Final Project Instructions

Project Overview

In this project we will have two 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.

Leaderboard/Evaluation

The project may have a leaderboard running on the codalab platform which will allow you to submit your code that trains and outputs results. Due to technical issues the leaderboard is not setup yet but you should already have everything you need to get going in the preliminary testbed that will be provided as a jupyter notebook. It is possible that we will not be able to use codalab in the end due to too many constraints on the training time of the models and not enough setup support. In that case I will provide towards the end of the semester a final format of your training and testing scripts and a final testing script and dataset that you will run yourself for your best performing model and submit the outputs of along with the code.

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

Challenge 1: Learning from Limited data

Few-sample learning on the CIFAR-10 dataset. We will consider a dataset with only 100 samples. For testing a larger dataset will be used (e.g. the CIFAR-10 test set).

Your job will be to train on 100 randomly selected (but class balanced) samples from the CIFAR-10 training set. 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 as long as it only utilizes the 100 examples for training.***

Final evaluation: In the final evaluation your model will be run on 5 random instances of this problem. Evaluation will occur both on the CIFAR-10 test set and other data which will not be

revealed at the moment. It may also be run on a similar drop in dataset with the same size images and number of output classes.

Note:

1. In the initial testbed there is a random validation set selected to allow you to measure your progress, note that your final training script will be trained on 100 samples and there will not be separately provided validation data.
2. You should try to limit the training time of your models to 2 hours using a colab GPU instance. This is not a hard limit however but may make it challenging to put the models up on codalab for comparisons. If you choose to have a submission which uses extensive training time (for example ensembling of multiple methods or the like) this is not explicitly prohibited but in this case also submit an alternative solution that respects reasonable training time constraints.
3. Note the variance in acc 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:

Consider the same exercise but now with the ability to use an external data or models ***not trained on CIFAR-10*** from the pytorch model repository or to train your own model on external data (you may not train a model on CIFAR-10 dataset or any derivative using these images. You may consider for example various forms of finetuning, meta-learning, semi-supervised learning with an external dataset and other related ideas.

Note:

4. You should try to limit the training time of your models to 2 hours using a colab GPU instance. This is not a hard limit however but may make it challenging to put the models up on codalab for comparisons. If you choose to have a submission which uses
5. If you would like to train custom models (potentially with custom loss functions) on external datasets this is allowed.
6. If you are unsure whether an idea violates the intended constraint you may ask for clarification. 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.

Requirements/Evaluation

Each team will submit their best working solution evaluated on the final testing sets. Note the final evaluation will train on 100 samples from a different set of data than the cifar-10 training set and test on a different set of data as well. If there is a codalab leaderboard you will be required to submit there as well. You will be evaluated as follows

	COMP 691	COMP 499
Performance	20%	20%
Report (excluding literature review)	30%	60%
Literature Review	30%	10%
Originality	20%	10%

Performance

Performance will take into account the overall accuracy of your model both on the CIFAR-10 test set and any potential additional data I may use for evaluation. 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. You will receive a portion of this credit for having a model above a certain baseline performance threshold.

Report (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 for each problem. 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 for each problem. 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 or use gradient boosting. The distinction of what counts as a "method" can be ambiguous; you can ask the instructor if you are unsure or submit a proposal of a maximum half page. 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

-- Graduate student teams - should provide a literature review including an ***in-depth review of at least 1 relevant published work per graduate 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.

-- Undergraduate teams - You may provide a short discussion (of minimum 1 page total) of where you sourced the methods you used and why you chose them. Your discussions of relevant methods can include journals or conferences but can also instead include blog posts, articles, kaggle forums, etc. In general it should document where you got your ideas.

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 fulfils 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.

(extra credit) Original Method (Max 15 points)

You may also gain additional credit beyond the 100% if you indicate in your report a description of a solution that was not necessarily your best performer but uses an original method, if it is your best performing model even better. Note the solution must have a certain minimum performance to be further evaluated. Examples of original solutions that may receive some of these additional points:

- New idea as backed up by a literature review
- Novel modification of an existing method
- Implementing a method from a paper which has not been explicitly used in a similar context
- Novel observations about the behavior of specific pre-trained models

Note the maximum grade for extra credit here is reserved for original research ideas. Only one such solution will be evaluated so indicate it clearly in your report.

Computing resources

I have reserved some computing resources for this project however they are limited and in many cases colab will likely be easier to use, particularly for Challenge 1. The computing resources

available are individual machines with a GTX 980 and 4GB of GPU RAM. These may however be useful if you want to train models for longer periods or create new pre-trained models on external datasets. There are only 8 of these machines available. If your team will need to use these please contact me and highlight why you will need to go beyond colab I will consider the requests

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 higher than for the default project and you should not reuse research projects you have already done.

The proposal must consist of an original research project in deep learning or it's applications. It should involve a literature review including an in-depth discussion of at least 3 papers as well as numerical experiments. If you are interested to pursue this option you must submit a (maximum) half-page initial proposal by **March 8**. 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.