

FINAL PROJECT PROPOSAL: [IFT6390]

MOHAMMAD BARDESTANI, JONATHAN GUYMONT, MARZI MEHDIZAD, LÉA
RICARD, JEFF SYLVESTRE-DÉCARY & JOSEPH D. VIVIANO

We propose to use two data sets for classification problems:

- **Wine Quality:** Using some attributes about a bunch of portugese wines, predict whether these wines are rated by experts as poor, normal, or excellent (3 class classification).
- **Drug Consumption:** Using demographic and personality measurements, predict whether each respondant has ever used each of 7 drugs (7 class classification).

And the following algorithms, all in the context of **adaboost**:

- MLP
- Linear SVM
- Decision Trees

0.1. General Idea. We want to investigate the difference between boosting using more weak learners versus boosting a smaller number of stronger learners. Specifically, we want to know if using a larger number of learners with lower individual capacity produces an ensemble with a different total capacity than an ensemble with a smaller number of learners with higher individual capacity.

0.2. Experiments Proposed. For all models, the number of learners trained in the context of adaboost will be manipulated. For larger ensembles, more regularization will be used on the individual weak learner models.

For MLP, we will manipulate the number of units per hidden layer and the strength of L1 and L2 regularization. These experiments will likely take the most time, and be the most interesting, as capacity in indivudal MLPs can be quite large. For Linear SVM, we will manipulate the hyperparameter C (number of allowed misclassified examples). For decision trees, we will manipulate the maximum tree depth.

We will report the results on training, validation, and test sets, using the notion of generalization gap between the validation set and test set (difference between classification loss on validation and test set) as proxy for poor, optimal, or too much model capacity. We will also keep track of the size of the ensemble model (in terms of number of total parameters and memory usage) to establish whether all models of similar capacity utilise the same resources.

UNIVERSITÉ DE MONTRÉAL

Email address: joseph@viviano.ca, jeffsd05@gmail.com, jonathan.guymont@umontreal.ca, lea.ricard@umontreal.ca

Date: Nov 12th 2018.