Peer-Review demo

Stage 1. Machine Learning - when and why?

Q1.1 Is the title suitable? The title should summarize the content of the project in a few words and be specific.

0 - No Title
1 - Bad Title
2 - Good title

Comment: Title is informative, indicates type of ML problem (classification) and application domain (sleep research on mice).

Q1.2. 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).

0 - No introduction or introduction with a major mistake(s)
1 - Yes, but poorly (multiple minor mistakes) or incompletely
2 - Yes

Comment: application domain is discussed in enough details, but it was not clear why sleep stage classification is important. Some examples on why researcher need to do this would be good.

Q1.3. Does the introduction explain the rationale for using ML approach?

0 - No explanation or explanation with major mistake(s)
1 - Yes, but poorly or incompletely
2 - Yes

Comment: Author states that ML approach will allow automation of the time and resource consuming manual process.

Q1.4. Does the introduction indicate the requirements of chosen ML system?

0 - No or yes, but with major mistake(s)
1 - Yes, but poorly or incompletely
2 – Yes

Comment: Author indicates that reliability, good generalization property and interpretability are the requirements and provides short explanation for each point.

Q1.5. Does the introduction discuss stakeholders and their goals? At least two stakeholders, ML engineer (i.e. student) and the user, should be discussed.

0 - No or yes, but with major mistake(s)
1 - Yes, but poorly or incompletely
2 - Yes

Comment: Author indicates some shared goals (accurate model with good generalization) and some conflicting goals, such as interpretability vs high accuracy, fast training vs fast inference.

How points for submission are computed (see formula here)?
grading levels {0,1,2}, weight =1, max points = 5

score (range 0-1) = (2/2*1 + 1/2*1 + 2/2*1 + 2/2*1 + 2/2*1) / (1+1+1+1+1) = 0.9
final points = 5*0.9 = 4.5

Stage 2. ML problem formulation – DATA

Q2.1. Does the section clearly describe dataset (source, collection process, sampling, biases, etc.)?

0 - No or yes, but with major mistake(s)
1 - Yes, but poorly or incompletely
2 - Yes

Comment: While the data described well, author could add more information about sampling methodology and distribution of labels. For example, “We used convenience sampling (i.e. used data which was available), as obtaining additional data was not possible. This means that dataset can be biased and may not represent overall data distribution. In practice, it means that model trained on our data may not generalize well to ECoG recording collected from other laboratories.
The data is highly unbalanced. While Wake and NREM represent majority of the recording, REM signal is much more rare and 'Artefact' is anomaly of the recording. Label 8 ‘not analyzed’ is very rare and we can filter out unlabeled epochs during data pre-processing step.
The dataset is static, i.e. the data collected and all data used for model training at once. We assume that data distribution will not change drastically with time, given that collection process is the same (ECoG recording on the same mouse line and same recording hardware).”

Q2.2. Does the section clearly define what data points represent?

0 - No or yes, but with major mistake(s)
1 - Yes, but poorly or incompletely
2 - Yes

Comment: One datapoint is labelled 4s of ECoG recording.

Q2.3. Does the section clearly define which properties of a datapoint are used as its features?

0 - No or yes, but with major mistake(s)
1 - Yes, but poorly or incompletely
2 - Yes

Comment: Feature vector of the datapoint is 800-dim vector, whose values are recorded voltage fluctuations.

Q2.4. Does the section clearly define which property of a datapoint is used as its label?

0 - No or yes, but with major mistake(s)
1 - Yes, but poorly or incompletely
2 - Yes

Comment: The labels are stages of the sleep and recording artefacts.

Q2.5. Does the section clearly indicate type of data used (continuous variable, categorical or ordinal values etc.) and units of measurement (if applicable).

0 - No or yes, but with major mistake(s)
1 - Yes, but poorly or incompletely
2 - Yes

Comment: features – real numbers, labels – categorical.

How points for submission are computed?
grading levels {0,1,2}, weight criteria 1-5=1, weight criteria 5-10=5, max points = 6

score = (2/2*1 + 1/2*1 + 2/2*1 + 2/2*1 + 2/2*1 +
1/2*5 + 2/2*5 + 2/2*5 + 2/2*5+ 2/2*5) / (5*1+5*5) = 0.9
final points = 6*0.9 = 5.4

Stage 3. ML problem formulation – MODEL and LOSS

Q3.1. 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.

0 – Models are not discussed or there is/are major mistake(s) in the discussion.
1 – Models are explained only partially or poorly.
2 – Models are explained clearly.

Comment: Two models SVM and decision tree are explained briefly.

Q3.2. Does the report clearly explain why certain the models (hypothesis spaces) were chosen (justification). It can include discussion on the assumption of the features-label relationship (linear, non-linear), computational efficiency, model's complexity, and interpretability etc.

0 – Justification is not provided or contains major mistake(s).
1 – Justification is provided, but partially or poorly.
2 – Justification is provided.

Comment: SVM – flexible, can handle high-dimensional data. DT - can handle numeric and categorical data, interpretable, computationally efficient.

Q3.3. Does the report clearly specify the loss function(s) used to evaluate the quality of a hypothesis? 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., metrics "accuracy" as the average 0/1 loss). If using several different loss functions, both should be described.

0 – The loss functions used for training and validating the ML methods are not defined or defined incorrectly.
1 - The loss functions defined but described partially or poorly.
2 - The loss function is explicitly defined.

Comment: The losses described well and references for precise formulation were provided.

Q3.4 Does the report explicitly discuss why these loss functions and metrics were chosen? For example, "The Huber loss is used as it is robust towards outliers." or "The accuracy (1/0 loss) is used as it is easier interpret in contrast to logistic loss used for parameters fitting.".

0 – Justification is not provided or contains major mistake(s).
1 – Justification is provided, but partially or poorly.
2 – Justification is provided.

Comment: Justification for SVM loss function: "This allows to maximize the margin, while incurring a penalty when a sample is misclassified or within the margin boundary", for DT: "By minimizing impurity of the node DT model can create a tree where datapoints with a same label are grouped together". Reasons to use different metrics (recall, precision) were also provided.

Q3.5 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?

0 – The construction of training and validation sets are not discussed at all or discussed incorrectly.
1 – The construction of training and validation sets are discussed superficially.
2 – The construction of training and validation sets are discussed very clearly. I would be able to reproduce this construction on my own.

Comment: Author will use shuffling and stratified sampling (train/val/test split, no cross-validation).

How points for submission are compute?
grading levels {0,1,2}, weight criteria 1-5 =1, weight criteria 5-10 =2, weight criteria 10-15 =10, max points = 14

score = (2/2*1 + 1/2*1 + 2/2*1 + 2/2*1 + 2/2*1 +
1/2*2 + 2/2*2 + 2/2*2 + 2/2*2+ 2/2*2 +
2/2*10 + 2/2*10 + 2/2*10 + 2/2*10+ 2/2*10) / (5*1+5*2+5*10) = 0.97
final points = 14*0.97 = 13.58