{(n)}, \mathbf{x}^{(m)})$$

$$\text{subject to} \quad a^{(n)} \geq 0, \ \forall n \in \{1, \cdots, N\} \text{ and } \textstyle\sum_{n=1}^{N} a^{(n)}y^{(n)} = 0,$$

where $k(\cdot, \cdot)$ is the kernel function.

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.