ML Programming Challenge (live)

Start Assignment

- Due Tuesday by 18:00
- Points 18
- Submitting a file upload
- Available 1 Jan at 13:00 - 11 Mar at 18:00

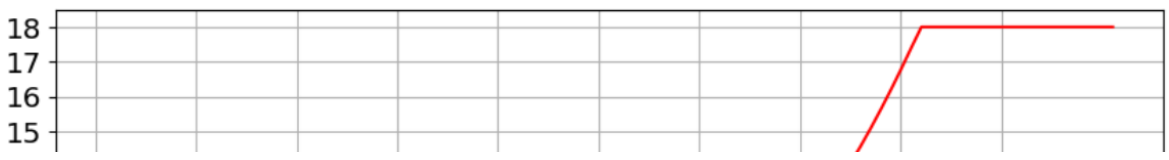
Update March 4: This is the live ML Programming Challenge. It will run for one week. See information below. Good luck! / Olov, on behalf of the teachers

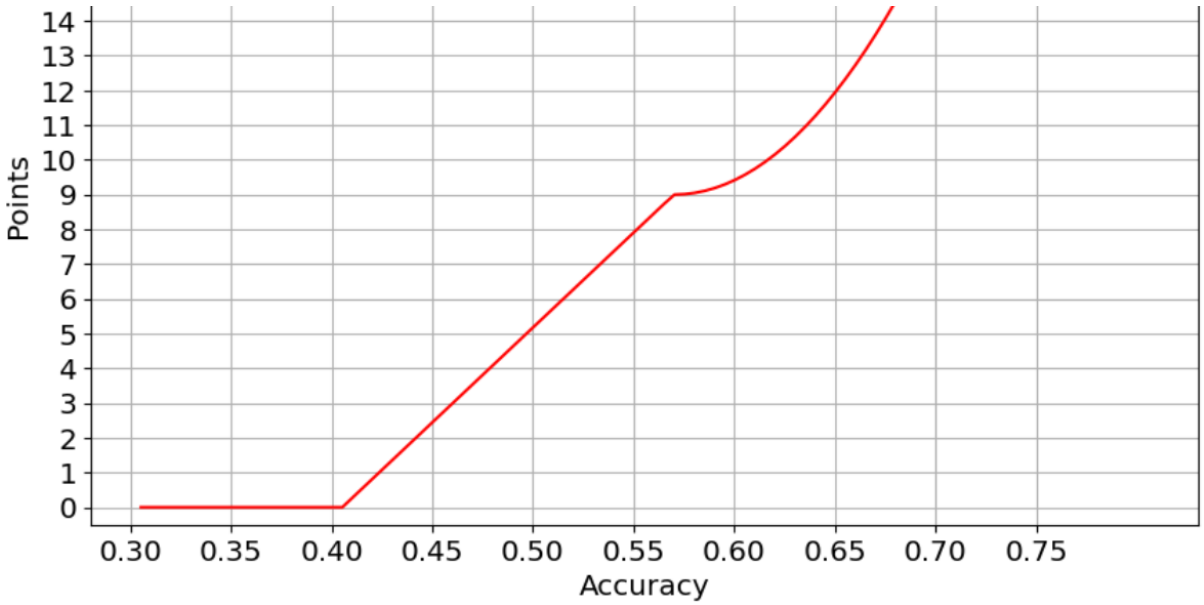
- You might notice a copy of this challenge assignment in a separate examination room that will open up soon. That one is for students re-taking the exam. These assignments will be identical and you can in theory submit to either of them, but please do not submit to both. If you submit to both challenge assignments, only your submission in the examination room will be used.
- As we mentioned during the lecture QA, there will be no update of your score in Canvas after each submission. To estimate your score, you have to instead do a train/test split on the training data much like in the real world. The final challenge score (to be added to your exam score) will be automatically calculated and added to your exam only after the exam. For those interested we will also release the correct labels then.

Challenge Instructions:

Here you must build and **train a classifier** given a labeled dataset and then use it to infer the labels of a given **unlabeled** evaluation dataset. You must then **submit the inferred labels in a specified format**, which will be compared to our evaluation data that has the correct labels (not shared, held out). This is made to mimic a real ML work assignment or challenge. If you want an unbiased estimate of how your model is performing, you can do your own train/test split using the supplied training data.

The accuracy achieved by your model on our (held out) evaluation data maps to the points you will receive for the challenge (to be added to your exam points) according to the curve below:





Some logistics:

- This challenge is to be done individually. The work you submit must be your own.
- You can use whatever programming language and libraries you want. You can use code you wrote for the labs. The challenge is designed such that it does not require high computational resources, but you can use **Google Colab (https://colab.research.google.com/)** if you feel the need.
- The training and evaluation dataset files are formatted as comma-separated values, with each line being an observation. Like real data, there may be problems with some of the entries in the training dataset file.
- You must submit two things: 1) your code (a zipfile is fine, but NO OTHER compression, e.g., rar); 2) a text file (with ".txt" as an extension) where each line is ONLY the label inferred by your system in the same order as that of the evaluation dataset file. If you do not submit these TWO files in these ways you will receive ZERO. If you send a rar compressed file you will receive a ZERO.
- If you submit only one zip file containing your code and predicted labels, you will receive ZERO.
- Be sure the labels your system generates are the same as those appearing in the training dataset. For instance, if the labels in the training dataset are {"olov", "atsuto", and "jörg"}, and the labels you predict are {"olov", "atsuto", and "jorg"}, all your predictions of "jorg" will be incorrect.
- Your label file should not contain a header or index column. If your label file has a header or index column, you will receive ZERO.
- Make sure the classifier you use in the end is the best you think you can create.
- No questions will be answered from the instructors. Ask the data!
- Use this opportunity to prepare for the exam!

The data will be published when the challenge starts, one week before the exam!

Here's the training data (variables x and example labels y): **TrainOnMe.csv (https://canvas.kth.se/courses/52576/files/9107573?wrap=1)** ⬇ **(https://canvas.kth.se/courses/52576/files/9107573/download?download_frd=1) (https://canvas.kth.se/courses/52576/files/9107572?wrap=1)**

Here's the evaluation data (only variables x, you have to infer labels y): **EvaluateOnMe.csv (https://canvas.kth.se/courses/52576/files/9107576?wrap=1)** ⬇ **(https://canvas.kth.se/courses/52576/files/9107576/download?download_frd=1)**

Here's the correct labels: To be released after the Challenge results have been posted.

Top 5 accuracy:

#1 TBD. Could be you!

#2 TBD

#3 TBD

#4 TBD

#5 TBD