Programming REchallenge

Submit Assignment

Due Monday by 0:01 **Points** 1 **Submitting** a file upload **Attempts** 0
Allowed attempts 1 **Available** 8 Dec at 0:00 - 14 Dec at 23:59 7 days

Welcome to the DD2421 ML REchallenge. In short, 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 the ground truth. The accuracy of your model is proportional to the grade you receive. (Scaling will be done according to a benchmark.)

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. 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 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 you will receive ZERO.
- The labels your system generates must be exactly the same as those appearing in the training dataset. If you do not do this you will receive ZERO.
- 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.
- You can only submit ONCE, so 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 take-home exam!

Here's the training data: **TrainOnMe_RE.csv**

Here's the evaluation data: **EvaluateOnMe_RE.csv**

Here's how the data was generated and the theoretical maximum accuracy calculated: