MSc in Advanced Computer Science
FHS Computer Science; Mathematics and Computer Science;
Computer Science and Philosophy.

UNCERTAINTY IN DEEP LEARNING

Michaelmas Term 2025

Submission deadline 12 noon, Wednesday 7th January 2026, via Inspera.

There is a total of 100 marks available for this paper, you should attempt all parts of the paper.

NB: You must not discuss this examination paper with anyone.

TURN OVER

UNCERTAINTY IN DEEP LEARNING (2025-26)

Lecturer: Prof Yarin Gal (yarin.gal@cs.ox.ac.uk)

Course webpage: https://www.cs.ox.ac.uk/teaching/courses/2025-2026/UDL

Important dates:

Project release	Friday, Week 8 of Michaelmas Term	12:00pm (noon)
Submission deadline	Wednesday, Week (-1) of Hilary Term	12:00pm (noon)

1 Overview

In this assessment you will attempt to replicate a well specified key aspect of a published machine learning research paper, and additionally extend it in a novel way. The project should be conducted individually and should take roughly 10hrs, excluding writing up. You will present your findings in a final report. The report should discuss your contribution, making clear that you understand the key aspects of the paper and your contribution. The paper to reproduce is given below.

Essentially, think of your role as an inspector verifying the validity of the experimental results and conclusions of the paper. In some instances, your role will also extend to helping the authors improve the quality of their work and paper. You are **not** expected to reproduce **all** experiments in the paper; you are **only expected to reproduce the experiments specified below**.

2 Paper to reproduce

The paper selected in this assessment to reproduce and extend, building on the techniques you've learnt in the course:

- Yarin Gal, Riashat Islam, Zoubin Ghahramani. "**Deep Bayesian Active Learning with Image Data** ", ICML, 2017.
  https://proceedings.mlr.press/v70/gal17a

The key aspect from the paper to reproduce is the experimental results in sections 5.1-5.2. You are expected to implement the models and training pipeline from scratch (using standard deep learning libraries, see below). **Note that your understanding of core aspects in the paper will form the basis of the marking, so ensure you understand the paper well.**

3 Paper extension

After you reproduced the original paper's results you are expected to extend the paper. When extending the paper you are expected to use the expertise you acquired in the course. Below is a **minimum required extension for a passing mark** (see marking criteria in section §8.2). You are encouraged to go above and beyond the minimum expectations and extend the paper in additional novel ways of your choosing (i.e. beyond the minimum required extension, see below).

As a minimum extension (i.e. in addition to any novel extensions you may propose), please add a new baseline to the MNIST experiments using a hierarchical parametrised basis function regression model (i.e. a frozen neural network with inference only on the last weight layer), and using predictive

variance as the acquisition function. For this baseline, please provide two inference methods for the last weight layer: one with analytic inference (i.e. extend the derivation in slides 49-56 in "Lecture Slides - slides 3-4" to multivariate outputs, and implement it), and one with analytic MFVI inference (i.e. no sampling, implementing the derivation in slides 59-68 in "Lecture Slides - slides 5-6"), and note any differences between the two (based on the results you expect to see, and based on the results you actually saw). Please dedicate the first page in the appendix to a detailed derivation for the two inference methods, using your own choice of derivation techniques. For this extension, please change the paper's classification task to a regression task where you may use the same data as the original paper, but interpreting the one-hot target vectors as continuous multi-dimensional outputs (so MNIST class 0's one hot encoding [1, 0, 0, ..] would be interpreted as a 10-dimensional vector in $\mathbb{R}^{10}$ with scalar elements). Please report RMSE instead of accuracy. You may additionally use additional regression data if you prefer.

You are encouraged to go beyond the above and extend the paper in additional novel ways. The different ways of extending the paper in a novel way may vary in terms of the level of technical content, the difficulty of the mathematical derivation, the difficulty of algorithmic implementation, and the number of experiments performed. Please take these factors into consideration when choosing a novel contribution to pursue and when targeting the aspect to reproduce.

As an example novel contribution, consider adding label noise to the pool data and analysing the new results, extending the experimental results e.g. to consider data and algorithms that were not part of the original paper's study, or answering one of the following questions: "if I try really hard, can I get the baselines to work better than the proposed method?", or "can we extend the approach to work with models with multiple outputs?". Alternatively, try coming up with a novel question of your own.

4 Code submission

You will need to put the code for the project on a public anonymous github repository (**with a link to the repo in the report**) so that we can look at the code should we need to. This is a strict requirement.

5 Use of online code and AI-assistance tools

You may use standard deep learning libraries such as PyTorch. You have to explicitly say what online code resources you used, and what you've written yourself.

It is NOT permissible to use code you find online that goes beyond standard libraries (i.e. code that was written to solve the reproducibility challenge of this paper).

It is NOT permissible to use AI-generated code or AI-generated text in the report.

6 Report

The report should describe the selected paper, the particular experiments for replication, and the extension. It should summarise aspects of the paper relevant to the experiment (both to set the experiment in context and to lay out relevant technical details). The report should detail the experiments you ran, compare the results observed to those in the original paper, and draw conclusions. The report may further discuss the target questions of the paper selected, experimental methodology, your implementation details, analysis and discussion of findings, and conclusions on reproducibility

of the paper. Finally, the report should explicitly detail any novel contributions. You should end your report by critically evaluating your work and suggest what you would have done given more time. We will assess your understanding and critical judgement of the approaches, plus any original work that you've added: new datasets, baselines, improvements, etc.

The report should be written in the style of a NeurIPS Workshop paper (more details below). The written submission will be assessed based on the following criteria:

- motivation, clarity, and presentation
- originality and novelty
- coherence and depth of the study (i.e. investigation of novel extensions)
- contextualisation within the existing literature
- relation to the concepts discussed throughout the course
- presentation of results and reporting decisions (e.g. performance metrics)
- the presence of a balanced critical self-evaluation

7 Computational resources

Please note that training deep learning models takes time, and several experiments may be required for your project. Be sure to plan ahead and budget enough time for experimental exploration. Your report will not be evaluated according to the number of experiments, but according to its merits with regard to how well you demonstrate your understanding of the paper and the proposed extension.

You may use the same computational resources you used for the practicals (you should have access to pro licenses for Google Colab). However, it is understood that you may not necessarily have sufficient access to GPUs or high-performance computational resources for some tasks. Your report will not be assessed on the basis of your use of computational resources, but rather on the quality and appropriateness of the experimental design towards addressing the novel extensions. Should it be helpful, please feel free to use only a subset of the dataset as long as this is clearly documented in the report.

8 Submission details and marking criteria

The assessment for this course comprises of a written **report**. All submissions will be made through Inspera by the deadline. The assessors will grade the clarity and coherence of the content, the presentation of the work, and the progress made in the investigation of your novel extension.

8.1 Report

The written report should follow the NeurIPS Workshops format and not exceed 3 pages of main content, the standard page limit for NeurIPS workshop papers (plus unlimited additional pages for appendices and references at the end of the report, although assessors are not expected to read beyond the first 3 pages). See examples here. The report should use the NeurIPS Latex template provided in `https://neurips.cc/Conferences/2023/PaperInformation/StyleFiles`. Please use the '[final]' option in the style file to compile a camera-ready version for submission (but make sure to keep the submission anonymised).

The report should consist of the following sections:

- Introduction: 1/4 page (the problem to be solved, importance, overview of your novel extension and why it is important)

- Background: 1/2 page (how do people approach the problem in the field – this will require you to conduct a literature review of more recent papers in the field, and do additional reading)
- Methodology: 1/4 page (explain in a technical way the methodology of the paper to be reproduced)
- Minimal extension: 1/2 page (explain in a technical way the above required extension, and give key mathematical derivations)
- Novel extension: 1/2 page (explain in a technical way any additional novel extensions you've done, and give key mathematical derivations; you're allowed to include additional information in the appendix)
- Experiments: 3/4 page (explain the experiments you reproduced and give the reproduction results as well as any additional experiments)
- Analysis & Conclusions: 1/4 page (analyse your experiment results – do these match the reported results? give a critical evaluation of your work)
- References and citations
- Appendix: should start with 1 page with the detailed derivation required above. Additional pages may be used for additional figures, experiment details, etc.

The written report should be submitted as a single PDF file with the appendix concatenated at the end, and any supplementary material as a separate PDF or ZIP file, as appropriate. All documents should be submitted via Inspera.

Along with your report, please submit the code implemented for your project as described above. Any other supplementary material that aids the understanding of your work (e.g. figures, animations) may also be submitted, and will not be counted in the 3-page requirement. The package of supplementary documents should not exceed 10 MB.

8.2 Report marking scheme

Marks for the course will be based 100% on the written submission. Qualitative descriptors for report marking are given below.

- First Class (70-100): The candidate has demonstrated an excellent understanding of almost all of the material covered in the course and in the selected paper, with a commensurate quality of presentation, and has completed almost all of the expectations above satisfactorily, going above and beyond minimum expectations (e.g. with a novel extension beyond the minimum extension); This is further subdivided by:

  90-100 The candidate has shown considerable originality and insight going well beyond the straightforward completion of the reproducibility task (e.g. by implementing new baselines and evaluating on new datasets, or extending the method suggested in the paper in novel ways).

  80-89 The work submitted shows a near-perfect understanding of the selected paper and completion of the reproducibility task, but does not meet all the additional requirements above, or does but has some defects in presentation.

  70-79 The work submitted is of a generally high order, but may have minor errors in content and/or deficiencies in presentation.
- Upper second class (60-69): The candidate has demonstrated a good or very good understanding of much of the material in both the course and the selected paper, and has completed most of the reproducibility task satisfactorily, without showing the level of excellence expected of the above USM range.
- Lower second class (50-59): The candidate has demonstrated an adequate understanding of the material in the course and the selected paper, and an adequate ability to apply their understanding, without showing the level of understanding expected of the above USM range.

- Third class (40-49): The work submitted, while sufficient in quantity, suffers from sufficient defects to show a lack of adequate understanding of the course or the selected paper, or ability to apply results.
- Fail:

30-39 The candidate, while attempting a significant part of the project, has displayed a very limited knowledge or understanding at the level required.

0-29 The candidate has either attempted only a fragment of a project or has shown an inadequate grasp of basic material

9 Honour Code

While you may consult papers and online references (e.g. publicly available implementations) for ideas, it is important that you give credit to the original authors, cite the sources incorporated into your proposed approach or algorithm, and clearly state your intellectual extension of the work.

Code Any code that was used as a base for your project must be clearly referenced and cited in the paper, either as a footnote or as a full bibliographic entry. This includes open-source GitHub implementations.

Authorship You are expected to work independently, and to submit an individual report.