Dynamic Sample Selection Strategy in Active Learning

# Introduction

Active learning is a special case of machine learning which gives the learning algorithm some control over the learning process, by letting it actively decide which examples it will learn from. More precisely, given a pool of unlabeled example the algorithm decides which one to label and to feed into the learning process. Active learning is most useful when labeling examples is costly, because then smart selection of which example to label can save resources and effort. There are few algorithms for active learning in neural networks, which we note in the related work section. Most of those algorithms do not integrate the learning procedure of the neural network with the active learning. This fact can cause the active algorithm to be efficient in the beginning of the process but not in later stages or vice versa, depending on the algorithm. For example, we can think of an active learning algorithm for SVM for classification of several classes. In the beginning it might choose examples in places in the space it has no information on, e.g. far from examples with known classes. But as the classification progresses, we might want the algorithm to choose examples with the smallest margin (i.e. closest to the decision boundary) and thus be more accurate. To perform such a task, traditionally we would need to set hyperparameters that will determine the algorithm's choice of the next example to classify – low-information areas vs. more accuracy in known areas. SelectiveNet‎ (Geifman & El-Yaniv, 2018) is a neural network architecture used for classifying images or for regression, which has an additional output. That output states the “confidence level” in the classification (or regression) of the example.

In this project we will try to use this mechanism to determine which example to label next, by training the net on labeled examples and using it on examples that are not labeled. Then we will choose the examples (from the unlabeled ones) which have the smallest “confidence level” in their prediction by the current classifier, to be labeled next. By using the mechanism of SelectiveNet, we let the neural network decide where it has the least information to determine a classification. Our hypothesis is that the decision will be adaptive to the stage of the active learning, and will achieve better results through the process, without the need to set any hyperparameters manually.

# Theoretical Justification

In the paper, Active Learning via Perfect Selective Classification (Ran El-Yaniv & Yair Wiener), reduction was shown between active learning to perfect selective classification. Where perfect selective classifier is a selective classifier with zero loss on the covered labels. In particular, it showed in the paper that if for hypothesis class there is a perfect selective classification with a fast rejection rate of O(polylog(m)/m), then CAL algorithm can choose examples for active learning such that it will learn the hypothesis class with exponential label complexity rate of O(polylog(1/ε)).

More recent results related to the connection of selective classifier and active learning was showed in (Gelbhart & El-Yaniv, 2019). This work discusses a selective classification scheme which makes fewer assumptions than previous results and applies this scheme to developing an active learning algorithm with exponential speedup.

# Related Work

1. SelectiveNet: A Deep Neural Network with an Integrated Reject Option – ***Yonatan Geifman****,* ***Ran El-Yaniv***; arXiv:1901.09192
2. Online Choice of Active Learning Algorithms – ***Yoram Baram****,* ***Ran El Yaniv****,* ***Kobi Luz***; J. Mach. Learn. Res. 5 (December 2004), 255-291. *A description of an approach to arbitrating between several algorithms for selecting the next sample in active learning*
3. Deep Active Learning over the Long Tail – ***Yonatan Geifman****,* ***Ran El-Yaniv***; arXiv:1711.00941. *An example for using Furthest-First (FF) algorithm for selecting the next sample*
4. Deep Bayesian Active Learning with Image Data – ***Yarin Gal****,* ***Riashat Islam****,* ***Zoubin Ghahramani***; arXiv:1703.02910. *Using dropout in order to approximate a NN bayesian model and then using inference to determine whether an example encapsulates new information. Example is considered as encapsulating new information if the model is uncertain about its label on average*
5. The Relationship Between Agnostic Selective Classification Active Learning and the Disagreement Coefficient – ***Roei Gelbhart****,* ***Ran El-Yaniv***; arXiv:1703.06536. *Discussing the applicability of improved selective classification algorithms for active learning*
6. Active Learning via Perfect Selective Classification - **Ran El-Yaniv & Yair Wiener**. Discussing the connection between active learning and perfect selective classifier.

# Problem Definition

The goal in active learning is to design and analyze learning algorithms that can effectively choose the samples for which they ask the teacher for a label. Utilizing this technique allows reaching high quality results while maintaining low data requirements. This is relevant in cases where data is expensive, scarce, cumbersome (as in hi-res images), and also when data is too plentiful, making training unnecessarily long when the same results could be achieved faster by choosing which datapoints to train on as part of the training process.

The algorithm used for choosing which samples to request the real label for is therefore a major parameter in the performance of the neural network. Additionally, many of the commonly used algorithms have varied performance across different learning problems, such as those with a "XOR"-like structure2. Estimating the performance of the selection algorithm or formulating a supervising algorithm to choose between several such algorithms, could be a place where development could lead to great improvements in accuracy and consistency across different learning tasks.

# Project Goals

Reduce training set size while preserving performance. Our main goal is to show that using active learning via SelectiveNet will achieve superior results than other active learning algorithms of neural networks for classification. The second goal is to show shifts in the type of examples chosen to label from areas with no information to higher accuracy areas. The third goal is to check whether adding another loss function related to information gain will reach better results (such as the COMB algorithm employed by Baram et al2.

# Project Plan

31/5/19 – working code for active learning for SelectiveNet.

14/6/19 – comparison between our baseline method to other active learning algorithms (graphs not only data).

30/6/19 – first version of improved active learning based on SelectiveNet and comparison to other algorithms.

14/7/19 - second version of improved active learning based on SelectiveNet and comparison to other algorithms.

28/7/19 – finished project paper.