Using Adaboost Algorithm to Enhance the Prediction Accuracy of Decision Trees

Outline

1. Adaboost Algorithm
2. The Dataset
3. Data Exploration and Visualization
4. Prediction Using a Single Decision Tree
5. Prediction Using with Random Forest
6. Prediction Using Adaboost
7. Can we improve even further?
8. Comparison
9. References

Bagging (Bootstrap aggregating)-Random Forest

- Take M bootstrap samples (with replacement)
- Train M different classifiers on these bootstrap samples
- For a new query, let all classifiers predict and take an average (or majority vote)

- If the classifiers make independent errors, then their ensemble can improve performance.

- Stated differently: the variance in the prediction is reduced (we don’t suffer from the random errors that a single classifier is bound to make).

Random Forest

- Ensemble consisting of a bagging of un-pruned decision tree learners with a randomized selection of features at each split.

- Grow many trees on datasets sampled from the original dataset with replacement (a bootstrap sample).
    - Draw K bootstrap samples of a fixed size
    - Grow a DT, randomly sampling a few attributes/dimensions to split on at each internal node

- Average the predictions of the trees for a new query (or take majority vote)

- Random Forests are state of the art classifiers!

Boosting

Train classifiers (e.g. decision trees) in a sequence.
A new classifier should focus on those cases which were incorrectly classified in the last round.
Combine the classifiers by letting them vote on the final prediction (like bagging).
Each classifier is “weak” but the ensemble is “strong.”
AdaBoost is a specific boosting method.

Adaboost Algorithm

- Adaboost is short for “Adaptive Boosting”
- Created by Yoav Freund and Robert Schapire in 1997
- Aaboost is a meta-algorithm (provides a sufficiently good solution to an optimization problem)
- Can be used with other learning algorithms to improve their performance
  - Makes the weak classifiers strong by combining them
  - Typically, decision trees (which we’ll use in this project)

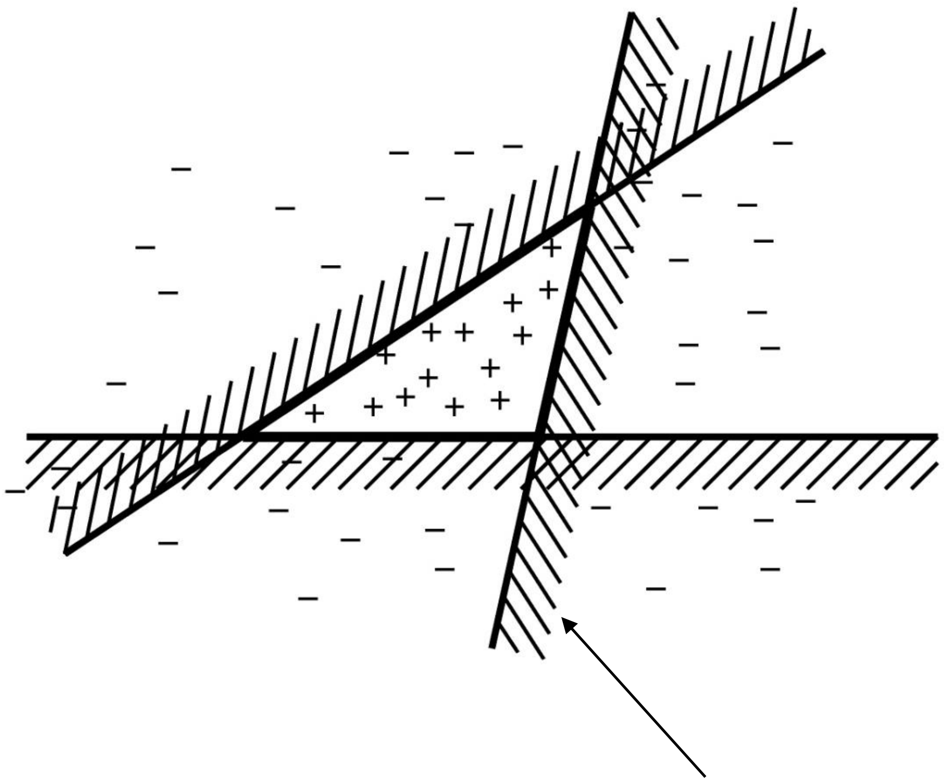
Adaboost Algorithm (cont.)

- Weak models are added sequentially after being trained on the training data
- The process continues until:
  - No further improvement can be made on the training dataset
  - A threshold number of weak classifiers is created
- Predication are then made by taking the weighted average of the weak classifiers
  - Each weak classifier has a weight (its classification result is multiplied by its weight)

Adaboost Algorithm (cont.)

- It is better to have these characteristics in the data before processing it with adaboost:
  - **Quality Data**:
    - Ensemble method attempts to correct misclassifications in the training data
    - Make sure that the training data is of high-quality
  - **Outliers**:
    - Outliers will force the ensemble to work very hard in order to correct for cases that are unrealistic
    - Best to remove them from dataset
  - **Noisy Data**:
    - Noise in the output variable can be problematic for ensemble
    - Isolate and clean the noisy data from your training dataset

Example



This line is one simple classifier saying that everything to the left + and everything to the right is -

Boosting Intuition

We adaptively weigh each data case.

Data cases which are wrongly classified get high weight (the algorithm will focus on them)

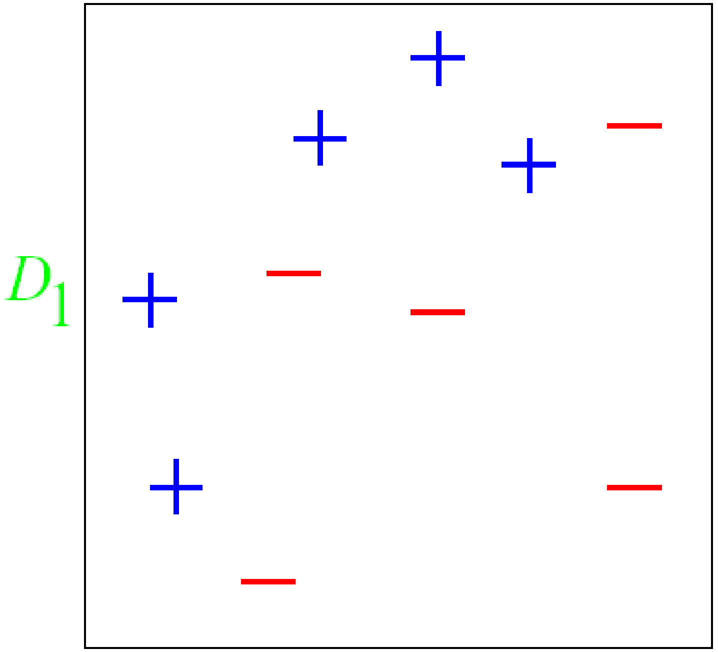
Each boosting round learns a new (simple) classifier on the weighed dataset.

These classifiers are weighed to combine them into a single powerful classifier.

Classifiers that that obtain low training error rate have high weight.

We stop by using monitoring a hold out set (cross-validation).

And in animation

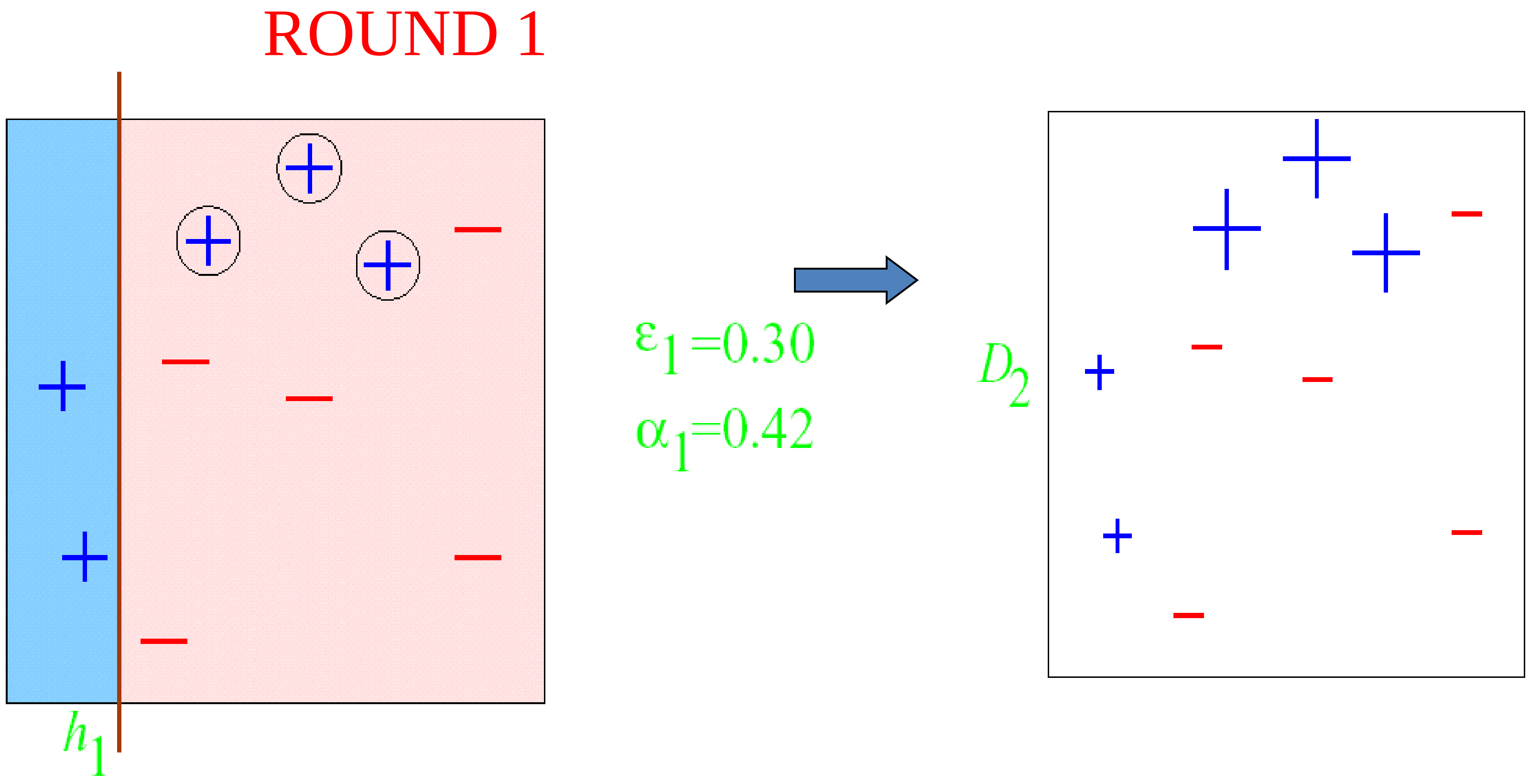


Original training set: equal weights to all training samples

AdaBoost example

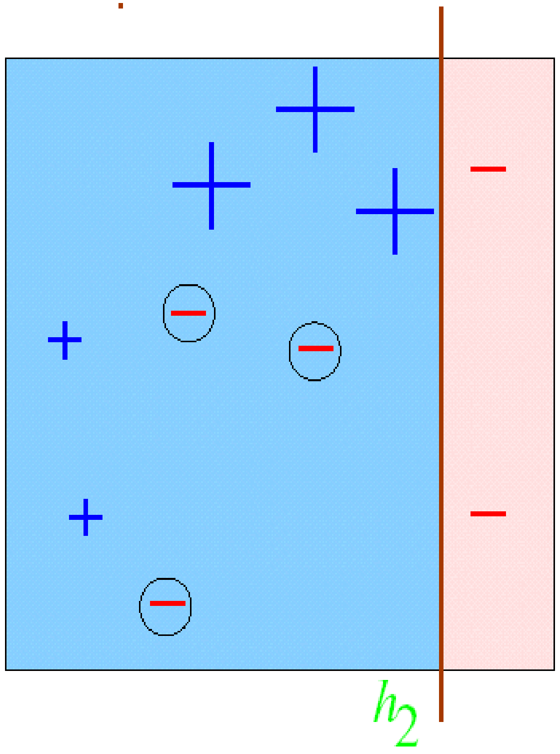
ε = error rate of classifier

α = weight of classifier



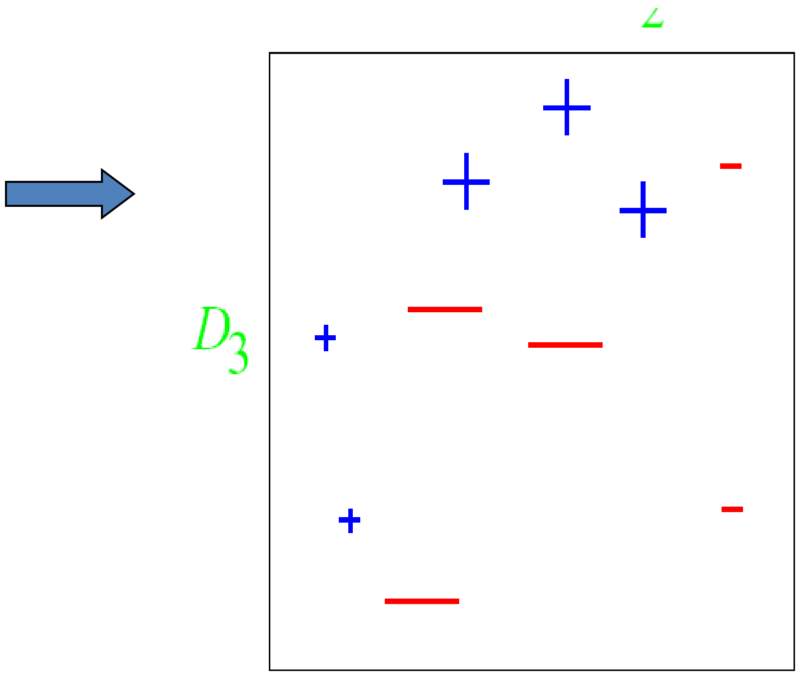
AdaBoost example

ROUND 2



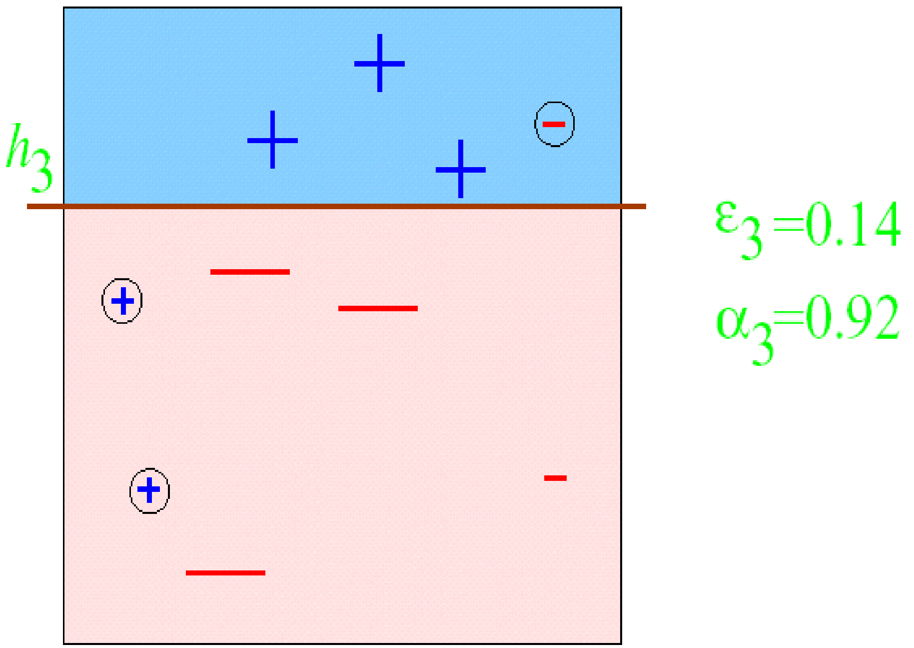
$\varepsilon_2=0.21$

$\alpha_2=0.65$

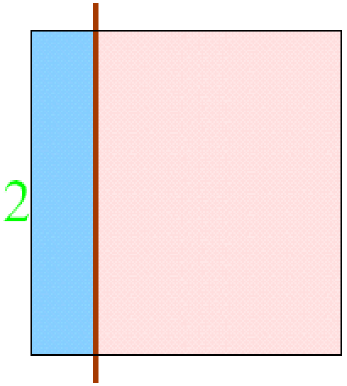
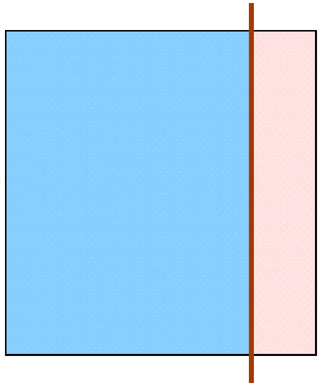
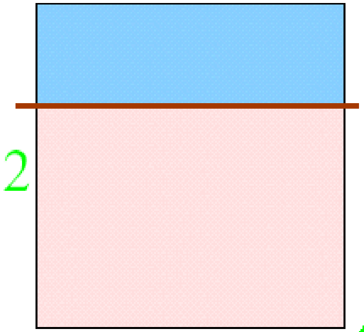


AdaBoost example

ROUND 3



AdaBoost example

$H_{\text{final}} = \text{sign}\Big($ 0.42  + 0.65  + 0.92  $\Big)$

The Data Set & Rattle()

For facilitations of comparisons:

1. We continue our discussion with the weather dataset
2. We shall also continue with the rattle() package

References

- Boosting and AdaBoost for Machine Learning
- What is AdaBoost - Quora
- Adaboost – Wikipedia
- Identify, describe, plot, and remove the outliers from the dataset
- Try-Out: Wine Quality Data Set - UCI Machine Learning Repository