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

LH Neural Computation

Main Summer Examinations 2023

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.

Question 1

Let $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}$ be vectors in $\mathbb{R}^d$. Consider the minimization of the following function

$$C(\mathbf{w}) = \frac{1}{2n} \sum_{i=1}^{n} \|\mathbf{w} - \mathbf{x}^{(i)}\|_2^2,$$

where $\|\cdot\|_2$ is the Euclidean norm, i.e., $\|\mathbf{w}\|_2^2 = \sum_{j=1}^{d} w_j^2$ for $\mathbf{w} = (w_1, \ldots, w_d)^\top \in \mathbb{R}^d$.

(a) What is the global minimiser of C? Give your arguments. **[5 marks]**

(b) Suppose we apply stochastic gradient descent to this problem with $\mathbf{w}^{(0)} = (0, 0, \ldots, 0)^\top$, step size $\eta_t = 1/(t+1)$ and $i_t = t+1$ $(t < n)$, i.e., when we derive $\mathbf{w}^{(t+1)}$ from $\mathbf{w}^{(t)}$ we use $\eta_t = 1/(t+1)$ and $\mathbf{x}^{(t+1)}$ to compute a stochastic gradient. Compute $\mathbf{w}^{(1)}, \mathbf{w}^{(2)}, \mathbf{w}^{(3)}$. Based on these computation, write a general formula for $\mathbf{w}^{(k)}$, $k \leq n$. Give your arguments to explain this general formula. **[11 marks]**

(c) Suppose we apply gradient descent to train a neural network. In which step of gradient descent do we use backpropagation? **[4 marks]**

Question 2

We have the following questions related to convolutional neural networks (CNNs). Please answer them with **justifications**.

(a) Given the convolution kernel $W \in \mathbb{R}^{3\times 3}$ and matrix $A \in \mathbb{R}^{5\times 5}$, we perform the following convolution with padding $= 0$ and stride $= 1$:

$$W \circledast A = B,$$

where $\circledast$ represents a 2D convolution, and A and W are respectively given as follows:

$$A = \{a_{ij}\} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 1 & 2 & 3 & 4 \\ 4 & 5 & 1 & 2 & 3 \\ 3 & 4 & 5 & 1 & 2 \\ 2 & 3 & 4 & 5 & 1 \end{bmatrix} \text{ and } W = \{w_{ij}\} = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}.$$

Please fill in the missing entries of the matrix $B \in \mathbb{R}^{3\times 3}$ below:

$$B = \{b_{ij}\} = \begin{bmatrix} \ldots & \ldots & 0 \\ \ldots & \ldots & \ldots \\ 5 & \ldots & \ldots \end{bmatrix}.$$

Note that you need to write down B on your answer sheet. **[7 marks]**

(b) Next, following the question (a) above, during convolution each element of A will be multiplied by some element of the convolution kernel $W \in \mathbb{R}^{3\times3}$ a number of times. Fill in the following matrix with the number of times each element is multiplied by an element of the weight matrix W:

$$\begin{bmatrix} 1 & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & 9 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}.$$

Note that you need to write down W on your answer sheet. **[7 marks]**

(c) We then calculate the following loss function f, which involves B computed in the question (a) above and some matrix $C \in \mathbb{R}^{3\times3}$:

$$f = \Sigma_{i=1}^{3}(b_{ii} - c_{ii})^2,$$

where c_{ij} denotes the entries of matrix C which are defined as follows:

$$C = \{c_{ij}\} = \begin{bmatrix} -10 & 0 & 0 \\ 0 & -10 & 0 \\ 0 & 0 & -10 \end{bmatrix}.$$

As is the case with neural networks, the weights of a CNN will be updated using backpropagation which relies on the chain rule to calculate the derivatives of the loss function with respect to the weights. Use the chain rule to calculate and evaluate the derivatives $\frac{\delta f}{\delta w_{11}}$ and $\frac{\delta f}{\delta a_{11}}$, where w_{11} and a_{11} are the first entry of W and A, respectively. Note that W and A are given in the question (a). **[6 marks]**

Question 3

(a) An auto-encoder (AE) consists of an encoding unit f_φ, a latent representation z, and a decoding unit g_θ. The goal of an auto-encoder is to learn to produce output $\hat{x} = g_\theta(z) = g_\theta(f_\varphi(x))$ for a given input x, such that $\hat{x} = x$ and where $z = f_\varphi(x)$.

(i) During the model training process, the auto-encoder model can naively learn the identity function making it a useless model. Describe briefly why we consider such a model as a useless model and how we make the auto-encoder learn a useful latent representation z instead of an identity function. **[5 marks]**

(ii) Given the trained auto-encoder consisting of f_φ, g_θ and z, consider that it has the ability for good self-reconstruction. Is this auto-encoder suitable for generating synthetic (or new) data? Justify your answer with brief reasoning. **[4 marks]**

(b) A variational auto-encoder (VAE) relies on the following loss function which consists of two terms:

$$\mathcal{L}_{VAE} = \mathcal{L}_{rec} + \mathcal{L}_{reg},$$

where $\mathcal{L}_{rec}$ represents the reconstruction loss and $\mathcal{L}_{reg}$ represents the regularisation loss. Briefly describe about what happens if we exclude the $\mathcal{L}_{reg}$ term from the VAE loss function such that $\mathcal{L}_{VAE} = \mathcal{L}_{rec}$ to train the VAE. **[5 marks]**

(c) A generative adversarial network (GAN) consists of two units: a generator G_θ to generate fake data samples and a discriminator D_φ to recognise whether a sample is real or fake. We need to design a good loss function to train the GAN discriminator unit for learning its parameters φ. We have designed the below loss function for discriminator learning:

$$\min_{\varphi} \mathbb{E}_{x \sim p_{data}(x)}\Big[-\log\big(1 - D_\varphi(x)\big)\Big] + \mathbb{E}_{z \sim p(z)}\Big[-\log\big(D_\varphi(G_\theta(z))\big)\Big],$$

where $\mathbb{E}$ denotes the expectation operator, $x \sim p_{data}(x)$ denotes input sample x drawn from real data distribution $p_{data}(x)$, and $z \sim p(z)$ denotes GAN generated data z drawn from fake data distribution $p(z)$.

Briefly describe what this loss function is doing and how it can be changed to help in GAN discriminator training. **[6 marks]**

This page intentionally left blank.

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.