Kaggle Competition 2 IFT3395/6390
Instructions for IFT6390 students
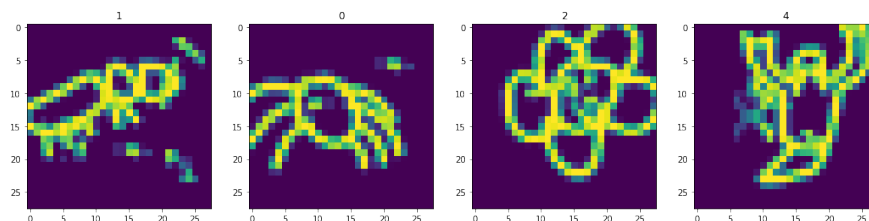
November 10, 2020

Figure 1: Example images from the training set

1 Background

For this project, you will take part in a 2nd Kaggle competition based on image classification. The goal is to design a machine learning algorithm that can automatically sort handdrawn objects into a pre-determined set of 6 categories. The dataset that we have prepared contains a subset of images extracted from the QuickDraw dataset. We have sampled 250 images from each category to be used as the training set, and 60,000 images as test set. You will implement and train a few different classifiers and will be evaluated on the test accuracy that your trained models achieve.

The competition, including the data, is available here: `https://www.kaggle.com/c/ift3395-6390-quickdraw`, but in order to access it directly, you will first need to join the competition using the link `https://www.kaggle.com/t/305a6210469244a5ae84fe38c928f61b`.

2 Important dates and information

- **Nov 21st 23:59** Deadline to enter the competition on Kaggle, form teams, and fill the Google form.

- **Dec 17th 23:59** Competition ends. No more Kaggle submissions are allowed.

- **Dec 21th 23:59** Reports and code are due on Gradescope .

Note on sharing and plagiarism: You are allowed to discuss general techniques with other teams. You are NOT allowed to share any of your code. This behavior constitutes plagiarism and it is very easy to detect. All teams involved in sharing code will receive a grade of 0 in the data competition."

3   Enter the competition and form teams

IFT6390 students must do the competition in **teams of 2**. Note that IFT3395 students can optionally participate in the competition, this is why there will also be teams of 3 in the leaderboard.

- Enter the competition by following the link:
  `https://www.kaggle.com/t/305a6210469244a5ae84fe38c928f61b`

- In the "Invite Others" section, enter your teammates' names, or team name.

- Your teammate has the option to accept your merge.

- Fill out the google form `https://forms.gle/B26SmEBTdxnsEnEi9` with your team information by **Nov 21st 23:59**. Any teams not registered or registered late will not be graded.

Important note: The maximum amount of submissions is 2 per day, per TEAM. Any team whose individual members have a submission count larger than what is allowed up to-date will be UNABLE to form a team. Example: Today is the first day of competition. A,B,C are three teammates who haven't formed a team yet.

- A submitted 0 times.

- B submitted 2 times.

- C submitted 1 time.

Because the maximum amount of submissions is 2 per team per day, the total possible submissions for a team is 2. However, the cumulative submission count for A,B,C is 3. Therefore, they will be unable to form a team (They will need to wait for tomorrow, and not submit any submissions for the next day).

4   Requirements

For this competition, you are allowed to use any method of your choice. To get full grades, you are asked to experiment with **at least 3 different algorithms**, and report their performances. You can implement any method of your choice, or use any library, either discussed during the course, or from your own researches. However, even if you use a library, it is important that you discuss each hyperparameter, and how you chose their values.

Here are some possibilities of methods:

- Kernelized SVM using image kernels

- Random Forests or other ensemble methods

- Hand-crafted features and logistic regression

- Neural networks

- any other algorithm of your choice...

The goal is to design the best performing method as measured by submitting predictions for the test set on Kaggle. Your final performance on Kaggle will count as a criterion for evaluation (see below), as well as the number of baselines that you beat. If a tested model does not perform well, you can still add it in your report and explain why you think it is not appropriate for this task. This kind of discussion is an important feature that we will be using to evaluate your final competition report.

5   Report

In addition to your methods, you must write up a report that details the preprocessing, validation, algorithmic, and optimization techniques, as well as providing results that help you compare different methods/models. The report should contain the following sections and elements. You will lose points for not following these guidelines.

- Project title

- Team name on Kaggle, as well as the list of team members, including their full name and student number.

- Introduction: briefly describe the problem and summarize your approach and results.

- Feature Design: Describe and justify your pre-processing methods, and how you designed and selected your features.

- Algorithms: Give an overview of the learning algorithms used without going into too much detail, unless necessary to understand other details.

- Methodology: Include any decisions about training/validation split, regularization strategy, any optimization tricks, setting hyper-parameters, etc.

- Results: Present a detailed analysis of your results, including graphs and tables as appropriate. This analysis should be broader than just the Kaggle result: include a short comparison of the most important hyperparameters and all methods (at least 3) you implemented.

- Discussion: Discuss the pros/cons of your approach & methodology and suggest ideas for improvement.

- Statement of Contributions. Briefly describe the contributions of each team member towards each of the components of the project (e.g. defining the problem, developing the methodology, coding the solution, performing the data analysis, writing the report, etc.) At the end of the Statement of Contributions, add the following statement: We hereby state that all the work presented in this report is that of the authors.

- References (very important if you use ideas and methods that you found in some paper or online; it is a matter of academic integrity).

- Appendix (optional). Here you can include additional results, more details of the methods, etc.

The main text of the report should not exceed 6 pages. References and appendix can be in excess of the 6 pages.

Your must submit your report and your code on Gradescope before **Dec 21th 23:59**.

Submission Instructions

- You must submit the code developed during the project. The code must be well-documented. The code should include a README file containing instructions on how to run the code.

- The prediction file containing your predictions on the test set must be submitted online at the Kaggle website.

- The report in pdf format (written according to the general layout described above) and the code should be uploaded on Gradescope.

6 Evaluation Criteria

Marks will be attributed based on the following criteria:

1. You will be assigned points if you beat the baseline.

2. You will be assigned points depending on your final performance at the end of the competition, given by your ranking in the private leaderboard.

3. You will be assigned points depending on the quality and technical soundness of your final report (see above).