

Grading criteria for student project

CS-EJ3211 - Machine Learning with Python

Maximum points – 30 (+2 bonus).

0. Minimum requirements

Does the report follow the required outline? I.e., 1. Introduction, 2. Problem Formulation, 3. Methods, 4. Results, 5. Conclusions/Discussion.

-Yes

-No

Is the final predictor used to solve the problem obtained by selecting one of at least 2 candidate models, and/or hyperparameter tuning of at least one model?

-Yes

-No. Only a single model is considered, and its hyperparameters are not tuned

Are the estimates of the performance of learned predictors (used for model selection and hyperparameter tuning) based on the validation error on either a single validation set, or some other validation method such as k-fold cross-validation?

-Yes

-No. The performance estimates are based on the training error

There should not be reasonable grounds to believe that the project is plagiarized. It is OK to use code-snippets from tutorials etc., and inspiration from blogs, Kaggle kernels, and other sources. However, if large parts of the project and its implementation seem to be taken from the same source, and/or there is a large mismatch between the implementation and its description, we kindly wish the grader to contact course staff. We can then investigate if the project is plagiarized.

1. Title and Introduction

Criterion 1

Is the title suitable?

The title should:

- 1) summarize the content of the paper in a few words
- 2) capture the reader's attention
- 3) be unique enough to differentiate the paper from other papers in the same field.

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"

0p - No Title

1p - Bad title

2p - Good title

Criterion 2

Is the introduction provided good? A good introduction briefly introduces the application domain and presents the problem on a non-technical level. In addition, the introduction should contain a very concise description of the contents of the report.

0p - No introduction or a very unclear introduction

1p - Introduction is provided but some parts are missing, or the quality could be better

2p - Good introduction

2. Problem formulation

Criterion 3

Is the problem formulated clearly as a machine learning problem? In essence, are the data points, features, and labels clearly explained for the problem?

0p - There is no explanation of what the data points, features, and labels are, or the concepts are used incorrectly

1p - The data points, features, and labels are correctly defined, but the explanation is not clear

2p - Data points, features, and labels for the problem are clearly and correctly explained

Criterion 4

Does the report clearly specify the loss/metric used to evaluate the quality of the learned predictor? Examples of metrics are mean squared error, accuracy score and Area Under the Curve (AUC).

0p - The metric is not defined in the text

1p - The metric is only mentioned by name

2p - The metric is defined along with a brief explanation of it

3. Methods

Criterion 5

Is the dataset presented, including information on the source and size of the dataset?

0p - No description of the dataset

1p - Information on either the source or size is missing

2p - The dataset is presented, including information on the source and size

Criterion 6

Are (eventual) pre-processing methods described accurately? These include, for example, feature selection or -engineering, imputation of missing data, and standardization methods.

0p – Pre-processing methods are used, but not mentioned at all in the text.

1p - The pre-processing methods are mentioned in passing or described inaccurately

2p – The pre-processing methods are described accurately in the report. Alternatively, it is clearly argued why pre-processing methods are not necessary. One possible indicator for adding pre-processing of the data could be warning messages by Python methods (such as a “ConvergenceWarning” when fitting models)

Criterion 7

Are the models described accurately and in sufficient detail? For simple models presented in this course, the description should contain information on the form of the predictor functions (= hypothesis space) and loss function of the models. If you are using methods or models beyond those discussed in this course, e.g. deep learning, it is not required to give a precise mathematical description of underlying hypothesis space. It actually is subject of current research to understand the hypothesis space of deep neural nets. If you are using such complex methods (e.g., deep nets), you should mainly try to justify your choice (e.g. you suspect that the relation between features and label is heavily non-linear and therefore requires a deep neural net). A short description of (eventual) tuned hyperparameters should also be included.

0p – The models are only mentioned by name, or the description is inaccurate

2p - The models are described accurately, but the description could have been more informative

4p – The models are described accurately and in sufficient detail.

Criterion 8

Is the model validation and -selection/ hyperparameter tuning process described accurately? The description presents the validation scheme, e.g., validation is done using a single set of labeled data, or validation is done via 10-fold cross validation.

0p – The process is not described or is described inaccurately

1p - The process is described accurately but the description is lacking or unclear

2p – The process is described accurately and clearly

4. Results

Criterion 9

Is the performance estimate of the final model presented clearly?

0p - No results are presented

1p - Results are presented, but the presentation is unclear/cluttered

2p - Results are presented clearly

Criterion 10

Is the performance estimate of the final model based on its performance on a separate test set, which is not used to tune the model parameters (for which training set is used) or to tune the hyperparameters (for which validation set is used)?

0p - No. The performance estimate of the final model is calculated based on data that was used for selecting and/or tuning the model

1p - Yes

5. Discussion/Conclusions

Criterion 10

Quality of discussion/conclusions.

0p - The results are not discussed and no conclusions are provided as defined in the project instructions.

1p - Discussion is present, but it is mostly repetition of the results.

2p - Results are discussed and conclusions are provided. The analysis is of good quality.

6. Overall Criteria

Criterion 11

Rate the use of language and the clarity of the report. Is the report well formatted (E.g. placing of images, code cells etc.)?

0p - The text is unclear and very poorly organized, or the report is nearly unreadable due to grammatical errors.

1p - The report is clear, but not well structured

2p - Some explanations/ descriptions could have been clearer in the report or report is too short

3p - The text and code are well organized and clear. The report is a pleasure to read :)

Criterion 12

Is the report technically correct?

0p - Most arguments are wrong and there are several profound methodological errors

1p - The report contains few minor and one significant methodological errors.

2p - The report contains few minor errors.

3p - The report contains one - two minor errors.

4p - The report contains no methodological errors

Criterion 13 (Bonus)

Is the machine learning problem original? How challenging is the project overall?

+1p - The project idea, methodology and implementation are original **or** the data required significant pre-processing/cleaning

+2p - The data required significant pre-processing/cleaning, the methodology **and** implementation are original and required significant effort.