Final Project Instructions

Project Overview

In this project we will have two challenges, you will do both challenges. They will both involve learning from limited data. Note a big part of this assignment will be properly reporting your approaches and development stages as well as a literature review for the graduate student section. You can also propose an alternative project (see instructions at the last page).

Leaderboard/Evaluation

The first challenge of the project will have a leaderboard which will be available later in the semester. If you want to get started before this you should already have everything you need to get going in the preliminary testbed that will be provided as a colab notebook. For the final submission you will include a script that reproduces the results for your best performing method

Teams

You will work in teams of 1-3. Exceptions will be considered case by case. Your final report should include a discussion of what each team member contributed. The number of team members will be taken into account for the report.

Challenge 1: Learning from Limited data

Few-sample learning on the CIFAR-10 dataset images. We will consider a dataset with only 25 samples per class and 2 object classes. For testing a larger dataset will be used (e.g. the CIFAR-10 test set).

Your job will be to make an algorithm that will train on 2 randomly selected classes from a CIFAR train set with 50 samples. You may not use any model trained on external data or any external data for your answer. On the other hand ***you may consider any kind of approach without constraint including non-deep learning methods, combinations of feature extraction, etc as long as it only utilizes the 50 examples for training.***

Final evaluation: In the final evaluation your model will be run on 10 random instances of this problem.

Note:

1. In the initial testbed there is a random development set selected to allow you to measure your progress, note that your final training script will be trained with just 10 samples, thus you cannot use the development set

2. You should take training time into consideration. It is recommended to limit the training time of your models to 2 hours for 5 seeds using a GPU.
3. You will need to submit a final script to reproduce your results in a specific format that will be provided close to the deadline
4. Note the variance in accuracy will be high in this setting, consider this in your evaluation (run on multiple subsets before concluding a certain approach is better over another)

Challenge 2: Learning with Limited Data w/ External Data or Pre-trained models:

This challenge will not be on the leaderboard.

Consider the same exercise but now with the ability to use some external data or models ***not trained on this same data.*** You may consider these external data to do for example various forms of finetuning, meta-learning, semi-supervised learning with an external dataset and other related ideas.

In order to bound the scope of this exercise these are the initial set of possible external sources you can use:

- Models pre-trained on imagenet in the pytorch model repository (see example in testbed)
- Imagenet, MNIST, STL-10, CUB, SCENES datasets

You are encouraged to propose other data sources to use, however you should clear them with the instructor or TA first.

Note:

5. There is no right answer and many possible approaches are reasonable, same for Challenge 1, but more so here.
6. If you are unsure whether an idea violates the intended constraint you may ask for clarification from the TA and instructor. For example, trying to find the same data categories from another dataset and augment the training set in this way would violate the spirit of the exercise. I may add further restrictions on the constraints for Challenge 2 as it evolves.
7. You are free to pre-train your own models on the allowed data sources

Requirements/Evaluation

Each team will submit code to reproduce their best working solution evaluated on 5 randomly selected problem instances and the details of the results obtained.

Performance/Quality	20%

Report (excluding literature review)	30%
Literature Review	30%
Originality	20%

Performance/Quality

Performance will take into account the overall accuracy of your model. This will also consider inference and training time. If you have a model that is excessively long in terms of inference or training time you may also submit a 2nd model for final evaluation and a discussion of the tradeoffs. The quality of your solution overall will be taken into account (e.g. using appropriate methods, exceeding some basic thresholds).

Report / Research Quality (excluding literature review)

1. Report should include details of all the methods/strategies you have tried and discussion of why they were appropriate, what were the results in each method in your validation framework.

2. Evaluate a minimum of 2 high level methods for Challenge #1. Add 1 method for teams over 2. For example if you have 3 team members you will need to try 3 methods. These are however minimums and you may study as many approaches as you like.

3. Evaluate a minimum of 1 high level methods for Challenge #2. Add 1 method for teams greater than 2. For example if you have 3 team members you will need to try 2 methods. These are however minimums and you may study as many approaches as you like.

4. Methods should be explored in detail trying various appropriate hyperparameters and improvements. For example a baseline method for challenge 1 may be to use a standard CNN model with cross-entropy loss. Exploring using different sizes of the same type of model, training schedules, would be counted as 1 method but explored in detail can yield potentially interesting results. Another method could for example use a specialized network architecture and regularization method. Another approach can study non-deep learning methods such as gradient boosting or logistic regression. The distinction of what counts as a “method” can be ambiguous; you can ask the instructor/TA if you are unsure. Overall the goal is to explore in detail approaches you take and to explore more than one general approach. Furthermore it is not a requirement to explore each high level direction in the same depth, you may choose to focus your efforts on a single approach, refining the strategy.

5. In general the size of the team will be taken into consideration when evaluating the depth and breadth of approaches studied.

6. Provide details of how you validated your model, hyperparameters etc.

7. You will be evaluated on the appropriateness and presentation (Clarity of presentation such as clearly marked/visible figures etc will be considered).

8. Include discussion of your observations on tradeoffs in training time, inference time, and performance.

Literature Review

Provide a literature review including an ***in-depth review of at least 1 relevant published work per student team member related to the problems posed,*** even if these do not make it to the final solution, these should represent methods or approaches which were evaluated as part of the report. The discussions should be detailed describing the relevant work, why it is relevant, any prior work that places it in context. Indeed it is expected your literature review will briefly discuss multiple works, but document some (based on team size) in detail. The discussion of each of the main papers should be around a half page to page per work. Discussed works should be framed in the context of your report as well, for example if it is a method you chose not to use you should describe why. Note the literature review must be for published papers at journals or conferences. You may also discuss unpublished methods or techniques you have found discussions about in online forums and which you used in your approach, but these do not count in the 1 work per team member total.

It is preferred that some of your selected papers for literature review are recent (published/released within the last 6 years). If your team has more than 1 team member then at least one paper must be recent.

Originality

To gain full credit for originality you should at a minimum apply methods not previously used for this exact problem and dataset or modify and attempt to expand existing methods that you find used for this and related problems. For example, if you find a work doing the exact same problem you can gain full credit for originality by exploring aspects they did not discuss (e.g. different model architectures, training approaches, loss functions, and regularizers). If you find a work applying an approach to a different problem and adapt it to this one this also fulfills the originality criteria. Undergraduate teams will be evaluated less rigidly on this part. Finally, you are free to and encouraged to use source code from other projects as a basis for implementing your methods but it should be cited in your report.

Computing resources

Colab should be sufficient to prototype your results for Challenge 1. If you need to run longer than colab permits for challenge 2 you can try to use the lab machines from our lab which have a GPU.

Alternative Project (must be pre-approved):

You may choose to do a research project of your choosing in deep learning. This option is largely intended for students already doing research in related areas applying deep learning methods or where deep learning methods make sense to be applied. The expectations in novelty will be slightly higher than for the default project and you should not reuse research projects you have already finished.

The proposal must consist of an original research project in deep learning or its applications. It should involve a literature review including an in-depth discussion of at least 1 paper per student as well as numerical experiments. If you are interested in pursuing this option you must submit a (maximum) half-page initial proposal by **February 25 by email.** You can also come discuss this during the office hours. Where appropriate you can also include some brief information about your ongoing research if it is needed as context for the proposal. An example of the expectations for a project http://cs231n.stanford.edu/2017/reports.html , since you do not have the full semester for this project the expectation is a bit below this but that should serve as a reference point for the kind of project proposals which are expected for this alternative.