University of Exeter

COURSEWORK SPECIFICATION

COM3023 – Machine Learning and AI

Module Leader: Dr Zeyu Fu

Academic Year: 2024/25

Title: **Machine Learning and AI**

Submission deadline: **21st of March**

This assessment contributes **30%** of the total module mark and assesses the following **intended learning outcomes**:

- Demonstrate the theoretical foundations of machine learning and AI methods;
- Choose appropriate analysis methods for new problems;
- Understand the principles underlying different machine learning and AI techniques;
- Describe and compare different theoretical approaches to a single problem;
- Learn a variety of machine learning and AI methods and apply them to real problems;
- Plan and write a technical report;
- Adapt existing technical knowledge to learning new methods.

AI-supported use is permitted in this assessment.

This is an **individual** assessment and you are reminded of the University's Regulations on Collaboration and Plagiarism. You must avoid plagiarism, collusion and any academic misconduct behaviours. Further details about Academic Honesty and Plagiarism can be found at `https://ele.exeter.ac.uk/course/view.php?id=1957`.

Important notes: The submitted ZIP file must include a **report** along with the **code**.

1 Assignment

This assignment aims to investigate Bayesian Neural Networks (BNNs), focusing on how they use Bayesian inference to manage uncertainties in artificial neural network parameters. Given the challenge of deriving a closed-form solution for the posterior distribution in BNNs, the assignment explores two approximation methods: Variational Approximation (VA) and Markov Chain Monte Carlo (MCMC) techniques. An important component of this assignment is to conduct a comparative study of VA and MCMC regarding their effectiveness in approximating posterior distributions in BNNs, considering aspects such as accuracy, computational efficiency, and applicability to real-world data. This comparison aims to deepen understanding of Bayesian inference in neural networks and enhance practical skills in applying these approximation methods.

For this comparative study, you are not restricted to using any programming language and can select any real-world data set (either a regression or classification problem) for demonstration. You can implement the methods by yourself or use a library or software (e.g., PyMC, PyTorch, TensorFlow or scikit-learn).

Initial references:

1. Andrieu, C., de Freitas, N., Doucet, A. and I. Jordan, Michael. An Introduction to MCMC for Machine Learning. Machine Learning 50, 5–43 (2003). `https://doi.org/10.1023/A:1020281327116`
2. Magris, M. and Iosifidis, A. Bayesian Learning for Neural Networks: an algorithmic survey. `https://arxiv.org/pdf/2211.11865.pdf`
3. Christopher M. Bishop. Pattern Recognition and Machine Learning. 2006.
4. David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. Variational inference: A review for statisticians. Journal of the American Statistical Association, 112(518): 859–877, 2017.
5. Li, Z., Liu, H., Luo, C., Fu, G. (2021). Assessing surface water flood risks in urban areas using machine learning. Water, 13(24), 3520.

Dataset Information: A flood-related dataset for classification has been provided on ELE and is available for this assignment.

- Flood hazard levels are categorized into five classes, making it a multi-class classification problem. There is an option to convert the provided flood-related dataset into a binary classification task by distinguishing between flood risk (hazard present) and no flood risk (hazard absent).
- The dataset contains multiple features. You may be interested in exploring feature selection techniques as a case study to identify the most relevant factors.

Alternatively, you may use other datasets for regression or classification tasks, but they must be sourced independently. Possible alternatives include datasets for house price prediction, stock price forecasting, or other applications where BNNs can be applied.

2 Requirement

This coursework requires the submission of a concise report along with the code for reproducibility. A specific list of requirements for the submitted report is given below:

- An introduction (or problem definition)
- A description of the background to the chosen technique(s), their advantages and disadvantages.
- A reasoned experimental design of the empirical work to be undertaken, including a description of the dataset used in this study, techniques to compare experimentally (if applicable), and how the results will be evaluated, etc.
- A presentation of the results obtained in an appropriate form (e.g. tables, plots, etc.).
- An analysis of the results.
- A conclusion.

The body of the pdf format report (excluding the title page and references) should be no more than 8 pages in length. It should be typeset in LaTeX, using the style file provided on the COM3023 ELE page.

Student declaration: You must prepare a declaration saying which of these applies.

- ❑ I have used GenAI tools for developing ideas.
- ❑ I have used GenAI tools to assist with research or gathering information.
- ❑ I have used GenAI tools to help me understand key theories and concepts.
- ❑ I have used GenAI tools to identify trends and themes as part of my data analysis.
- ❑ I have used GenAI tools to suggest a plan or structure for my assessment.
- ❑ I have used GenAI tools to give me feedback on a draft.
- ❑ I have used GenAI tools to generate image[s], figures or diagrams.
- ❑ I have used GenAI tools to proofread and correct grammar or spelling errors.
- ❑ I have used GenAI tools to generate citations or references.
- ❑ Other [please specify].
- ❑ I have not used any GenAI tools in preparing this assessment.

Please note that submitting your work without an accompanying declaration, or one with no ticked boxes will be considered a declaration that you have not used GenAI tools in preparing your work.

3 Marking criteria

The assessment will be marked using the following criteria.

Problem definition.	The degree to which the problem is concisely and clearly defined, including the motivation for the importance of Neural Networks, Bayesian approach and approximation inference and how these methods are linked to each other.	/15
Description of technique(s).	The degree to which the theoretical explanations have been well presented and compared with relevant references.	/25
Experimental design and results presentation.	The degree to which the experimental programme presented is clear and well-aligned to the question(s) being addressed. The degree to which the presentation of results is clear and supports the research topic being investigated.	/25
Analysis of results.	The degree to which the analysis of the results is appropriate, supported by the data, and provides insight into the question addressed.	/25
Presentation.	The degree to which the report as a whole is clear, concise, well-written, well-structured, well-formatted and devoid of grammatical error.	/10
Length penalty.	A penalty of 10 marks will be applied for each page (or part thereof) that the document body is over 8 pages (excluding references and title page).	
Code Requirement.	While the code itself is not graded, a 10-mark penalty will be applied if it is missing.	
	Total	/100
Feedback		

Note: You may consider including results (e.g., tables or graphs) to support your main report in an appendix, but the main report (8 pages) must be complete *without* the appendix. Marks will be awarded only for content in the main report and you must not use the appendix as a way of circumventing the 8-page limit.