Peer-Review Questions for

ML Student Projects CS-C3240 – “Machine Learning”

DRAFT - 01 Mar 2021

Category 1. Minimum requirements

*Q1.1 Does the report follow the required outline? I.e., 1. Introduction, 2. Problem Formulation, 3.Method, 4. Results, 5. Conclusions. (Bibliography/References)*

* No -> -10p
* Yes -> 0p

*Q1.2 Does the report discuss and compare at least two different candidate models (hypothesis spaces)? If only one model (hypothesis space) is considered does the report discuss why this model is considered optimal (in some sense)?*

* No -> -10p
* Yes -> 0p

*Q1.3 Does the report state the performance (average loss) of the learnt hypothesis on some test set? A test set consists of data points that are not used for any tuning or learning of model parameters such as weights of a linear hypothesis or number of features used.*

* No. The report only presents the average loss on the training set which is also used to tune model parameters. -> -10p
* Yes. -> 0p

*Q1.4 Does the report contain entire paragraphs which seem to be copy-pasted from another source (other student projects, Wikipedia, research articles, Kaggle, https://stackoverflow.com/, github repositories, …) without proper indication of the source.*

* No, I did not find indicators for copy-paste of other sources. -> 0p
* Yes, I found parts in the report which have been copy pasted. -> **-30p and reporting to course staff**

Category 2. Title and Introduction

*Q2.1 Is the title suitable? The title should: 1) summarize the content of the project in a few words*

*2) capture the reader's attention 3) be specific. 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 bad titles: "Spam filtering with extras"; "CS-EJ3221 Term project, final report"*

* No Title -> 0p
* Bad Title -> 1p
* Good title -> 2p

*Q2.2. Does the introduction discuss the application domain? The application domain could be, e.g., medical diagnosis (classifying lung images into “Covid-19” vs. “No Covid-19”) or cross-country skiing (predict maximum daytime temperature to choose right ski wax).*

* No -> 0p
* Yes -> 1p

*Q2.3. 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 …”)*

* No -> 0p
* Yes -> 1p

Category 3. Problem Definition

*Q3.1 Does the report clearly define what data points represent? Datapoints could represent drugs, images, sound recordings, cats, dogs, animals …*

0p – The meaning of datapoints is not defined at all.

1p – The meaning of datapoints is defined only superficially.

2p – The meaning of datapoints is defined very clearly.

*Q3.2 Does the report clearly define which properties of a datapoint are used as its features? E.g., if a datapoint represents a dog, then its features could be its weight and its age.*

0p – The report does not include any description of the features that characterize a datapoint.

1p – The report includes a description of the features that characterize a datapoint.

*Q3.3 Does the report clearly define which property of a datapoint is used as its label? E.g., if a datapoint represents a diamond, then its label could the amount of money sbdy is willing to pay for it (its “prize”).*

0p – The report does not define the label of a datapoint.

1p – The report clearly defines the label of a datapoint.

Category 4. Methods

Q4*.1 Does the report clearly describe the dataset which is used for training in the ML method? The description should include the source of the dataset and the number of datapoints contained in that set.*

0p – No discussion of the used dataset.

1p – Description of dataset is unclear.

2p – Description of dataset is clear.

*Q4.2 Does the report clearly explain the models (hypothesis spaces) underlying all ML methods that are used in the project? Chapter 3 of mlbook.cs.aalto.fi discusses the models used by some well-known ML methods.*

0p – Models are not discussed.

1p – Models are explained only partially.

2p – Models are explained clearly.

Q4.3 Does the report clearly specify the loss function(s) used to evaluate the quality of a hypothesis? 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).

0p – The loss functions used for training and validating the ML methods are not defined.

1p - The loss functions are only mentioned by name.

2p - The loss function is explicitly defined and some motivation for the used loss function is given (E.g., “ We used the logistic loss (as defined in Ch. 2 of mlbook.cs.aalto.fi) since this allows to use a ready-made library for logistic regression…”.

*Q4.4 Does the report explicitly discuss how the validation set is constructed, e.g., using a single split into training and validation set of or k-fold CV? (see Section 6.2 of mlbook.cs.aalto.fi)*

0p – The construction of training and validation sets are not discussed at all.

1p – The construction of training and validation sets are discussed superficially.

2p – The construction of training and validation sets are discussed very clearly. I would be able to reproduce this construction on my own.

Category 5. Results and Conclusion.

*Q5.1 Does the report clearly state the training and validation errors obtained for each ML method considered in the report? (See Section 6.2-6.3 of mlbook.cs.aalto.fi between different models (e.g., linear models that use different sets of features, or different ANN structures…).*

0p – No.

2p – Yes.

*Q5.2 Does the report clearly state a test error of the finally chosen ML method? The test error is the average loss incurred on a test set which consists of datapoints that have neither been used for the training of individual ML methods (training set) nor for choosing between different ML methods (validation set).*

0p – No.

2p – Yes.

*Q5.3 Rate the quality of the discussion of the obtained results (“conclusion”). Does the report discuss if the obtained results seem to be optimal or if there is room for improvement? Does the report speculate about future directions for how to further improve the ML method (collect more datapoints, use more/different features of datapoints, use different models, use a different loss function for training…)*

0p – No discussion of results can be found.

1p - Discussion is mostly a repetition of the numeric results.

2p – The training and validation errors obtained for the chosen ML methods are interpreted and conclusions are provided (see Section 6.6 of mlbook.cs.aalto.fi). E.g. “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.”

Category 6. Overall Criteria.

*Q6.1 Rate the use of language and the clarity of the report.*

0p – The quality of presentation and use of language is very poor.

1p – The report is clear enough. I could probably reproduce the project based on the report.

4p – Almost no typos or grammar errors and well-structured. The report is a pleasure to read.

*Q6.2 (Bonus) Do you find the machine learning problem very original, do you consider it a straightforward variation of well-known ML problems, e.g., examples from Python documentation, Kaggle.com or other webforums?*

+2p – I would never have thought about such a ML problem.