Peer–Review Questions for

ML Student Projects CS–C3240 – Machine Learning

Stage 2 – Full report

Opens: 27 Sep 2023, 12:00

Closes: 11 Oct 2023, 23:59

Stage 2 objectives

- Apply **at least 2** machine learning methods to the problem you formulated at stage 1
- Compare and analyze the results, to choose a final best method

Point distribution (out of 47 point total for the project)

- Submission points: 26
- Peer grading points: 11

Late submission policy

- Late submission open until **12 Oct 2023, 23:59**
    - Late submissions will incur a penalty of 30% on the stage 2 submission points
- Late submissions after the 12 Oct 2023 deadline will only be accepted due to illness (must provide valid medical certificate)
    - 0 points for peer grading

Overview

The objective of stage 2 is to submit the full report of your ML project. You need to apply and discuss **at least two methods**, compare them, analyse the results with which you will choose a final (best) method.

Check the report outline for the report format you are expected to submit at stage 2. **Make sure that your report addresses the questions outlined in the peer grading section below, because this is the criteria that your report will be assessed on.**

NOTE: You are expected to stick to the ML problem chosen for stage 1.

Peer-grading questions

Category 1. Introduction

Q1.1. Does the introduction clearly discuss the **application domain**?

Some examples of application domain are (1) medical diagnosis (classifying lung images into "Covid-19" vs. "No Covid-19"); (2) cross-country skiing (predict maximum daytime temperature to choose right ski wax).

- 1p – Yes, the application domain is very clearly explained.
- 0p – No discussion or explanation

Q1.2. Does the introduction section give a concise **summary or overview of the report** (E.g., "Section 2 discusses .., Section 3 then … and in Section 4 …")

- 1p – Yes
- 0p – No

Category 2. Problem formulation.

Q2.1 Does the dataset have a **multidimensional input** (i.e. more than one feature variable) and is the meaning of the **data points** clearly explained? Are the **types of data** also explicitly stated (binary, categorical, continuous, etc.)?

- 1p – Yes
- 0p – No, something is missing.

Category 3. Methods

Q3.1 Does the report clearly state **where the dataset was collected from**, the **number of data points** and give a **brief description** of the dataset?

- 1p – Yes, it is clearly stated where the dataset was obtained, and the description gives me a general understanding of the dataset.
- 0p – No, the source and the dataset are either not described at all or something is missing.

Q3.2 Does the report explain the **process of feature selection**? Note that theoretical justifications are not necessary, but instead we focus on the process of how the features were selected. It could be based on data visualisation, domain knowledge and other strategies.

- 1p – Yes, I fully understand the process of how the final features were chosen.
- 0p – No, it is not mentioned at all OR it is still unclear to me how the features were chosen.

Q3.3 Does the report discuss **at least two ML methods**?

If you choose "0p - No", please grant maximum 1p for Q3.4 - Q3.6.

- 2p – Yes, at least two ML methods are discussed.
- 0p – No.

Q3.4 Does the report clearly state **the models (hypothesis spaces)** and explain the **motivation** behind using them for the ML methods? Chapter 3 of mlbook.cs.aalto.fi discusses the models used by some well–known ML methods.

For example, *"Linear predictor maps are used as the visualisation shows a linear relationship between the features and the labels."*

- 4p – All chosen models and the motivation for using them are explained clearly.
- 2-3p – Some or all chosen models are explained only partially.
- 2-3p – I still do not understand why the author thinks they are reasonable design choices.
- 2-3p – The explanations are inconsistent with other parts of the report.
- 1p – The discussion is clear, but it only covers one method.
- 0p – Models are not discussed.

Q3.5 Does the report clearly specify the **loss functions** and explain the **motivation** behind using them to evaluate the quality of a hypothesis?

For example, "*The logistic loss is chosen as it allowed the use of a ready-made library for logistic regression"; "The Huber loss is used as it is robust towards outliers.*"

Other examples of loss functions can be found in Chapter 2 and Chapter 3 of mlbook.cs.aalto.fi. Note that it might be useful to use a different loss function for learning a hypothesis (e.g., logistic loss) than for computing the validation error (e.g., "accuracy" as the average 0/1 loss).

- 4p – All chosen loss functions and the motivation for using them are explained clearly.
- 2-3p – Some or all chosen loss functions are explained only partially.
- 2-3p – I still do not understand why the author thinks they are reasonable design choices.
- 2-3p – The explanations are inconsistent with other parts of the report.
- 1p – The discussion is clear, but it only covers one method.
- 0p – No, the loss functions are not discussed.

Q3.6 Does the report explicitly discuss how **the training and validation set** are constructed, the **size** of each set, and the **reason(s)** behind such design choice? If the project is an unsupervised learning task, does the report outline an alternative validation strategy?

Some examples are (1) using a single split into training and validation set, (2) k –fold cross validation, etc. (See Section 6.2 of mlbook.cs.aalto.fi

- 2p – The construction of training and validation sets are discussed very clearly. I also understand why the author thinks this is a reasonable design choice.
- 1p – The construction of training and validation sets are discussed superficially.
- 1p – The discussion is inconsistent with other parts of the report.
- 0p – The construction of training and validation sets are not discussed at all.

Category 4. Results and Conclusion

Q4.1 Does the report clearly (1) state and (2) compare the **training and validation errors and/or other metrics** obtained for each ML method considered in the report, and thereby (3) decide which is the **final chosen method**? For unsupervised projects, does the report provide a justified alternative evaluation or qualitative comparison to inform model selection?

See Section 6.2-6.3 of mlbook.cs.aalto.fi for comparison between different methods.

- 4p – Yes, all aspects are well-discussed.
- 1-3p – Partially, some elements are missing or inconsistent.
- 0p – Not discussed OR less than two methods are discussed, therefore no comparison.

Q4.2 Does the report (1) explain how the **test set** is constructed and (2) clearly state a **test error** of the final chosen ML method? For unsupervised projects, does the report include an alternative evaluation or qualitative comparison and discuss how the findings might generalize to new data?

A test set should consist of data points that have neither been used to train the ML method (training set) nor to choose between ML methods (validation set). The test error is the average loss incurred on a test set.

- 4p – Yes, test set is well-discussed.
- 1-3p – Partially, some elements are missing or inconsistent.
- 0p – No discussion.

Q4.3 Rate the quality of the **discussion of the obtained results** ("conclusion"). The conclusion should (1) briefly summarise the report and interpret the results; (2) discuss if the obtained results seem to be optimal or if there is room for improvement; (3) speculate about future directions on how to further improve the ML method.

A brief example: "*The training error was much smaller than the validation error which hints at overfitting. (...) As promising directions for future work, we consider collecting more training data.*"

Examples of other strategies to improve the ML method: use more/different features of datapoints, use different models, use a different loss function for training, etc.

- 4p – The performance (average loss) of the ML methods is interpreted and conclusions are provided with appropriate depth.
- 1-3p – Some elements are missing, e.g. the conclusion is mostly a repetition of the numeric results.
- 0p – No discussion.

Category 5. Overall Criteria.

Q5.1 Does the report follow the required outline? I.e., 1. Introduction, 2. Problem Formulation, 3. Methods, 4. Results, 5. Conclusions, 6. Bibliography/References (if any) and 7. appendix (code).

- 2p – Yes.
- 0p – No.

Q5.2 Rate **the quality of scientific writing in the report**. Are the report format and language use professional and clear? Is the report free of typos and incomplete sentences?

- 2p – The report is well-structured and easy to follow, the language is clear and concise, and there are almost no typos.

- 1p – The report is well-written overall, but it could be improved in some respects (please provide examples).
- 0p – The writing is not professional enough for a scientific report, e.g., there are a lot of incomplete sentences and typos.

Q5.3 Is the code file submitted as an appendix?

- 2p – Yes.
- 1p – Yes, but it seems to be incomplete.
- 0p – No.

Q5.4 To the best of your knowledge, does the report contain any existing material – either from this course, Kaggle, or other sources - **without clearly indicating the source**? Please report any suspicions of plagiarism (e.g. direct copying of text from other sources) to course staff.

- 1p – I have not seen the same ML problem or discussion in any of the mentioned places.
- 1p – I have seen the exact same ML problem in one of the mentioned places, but the source is clearly indicated in the report.
- 0p – I have seen the exact same ML problem or discussion in one of the mentioned places, but the source is not indicated in the report.