

# ADABOOST Report

## Introduction to Adaboost and algorithm used:

Adaboost is defined as a booster and it works in a number of rounds. At round the 't' booster defines a distribution of weights over training examples, and the weak learner at 't' iteration produces a weak hypothesis this implementation can be defined using an algorithm.

### Algorithm:

input:

training set  $S = (x_1, y_1), \dots, (x_m, y_m)$

weak learner WL

number of rounds T

initialize  $D(1) = (1/m, \dots, 1/m)$ .

for  $t = 1, \dots, T$ :

invoke weak learner  $h_t = WL(D(t), S)$

compute  $error\_t = \sum_{i=1}^m D_i(t) * 1[y_i \neq h_t(x_i)]$

let  $weight\_t = 1/2 * \log(1 - error\_t / error\_t)$

update  $D(t+1)_i = D(t)_i * \exp(-weight\_t * y_i * h_t(x_i)) / \sum_{j=1}^m D(t)_j * \exp(-weight\_t * y_j * h_t(x_j))$  for all  $i = 1, \dots, m$

output the hypothesis  $h_s(x) = \text{sign}(\sum_{t=1}^T weight\_t * h_t(x))$ .

By using the above defined algorithm, to calculate the error and update the weight, weak learners can be used in adaboost to produce a much better hypothesis.

## Implementation of Code:

1. In the code defined I have defined a class which contains the different methods used for fitting the model and predicting the output of the defined model.
2. Used Decision tree classifiers as weak classifiers at each iteration to create appropriate decision stumps.
3. Created another function named plot\_adaboost to plot the decision boundary of the adaboost classifier.
4. Used make\_gaussian\_quantiles to create a random data set of defined data points.
5. **For different values of data points change the value of variable 'n' and for different numbers of weak classifiers change the value of variable 'i'**
6. For the comparison with different values of number of iterations with constant data points I have calculated the minimum number of iterations or weak classifiers needed to reach the minimum performance error for both training set and test set.

## Observation:

1. For a constant Set of data points, As the number of data points increases the number of iterations increases.

2. It can also be observed that in some cases the minimum number of iterations or weak classifiers for training set is greater than that of the test set while in some cases it's the opposite.