

Classical Machine Learning: Classification and Regression (III)

Learning Objectives

- Learn the basic concepts of ensemble classifiers.

Classification Algorithm Walkthrough: Ensemble Classifiers - Boosting

C-S David Chen, Department of Civil Engineering, National Taiwan University

Boosting

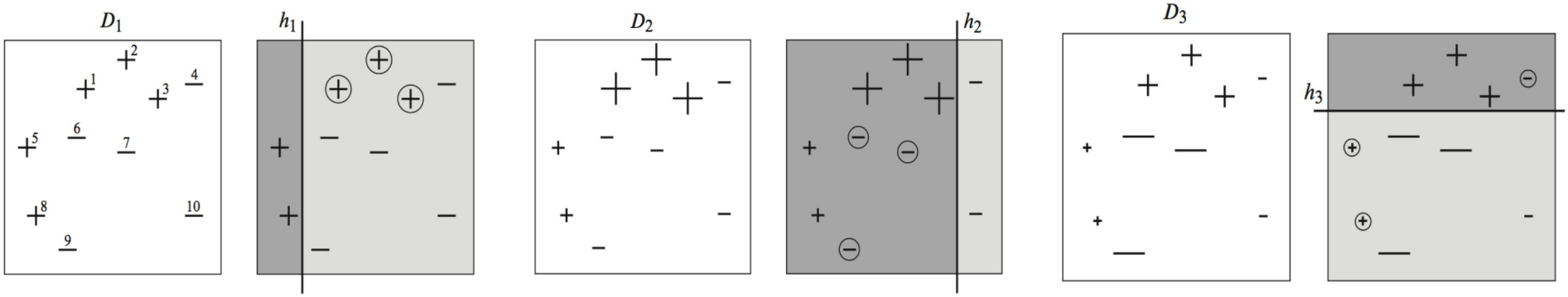
- An iterative procedure to adaptively change distribution of training data by focusing more on previously misclassified records

- Two popular versions: adaptive boosting (AdaBoost) and gradient boosting.

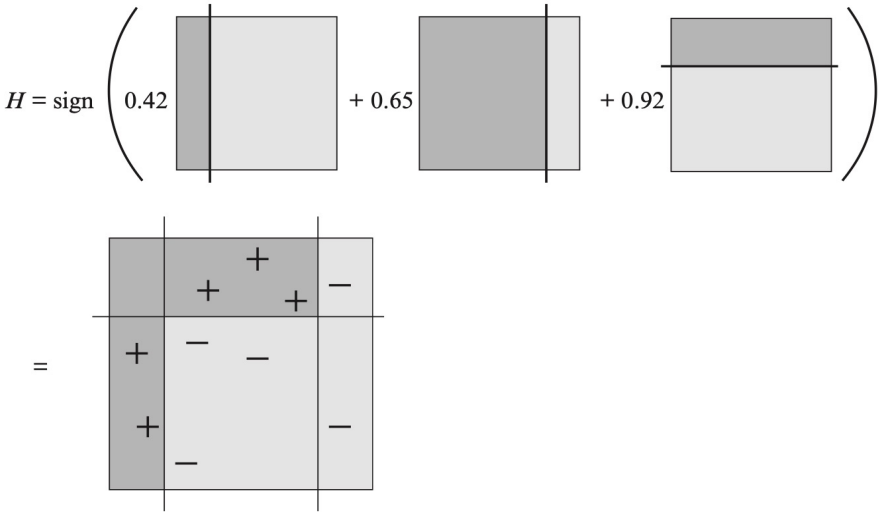
AdaBoost: theoretical minimum and example

- The phrase “theoretical minimum” is taken from a very successful book series written by Leonard Susskind, a great physicist at Stanford University.
- “Theoretical minimum” means just the minimum theories and equations you need to know in order to proceed to the next level.
- See Ensemble_AdaBoost.pdf

AdaBoost: Summary



- Fit an additive model (ensemble) $\sum_t \alpha_t h_t(\mathbf{x})$ in a forward stage-wise manner.
- In each stage, introduce a weak learner to compensate the shortcomings of existing weak learners.
- In Adaboost, "shortcomings" are identified by high-weight data points.



credit: Cheng Li, A Gentle Introduction to Gradient Boosting.

Review

Summary

Know the modern ML landscape

- Scikit-Learn and Keras (now part of TensorFlow) are mostly widely used ML software frameworks by ML professionals.
- From 2016 to 2020, the entire machine learning and data science industry has been dominated by these two approaches: deep learning and gradient boosted trees. Specifically, gradient boosted trees is used for problems where structured data is available, whereas deep learning is used for perceptual problems such as image classification.
- Users of gradient boosted trees tend to use Scikit-Learn, XGBoost or LightGBM. Meanwhile, most practitioners of deep learning use Keras, often in combination with its parent framework TensorFlow.
- The common point of these tools is they're all Python libraries: Python has is by far the most widely-used language for machine learning and data science.

Gradient Boosting = Gradient Descent + Boosting

- Fit an additive model (ensemble) $\sum_t \alpha_t h_t(\mathbf{x})$ in a forward stage-wise manner.
- In each stage, introduce a weak learner to compensate the shortcomings of existing weak learners.
- In Gradient Boosting, "shortcomings" are identified by gradients.
- Recall that, in Adaboost, "shortcomings" are identified by high-weight data points.
- Both high-weight data points and gradients tell us how to improve our model.

credit: Cheng Li, A Gentle Introduction to Gradient Boosting.

Gradient Boosting: theoretical minimum and example

- The phrase “theoretical minimum” is taken from a very successful book series written by Leonard Susskind, a great physicist at Stanford University.
- “Theoretical minimum” means just the minimum theories and equations you need to know in order to proceed to the next level.
- See Ensemble_GradBoost.pdf

Ensemble classifiers

- **Bagging**
- **Random Forest**
- **AdaBoost**
- **XGBoost**



Ensemble_compare_02.ipynb

Battle Between XGBoost, LightGBM, CatBoost



- XGBoost（eXtreme Gradient Boosting） 特點是計算速度快，模型表現好，可以用於分類和回歸問題中，**號稱「比賽奪冠的必備殺器」**。
- LightGBM（Light Gradient Boosting Machine）的**訓練速度和效率更快**、使用的內存更低、**準確率更高**、並且支持並行化學習與處理大規模數據。
- Catboost（ Categorical Features+Gradient Boosting）採用的策略在降低過擬合的同時保證所有數據集都可用於學習。性能卓越、通用性更好、易於使用而且更實用。

原文網址：https://kknews.cc/tech/vlxj8n2.html https://lavanya.ai/2019/06/27/battle-of-the-boosting-algorithms