

# Assignment 2: Adaboost

Moaz Mohamed  
The University of Adelaide  
Adelaide SA 5005

a1779177@student.adelaide.edu.au

## 1. Introduction

Over the last few years, Machine learning had a significant impact in the current technological landscape. The application of numerous Deep learning algorithms have been implemented in different sectors. In this paper Adaboost algorithm will be explored and its performance will be investigated on classification task on the provided dataset. In addition to a complete review of the Adaboost algorithm describing its strength and weakness.

## 2. Methodology and Background

### 2.1. AdaBoost

Boosting is the idea of utilizing the a number of weak learner and combine them to make a strong learner that is capable of reaching a training error of zero, hence a weak learner alone is able to reach zero training error. instead of using a strong learner like Support Vector machine (SVM) adaboost is an ensemble of weighted weak learners Eq (1).

$$H(x) = \text{sign} \left( \sum_{t=1}^T \alpha_t h_t(x) \right) \quad (1)$$

Some learners do perform better than others, hence the weighting is depended on the accuracy of the learner.

$$\alpha_t = \frac{1}{2} \ln \left( \frac{1 - \epsilon_t}{\epsilon_t} \right) \quad (2)$$

In every new iteration a new weak learner is used to minimize the error of Eq 3. in addition to a weight matrix for the contribution of error for every data point. which increase and decreases for every data point per iteration Eq 4

$$\epsilon_t = \sum_{i=1}^n \mathbf{1} [h_t(x_i) \neq y_i] w_i^{(t)} \quad (3)$$

$$w_i^{(t+1)} = w_i^{(t)} \cdot e^{-\alpha_t y_i h_t(x_i)} \quad (4)$$

### 2.2. Weak Learner

A weak learner can be any learner that is better than a coin tossing. Decision stumps have been used as a weak learner in most Adaboost implementations. Decision stumps can be treated as decision trees with max depth of 1, increasing the depth of the tree can be used to study the affect of having stronger learner in Adaboost algorithm. Decision stumps can also be implemented as using greedy search to test different unique values in each feature in the dataset to come up with the best threshold for the split. in this project sklearn implementation has been used for ease of use and the added benefit of using optimized code to run multiple runs of experiments efficiently.

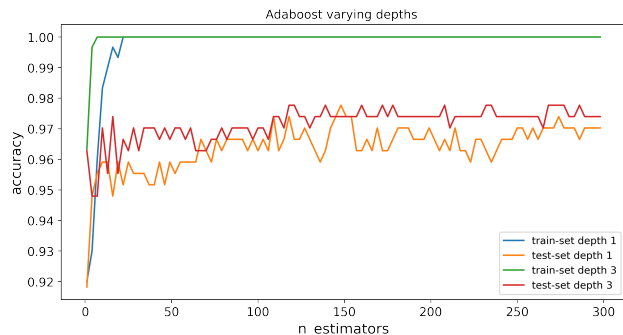
## 3. Experimental Analysis.

### 3.1. Training and Testing Curves

in figure 1. Adaboost reaches a high accuracy in relatively small number of iteration. that is due to the weighting in Eq 2. As high weighting is applied to the first learners especially with the learners with low error or loss and as the boosting continues  $\alpha_t$  gets smaller and smaller as the contribution of additional learners gets smaller and hence why Adaboost tends not to overfit easily. even with increasing the weak learner ( decision stumps ) depth to 3. the training accuracy increases faster than depth of 1 while not showing any signs of overfitting. hence Adaboost does fix the successfully combines classifiers with high bias ( decision stumps ) and at the same time has an inbuilt feature of combating overfitting.

### 3.2. Performance on provided data-set

In table 1 Adaboost does outperform SVM without and with no regularization. there are certain advantageous that goes Adaboost an edge. Adaboost makes no assumptions about the data but in SVM since the objective is finding the largest margin. hence feature vectors with high values can hinder SVM ability to find the margin while in Adaboost that isn't a problem. the mentioned Adaboost algorithm is



A

weak learner can be any learner that is better than a coin tossing. Decision stumps have been used as a weak learner in most Adaboost implementations. Decision stumps can be treated as decision trees with max depth of 1, increasing the depth of the tree can be used to study the affect of having stronger learner in Adaboost algorithm. Decision stumps can also be implemented as using greedy search to test different unique values in each feature in the dataset to come up with the best threshold for the split

Figure 1: Training vs testing

sensitive to noise. as in Eq 4 for noisy datasets the weight for the misclassified data points will keep increasing and new learner will try to classify such points correctly but in SVM regularization parameter can be adjusted to cater for the inherit noise in the dataset.

Algorithm	accuracy
Adaboost scratch	96.282%
Adaboost sklearn	96.654%
SVM-C=0	95.539%
SVM-C=1000	97.026%

Table 1: SVM vs Adaboost

## 4. Conclusion

From bias and variance prospective. Adaboost can be seen as method to take an ensemble of high bias learners combine them in such way to decrease the bias and increase the variance by adding learners and yet not overfit easily. Of-course no algorithm is perfect. due to Eq 4 Adaboost will have such emphases on the misclassified data points and for each iteration will try to adjust the decision boundary to correctly classify every data point correctly.