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 Machine Learning and Intelligent Data Analysis

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 Regression

(a) The least squares error function is defined as

$$\mathcal{L}(\mathbf{w}) = \sum_{i=1}^{N} (t_i - f(x_i, \mathbf{w}))^2 .$$

This function is commonly used to measure how well a function $f(x, \mathbf{w})$ parameterised by $\mathbf{w}$ fits a set of N data points $\mathcal{D} = \{(x_i, t_i)\}_{i=1}^{N}$.

The likelihood that a data point t was generated by model $f(x, \mathbf{w})$ is $p(t|f(x, \mathbf{w}))$. Explain how, and under what assumptions, the least squares error is derived from the likelihood. You do not need to reproduce all of the mathematical steps of the derivation. **[6 marks]**

(b) Given some dataset, the expected value of the LSE can be written as

$$\mathbb{E}[\mathcal{L}] = \sigma^2 + \text{var}[f] + (h - \mathbb{E}[f])^2,$$

where σ^2 is the variance of the data, f is the estimated fit, and h is the true data generating function. Explain the terms in this expression and its relevance for learning. **[6 marks]**

(c) Given the data point $(2, 1)$, sketch a diagram of the likelihood in parameter space that this data point was generated by functions of the form $f(x, \mathbf{w}) = w_0 + w_1 x$. Your sketch should cover the domain $\{w_0, w_1\} \in [-1, 1]$. **Explain your reasoning.** **[8 marks]**

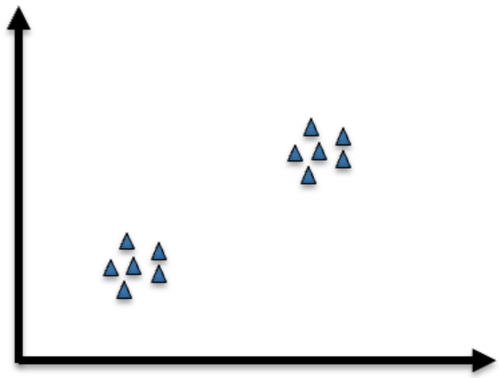
Question 2 Clustering, Dimensionality and Text Analysis

(a) Cluster the data in the table below using hierarchical clustering with Euclidean distance and single linkage, and draw the dendrogram.

Label	A	B	C	D	E
Coordinates	(4,2)	(7,8)	(3,2)	(3,4)	(8,7)

[6 marks]

(b) The graph below shows a two dimensional dataset.



(i) Reproduce the plot and draw the first and second principal components. Your drawing does not need to be completely accurate but should capture the key features. **Explain your reasoning**

(ii) If you were to use PCA to reduce the dimensionality of this data to just 1 dimension, show how the points will be mapped onto the new dimension (your drawing should give the general idea of the mapping).

(iii) Describe one way to determine how many dimensions should be kept in PCA.

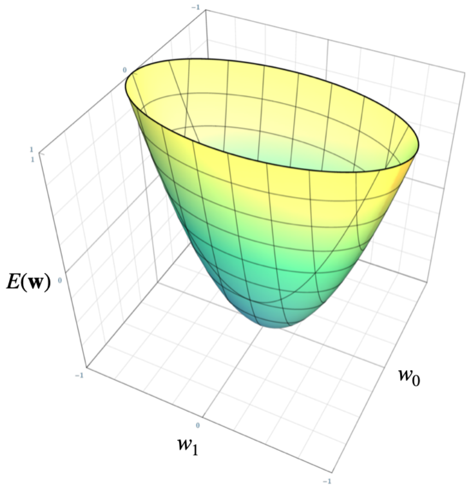
[6 marks]

(c) Document vectorisation and Page Rank are both methods for ranking documents. Document vectorisation allows documents to be ranked by their similarity to a query document. Page Rank provides a way to rank a collection of *linked documents* by considering their *authority* which is derived from the connectivity of each document.

Explain briefly how these two methods might be combined to implement a basic search engine for a collection of linked documents that takes both document content and document authority into account. Briefly discuss any limitations of your approach. **[8 marks]**

Question 3 Classification

(a) Which algorithm (Gradient Descent or Iterative Reweighted Least Squares) would be better to learn the weights w_0 and w_1 of a logistic regression model for a problem with the loss function $E(\mathbf{w})$ below, which is an elliptic quadratic function? **Justify** your answer by explaining how these two algorithms would work **in the context of this problem**. **[6 marks]**



(b) Logistic regression models for binary classification can be trained by maximising the log-likelihood:

$$ln(\mathcal{L}(\mathbf{w})) = \sum_{i=1}^{N} y^{(i)} \; ln \; p_1(\mathbf{x}^{(i)}, \mathbf{w}) + (1 - y^{(i)}) \; ln \; (1 - p_1(\mathbf{x}^{(i)}, \mathbf{w})).$$

where $\mathbf{w}$ are the weights of the logistic regression model, $y^{(i)} \in \{0, 1\}$ is the output variable of training example i, $\mathbf{x}^{(i)} \in \mathcal{X}$ are the input variables of training example i, $\mathcal{X}$ is the input space, N is the number of training examples, and $p_1(\mathbf{x}^{(i)}, \mathbf{w})$ is the probability of example i to belong to class 1 given $\mathbf{x}^{(i)}$ and $\mathbf{w}$.

How would you modify the log-likelihood function above so that it also works for problems with $M > 2$ classes? **Explain** your function.

PS: Please create a **single** log-likelihood function and make sure to define any variable or symbol that is different from the ones defined above.

[7 marks]

Question 3 continued over the page

(c) Prove that the kernel below is a valid kernel based on the kernel composition rules below and the fact that $\mathbf{x}^T\mathbf{z}$ is a valid kernel.

$$k(\mathbf{x}, \mathbf{z}) = 10(e^{(\mathbf{x}^T\mathbf{z})})^2 + 2 + \mathbf{x}^T\mathbf{z}$$

Kernel composition rules, given two valid kernels $k_1(\mathbf{x}, \mathbf{z})$ and $k_2(\mathbf{x}, \mathbf{z})$:

1	$k(\mathbf{x}, \mathbf{z}) = ck_1(\mathbf{x}, \mathbf{z})$, where $c > 0$ is a constant
2	$k(\mathbf{x}, \mathbf{z}) = f(\mathbf{x})k_1(\mathbf{x}, \mathbf{z})f(\mathbf{z})$, where $f()$ is any function
3	$k(\mathbf{x}, \mathbf{z}) = q(k_1(\mathbf{x}, \mathbf{z}))$, where $q()$ is a polynomial with non-negative coefficients
4	$k(\mathbf{x}, \mathbf{z}) = e^{k_1(\mathbf{x}, \mathbf{z})}$
5	$k(\mathbf{x}, \mathbf{z}) = k_1(\mathbf{x}, \mathbf{z}) + k_2(\mathbf{x}, \mathbf{z})$
6	$k(\mathbf{x}, \mathbf{z}) = k_1(\mathbf{x}, \mathbf{z})k_2(\mathbf{x}, \mathbf{z})$

[7 marks]

End of Paper

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.