Feedback/Comment

Table of Contents

[Feb 10, 2023 1](#_Toc128740153)

[Mar 3, 2023 3](#_Toc128740154)

# Feb 10, 2023

**Overall**

1. Keep naming conventions consistent between modules teams.
2. **Folder structure:** consider separating drafts, ready for review, validated, … etc and make sure the folder names are self-explanatory.
3. If there are dependencies, considering numbering them in sequence.
4. **Communication:** set expectations
   1. Summarize questions in a central place (for example, in this log) so that we don’t miss any of your questions.
   2. Please be explicit and let us know which one(s) you’d like us to review. In this case, it seems that “Final Modules” would contain the ones for review or will be published?
5. Consider adding table of contents to live scripts for easier navigation.
6. Where available, text description is very well written. Great job!
7. Consider adding graphics and equations where applicable, to make modules more self-serve and visually easier to digest.

**Module\_Team1**

Please avoid punctuations in file names

**GatherData\_forvalidation\_validated\_needs\_revision**

1. You can also show import tool here as a way to bring data interactively into MATLAB or you can link documentation.
2. Yes, there is a way to webscrub in MATLAB, you can use “webread” <https://www.mathworks.com/help/matlab/ref/webread.html>
3. Rightjoin – are you asking about merging tables, if yes, you can do that <https://www.mathworks.com/help/matlab/ref/table.join.html>
4. You can include doc links
5. You can use “summary” function to show overall summary of the data

* <https://www.mathworks.com/help/matlab/ref/table.summary.html>

Other suggestions:

* Add table of contents to live scripts
* You can add additional resources links such as MATLAB Onramp, MATLAB Fundamentals etc. and point out relevant sections.
* Maybe timetable and datastore would be nice to mention
  + <https://www.mathworks.com/help/matlab/ref/timetable.html>
  + <https://www.mathworks.com/help/matlab/datastore.html>

**Clean.mlx**

* Are you planning to add any comments to this script?
* It will be also nice to show preprocessing interactively using apps or at Import Tool when you are importing data.
* Good use of logical indexing

**Module\_Team2**

**Feature\_Engineering\_Module**

* Check spelling – there are spelling mistakes
* Maybe you can link this: <https://www.mathworks.com/discovery/feature-extraction.html>
* Feature hashing plot – investigating (maybe Jianghao can comment)
* Text reflects Python (under Feature Collection and Generation) – needs revision
* Are you looking for chi square goodness of fit test? (maybe Jianghao can comment) <https://www.mathworks.com/help/stats/chi2gof.html>

**Module\_Team2\_splitdata**

<https://www.mathworks.com/help/matlab/matlab_prog/split-data-into-groups-and-calculate-statistics.html>

<https://www.mathworks.com/help/matlab/ref/matlab.io.datastore.imagedatastore.spliteachlabel.html>

* You can also set split ratios when you are training the models
* Cvpartition is good

**Module\_Team3**

Feature engineering:

**Question from the team:** “How can we get the K scores in matlab? Dr. Hedengren uses a python package, sklearn.feature\_selection to get k scores and then he plots those scores for each feature. We cannot find an equivalent for this in MATLAB.”

**Answer:** K-score is not a terminology. In this case, I think it’s choosing the top K highest scoring features. K is a number that falls in 1:total number of features. You can decide on what test you run (i.e. a scoring function) to calculate a score, and then sort() and return the indices of the K highest scores.

**Final Modules**

Stats\_Module.mlx – need a title

Visualize\_validated.mlx – need more text description/comments in the main part of the script

# Mar 3, 2023

**Overall:**

* Great work!!
* Table of content needs to come after titles (and possibly after your initial problem description)
* Okay to reference non-built-in functions as long as they are included in the repo
* I’m not able to run any code because for many, data is not available
* Many functions work with the native “table” data type, i.e. no need to covert to array for functions to work. Go through the script and refer to documentation to optimize the scripts. Some examples are provided below
* How are we doing on the overall progress?
* For each check-in meeting, would you provide an update to the files reviewed in the previous meeting?

**Imbalanced\_Validated**

1. Line 2 & 51 & 154: make sure data is included in the final submission, I suppose in the same Git repo works.
2. Plot data: scatter() and many other MATLAB functions work with the “table” data structure. Take a look and consider avoiding loops: <https://www.mathworks.com/help/matlab/ref/scatter.html>
   1. Search for “tbl” and try relevant examples
3. SMOTE: consider using this one instead: [https://www.mathworks.com/matlabcentral/fileexchange/75168-oversampling-imbalanced-data-smote-related-algorithms](https://www.mathworks.com/matlabcentral/fileexchange/75168-oversampling-imbalanced-data-smote-related-algorithms?s_tid=FX_rc2_behav)

It’s also okay to include the functions directly in your repo as long as they are properly referenced

1. Separate “Activity” from the rest of the code, so that it’s clear all other is lecture, while activity is homework (it also seems that the code here is the solution – make sure you also have a corresponding work file)

**Linear Regression Validated**

1. Two examples demonstrate multiple Python methods for (1) univariate linear regression and (2) multiple linear regression. Should be MATLAB I suppose?
2. Maybe good to break down and blend text description and code together. I will explan when we meet.
3. I like how things are broken down in the original Python code, i.e. solution file shows how you can write everything using NumPy to how you can directly call a built-in function. Please consider doing something similar here.
4. Okay to include supporting functions. Great that you include it at the end. Perhaps make it clear that they are supplementary by bolding and renaming

**Polymer\_Melt\_Flowrate\_Validated**

Great work. No major feedback.

1. Condense table of content,
2. KNN – okay to use custom function, just be sure to include the actual function in the repo
3. Feature selection
   1. Feature selection in general: <https://www.mathworks.com/help/stats/feature-selection.html>
   2. fscchi2 works with table. Have you tried that? It may solve the problem you asked in the other document.
4. fitlm works with table (line 99 – 103)

Scale data for machine learning

Great problem description and script.

1. Add table of content (after problem description)
2. Separate activity from example, and include work/solution copies for activity

**Sonar Detection**

* Script is incomplete, I will review when it’s done
* Also for case studies in general, make sure Ayan has a pass at it as she’s the chemical engineering expert.