Grading Criteria (16.6 points per category)

Category 1. Title and Introduction

Q1.1 Is the **title** suitable? The title should (1) summarize the content of the project in a few words, (2) capture the reader's attention and (3) be specific.

Some examples of good titles: "Comparing Logistic Regression and naive Bayes' classifier in spam detection"; "Using Support Vector Machines for analyses of histological samples and cancer prediction"

Some examples of bad titles: "Spam filtering with extras"; "CS-C3240 ML project, final report"

- 1p – Good title
- 0p – Bad title

Q1.2. 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

Q1.3. 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 Definition.

Q2.1 Are the **data points**, **features**, **labels** clearly explained? Are the **types of data** also explicitly stated (e.g., integers, binary categories 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**? The methods could be polynomials with different degrees, or linear regression with and without regularization (ridge regression).

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?

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** obtained for each ML method considered in the report, and thereby (3) decide which is the **final chosen method**?

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?

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 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.

Q5.2 Is the code file submitted as an appendix?

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

Category 6. Overall assessment

Q6.1 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.

Q6.2 Does the report contain existing material – either from this course, Kaggle, or other sources - **without clearly indicating the source**?

- 1p – Yes, I have seen the exact same ML problem in one of the mentioned places, but the source is clearly indicated in the report.
- 1p – No, I have not seen the same ML problem or discussion in any of the mentioned places.
- 0p – Yes, 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.

Q6.3 If you answered 0p - Yes to the question above, does it also use the same model and loss function **without clearly indicating the source?**

- 5p – I chose "1p" in the question above.
- 5p – Different model and/or loss functions are used.
- 5p – The same model and loss function are used, and the source is clearly cited.
- 0p – The same model and loss function are used, but the source is **not** clearly cited.

Q6.6 Does the report contain paragraphs which are **copy-pasted** from other sources - such as the example projects, teaching material, Wikipedia, Kaggle, Stack Overflow and so on.

- 10p – No, I do not suspect any copy-pasting from other sources.
- 5p – Yes, a large part of the report is paraphrased from some source texts, but with indication of the source (students need to use their own words in the report).
- 0p – Yes, some parts of the report are copy-pasted without proper indication of the source (Please report this to course staff!).

Q6.7 (**Overall**) This part is for any aspects of the report that are not included in other grading criteria. Do you find the ML problem worth some extra points? For example, is the problem formulation **highly original** or did the student **explain the use of ML method outstandingly well**?

- 5p – I think the problem is very original
- 5p – I am impressed by how well the author explained the chosen ML methods.