Boosting
Seminar: Foundations of Data Science

Christian Peters, enrolment no.: 213996
Faculty of Statistics
TU Dortmund

December 30, 2020
winter term 2020/21

Abstract

A short summary of about two to three sentences that briefly and concisely outline the content ...

1 Introduction

Training machine learning models in practice is not always as simple as it might seem from a theoretical standpoint. In [1, chapter 2], the empirical risk minimization (ERM) rule was introduced as the learning algorithm of choice. However, this theoretical principle of choosing a hypothesis $h \in \mathcal{H}$ such that $L_S(h) = \min_{h \in \mathcal{H}} L_S(h)$, where $L_S(h)$ is the error of h on the training sequence S, can be impossible to use in practical applications due to the sometimes enormous computational complexity of searching through interesting hypothesis classes $\mathcal{H}$.

This problem leads to the question if it is possible to arrive at a strong learning algorithm in a way that doesn't require the computational cost of searching through complex hypothesis classes. Is it perhaps possible to create a strong learner by finding a way to combine "weak" learners that are potentially easier to compute? The algorithmic paradigm of boosting deals with exactly this question, resulting in the widely used AdaBoost algorithm that shows how "weak" learners can be combined in order to obtain a strong learning algorithm.

The goal of this article is to first lay down the foundations of boosting by explaining the concept of weak learnability which will be used to arrive at the AdaBoost algorithm followed by a discussion of its implications as well as a practical example of how it can be used in the domain of image classification.

2 Weak Learnability

This and that...

3 AdaBoost

Here we work with the above definitions and notations and derive important results such as the following Theorem on the least-squares solution

Theorem 1. *(useful theorem) Let $X \in \mathbb{R}^{n\times d}, Y \in \mathbb{R}^n$. Further define* $\beta^* = \mathrm{argmin}_{\beta\in R^d} \|X\beta - Y\|^2$*. Then*

$$\|Y\|^2 = \|X\beta^*\|^2 + \|X\beta^* - Y\|^2.$$

Proof. The proof is left as an exercise. □

Sometimes figures help to illustrate a formalism. This is completely unrelated to Theorem 1.

4 Conclusion

Even after centuries of research in the field of data science, there is nothing more versatile than the useful theorem of chapter 3. It is used everywhere and has led to the greatest and most intriguing results, cf. [2]. By the way, the book for the seminar [1] is a great reference and should be cited. Further literature can be found in the respective *Bibliographic Remarks* sections and of course you are welcome to search and add your own references.

Note: BibTEX entries can often be found in the DBLP collection. Google Scholar also offers BibTEX entries, which can be copied into the .bib file and may need some minor adjustments.

References

[1] S. Shalev-Shwartz and S. Ben-David. *Understanding Machine Learning - From Theory to Algorithms.* Cambridge University Press, 2014.

[2] J. Someone and J. Someoneelse. Useful theorems, 2003.