A37253

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

Third Year Undergraduate

06-32167

LH Neural Computation

Main Summer Examinations 2025

Time allowed: 2 hours

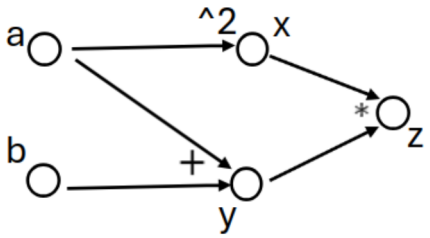
[Answer all questions]

Note

Answer ALL questions.

Question 1

(a) Consider the computation graph below. The + and * symbol correspond to addition and multiplication respectively. The symbol '^2' indicates that the variable is squared.



(i) Write down the equation corresponding to the computation graph, describing the calculation of the output variable z from the inputs a,b. Which part of the equation corresponds to the intermediate variables x and y? **[2 marks]**

(ii) Consider the graph from above and $a = 1$ and $b = -2$. Use the forward pass of the backpropagation algorithm to compute x, y, and z. **[3 marks]**

(iii) Use the backward pass to compute the partial derivatives of z in the direction of x, y, a, and b. **[14 marks]**

(b) (i) Consider the following linear regression problem: Given a set of training data, you want to predict the price of a real estate property from the number of bedrooms in a home using a squared error loss function. Sketch the computation graph for the prediction and the resulting loss. **[8 marks]**

(ii) Mark which nodes in your graph from (b)(i) correspond to the number of bedrooms x, the predicted price $\hat{y}$, the true price y, the loss L and the learnable parameters w and b. **[6 marks]**

Question 2

(a) A convolutional neural network (CNN) model is applied to process an RGB image of 32 x 128 pixels to solve a binary classification problem (e.g., classify whether there is a cat in the image). The CNN consists of a two-dimensional convolutional layer, followed by a max-pooling layer, a non-linear activation function, a flattening operation, a fully connected layer, and a sigmoid function.

(i) The convolutional layer consists of 32 filters of size 5×5, with a stride of 1 and a padding of size 2 on each side of the horizontal and vertical dimensions, including bias terms.

(ii) The max-pooling layer has a pool size of 2 x 2, a stride of 2 and does not use padding.

(iii) The non-linear activation function is the ReLU function.

(iv) The fully connected layer produces a single output, and includes a bias term.

Please calculate **the number of parameters** and **the output sizes** for each of (i) the convolutional layer, (ii) the pooling layer, (iii) the non-linear activation function, and (iv) the fully connected layer. **[8 marks]**

(b) (i) A vanilla recurrent neural network (RNN) uses the following equations, where the hidden-state vector h_t, the output vector $\hat{y}_t$ and the input vector x_t are column vectors:

$$h_t = \tanh(W_{hh}^\top h_{t-1} + W_{xh}^\top x_t) \tag{1}$$

and

$$\hat{y}_t = W_{hy}^\top h_t. \tag{2}$$

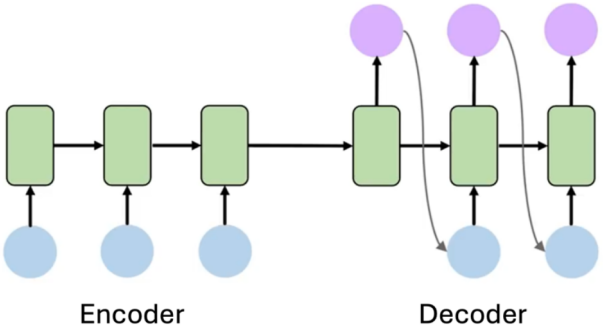
Note there are matrix transpose symbols.

Suppose we obtained the weight matrices through training as

$$W_{hh} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 2 & 1 & 3 \end{bmatrix}, W_{xh} = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 1 & 3 \end{bmatrix}, W_{hy} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 3 & 3 \end{bmatrix} \tag{3}$$

What are the dimensions of the hidden-state vector h_t, the output vector $\hat{y}_t$ and the input vector x_t of each time step? **[3 marks]**

(ii) Consider the below RNN encoder-decoder model:



Such an RNN network can be used to translate a French word (as a sequence of French letters) to its corresponding English word (as a sequence of English letters). The input French words are collected from the raw-text conversations of an online social platform, which are then translated by experts to provide the input-output pairs for translation.

In this scenario, why could an RNN be needed for translating a French word into an English word? Hint: Consider why a simple French-to-English dictionary lookup might be insufficient. Please keep your answer brief. **[3 marks]**

(iii) For the French-to-English word translation task described in Question (ii), briefly explain the process for the **RNN decoder** to generate the output sequence **during inference**, after encoding is finished.

Note that we can use the <SOS> and <EOS> tokens to represent the start of sequence and the end of sequence, respectively. Please keep your answer brief.

- How is the hidden state initialized?
- What are the input and output at the first time step?
- What are the input and output of the remaining time steps?
- What is the stopping criterion?

[8 marks]

(c) We want to design a deep learning model to output English words from images containing handwritten English words. Each handwritten word is written from left to right, and is presented as a grey-scale image of fixed size (32 x 128 pixels).

Please provide a model that uses a combination of CNN and RNN to achieve this task. Address the following points in your explanation:

(i) Describe the inputs and outputs of the CNN and RNN components, and explain how the CNN and RNN components are connected. Briefly highlight any key aspects of the CNN and RNN components that are essential for the task. Please keep your answer brief. **[8 marks]**

(ii) State at least two advantages of your approach. Please keep your answer brief. **[4 marks]**

Question 3

(a) A company is predicting house prices based on the number of bedrooms and the size of the house (in square metres). The following is the collected data for training the model.

Bedrooms (x_1)	Size (x_2, in m^2)	Price (y, in \$1000)
2	50	150
3	75	200
3	80	210
4	100	260
5	110	310

A linear regression model has been trained using the above data to predict house prices. The model weights are:

- Model weight corresponding to Bedrooms (x_1) feature: $w_1 = 43.5$
- Model weight corresponding to Size (x_2) feature: $w_2 = 0.5$
- Model intercept / bias: $b = 37$

The company applies this trained model to predict the price of a luxury house with 8 bedrooms and 350 m^2 in size. According to a real estate report, this house is priced at \$920,000.

Answer the following questions:

(i) Use the model to predict the price of this luxury house. Show your calculations. **[5 marks]**

(ii) Based on numerical values, briefly compare the model's predicted price with the reference actual price (\$920,000). Using this example, briefly discuss whether the model can generalise well to luxury houses with significantly higher values, and in the context of feature values beyond the training data range. Provide at least two reasons to support your answer. Please keep your answer brief. **[5 marks]**

(b) (i) A machine learning engineer is working on optimising a linear model using gradient-descent based methods. The dataset consists of $1,000,000$ training samples, and they are considering different approaches. The engineer wants to use gradient descent (GD) but is worried about computation time. Would stochastic gradient descent (SGD) be a better alternative? What could be a downside of using SGD? Justify your answer. Please keep your answer brief. **[4 marks]**

(ii) Consider a simple linear model with a single parameter w. You are given a training dataset consisting of three data points:

$$(x_1, y_1) = (1, 2), (x_2, y_2) = (2, 3), (x_3, y_3) = (3, 4). \tag{4}$$

The loss function (mean squared error) is defined as:

$$L(w) = \frac{1}{n}\sum_{i=1}^{n}(wx_i - y_i)^2, \tag{5}$$

with n denoting the number of data points. With respect to an individual data point (when $n = 1$), provide the mathematical expression for computing its respective squared loss and derive the mathematical expression for the gradient (or derivative) of this loss function. **[4 marks]**

(iii) Using only the first data point $(x_1, y_1) = (1, 2)$ from above and the derivative computed from above, compute the gradient at $w^0 = 0$ and perform one weight update step with the learning rate $\alpha = 0.1$ to compute w^1 using the stochastic gradient descent method. **[6 marks]**

(c) A company is using a standard denoising diffusion probabilistic model (DDPM) to generate high-resolution realistic images. However, the company has determined that the image generation process is too slow during the inference (sampling) phase for its production needs. To improve inference speed for image generation, they are considering switching to a latent diffusion model (LDM).

Answer the following questions:

(i) Would switching to LDM solve the inference speed issue? In your answer, briefly describe how operating in a latent space might affect the overall sampling speed compared to DDPM and the role of auto-encoders in inference speed. Please keep your answer brief. **[4 marks]**

(ii) What trade-offs (or disadvantages) might the company face when adopting LDM instead of standard DDPM? Briefly justify your answer by considering potential impacts on image quality, as well as any changes in model complexity. Please keep your answer brief. **[5 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.