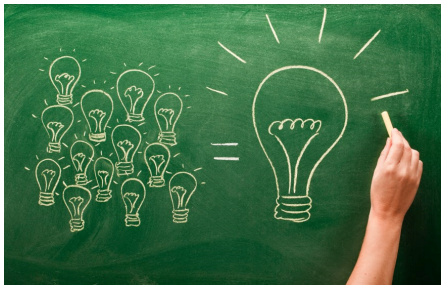
Random Forest

Akito Sakurai

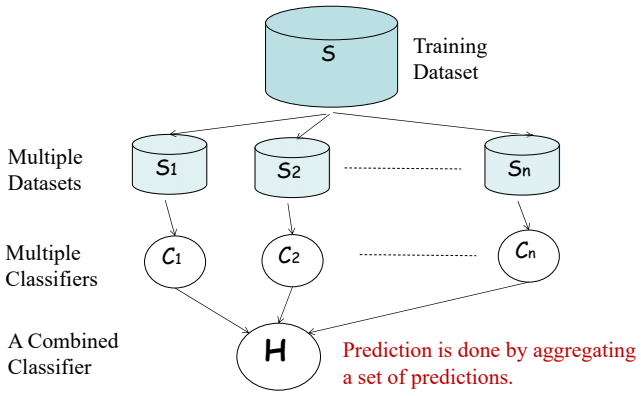
Many heads are better than one



三个臭皮匠合成一个诸葛亮　　三人寄れば文殊の知恵

http://www.leanleader.org/2012/04/many-heads-are-better-than-one.html

General idea of ensemble methods



Prediction is done by aggregating a set of predictions.

Ensemble Methods

Bagging by Breiman 1994,…

bootstrap **agg**regat**ing**

Boosting by Freund and Schapire 1995, Friedman et al. 1998,…

converts weak learners to a strong one

the connection (between boosting and bagging) is at best superficial and that boosting is fundamentally different (from bagging).

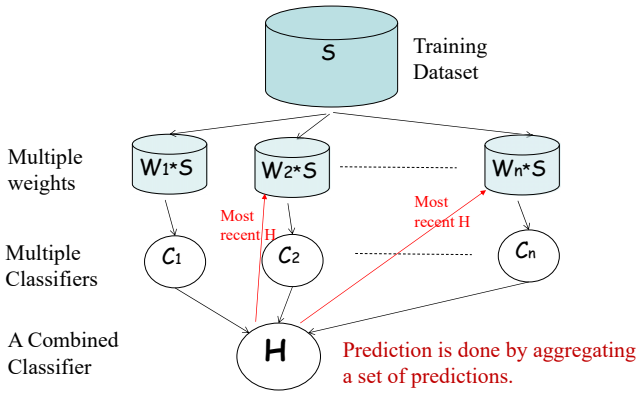
(Hastie et al., Elements of Statistical Learning)

Random forests by Breiman 2001,…

a combination of "bagging" and random selection of features

a trademark as of 2019, owned by Minitab, Inc.

General idea of boosting



Prediction is done by aggregating a set of predictions.

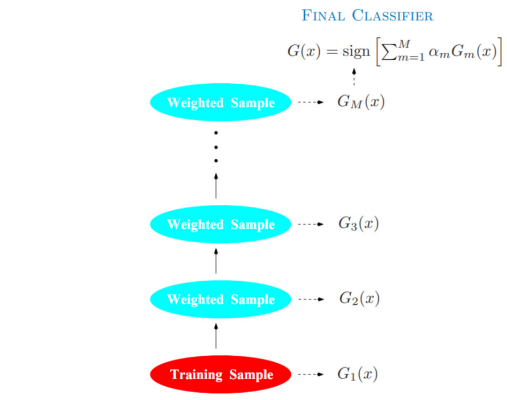


FIGURE 10.1. *Schematic of AdaBoost. Classifiers are trained on weighted versions of the dataset, and then combined to produce a final prediction.*

Hastie et al.,"The Elements of Statistical Learning: Data Mining, Inference, and Prediction", Springer (2009)

Random Forest in a nutshell

- An ensemble of decision trees.
- A bootstrapped training dataset is used for training a tree.
- A random subset of features are used for training a tree.
- Trees are not pruned.
- Output is an average or a majority.

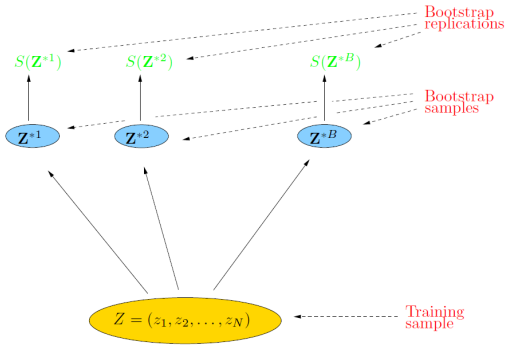


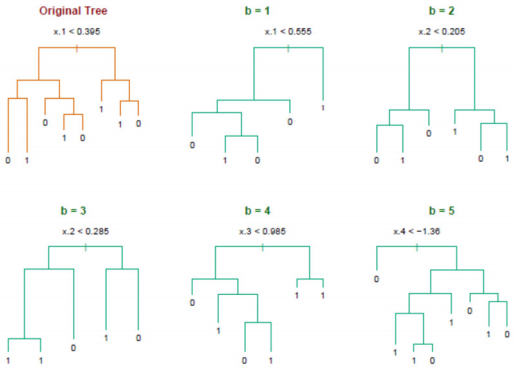
FIGURE 7.12. *Schematic of the bootstrap process. We wish to assess the statistical accuracy of a quantity $S(\mathbf{Z})$ computed from our dataset. B training sets $\mathbf{Z}^{*b}$, $b = 1, \ldots, B$ each of size N are drawn with replacement from the original dataset. The quantity of interest $S(\mathbf{Z})$ is computed from each bootstrap training set, and the values $S(\mathbf{Z}^{*1}), \ldots, S(\mathbf{Z}^{*B})$ are used to assess the statistical accuracy of $S(\mathbf{Z})$.*

Hastie et al.,"The Elements of Statistical Learning: Data Mining, Inference, and Prediction", Springer (2009)

Bootstrap sample

- $\mathrm{x}^* = (x^*_1, \ldots, x^*_n)$: a set of random samples of size n being the same as the original sample, drawn with replacement from the original sample set $\mathrm{x} = (x_1, \ldots, x_n)$
  - Each sample in x can appear in x^* zero times, once, twice, etc.
- Ratio of unique instances in x^*: around $1 - 1/e = 63.2\%$
  - i.e., about $1/e \approx 1/3$ is out-of-bag (OOB)

Note: trees learned from bootstrap samples are different from the original tree



Hastie et al.,"The Elements of Statistical Learning: Data Mining, Inference, and Prediction", Springer (2009)

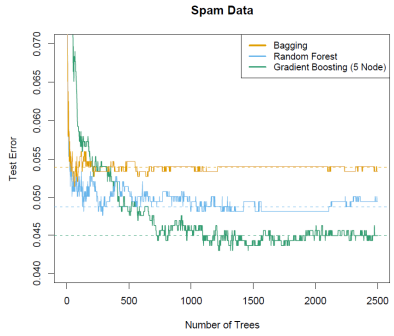


FIGURE 15.1. *Bagging, random forest, and gradient boosting, applied to the spam data. For boosting, 5-node trees were used, and the number of trees were chosen by 10-fold cross-validation (2500 trees). Each "step" in the figure corresponds to a change in a single misclassification (in a test set of 1536).*

Hastie et al.,"The Elements of Statistical Learning: Data Mining, Inference, and Prediction", Springer (2009)

Out of bag error estimation

- No cross validation is required to estimate validation error
- In bootstrapping we sample with replacement, and therefore on average 1/3 of them are not used for training.
- These are called out-of-bag samples (OOB)
- We can treat OOBs as if it is a validation dataset and calculate overall OOB MSE or classification error

Advantages of random forest

- RF and SVM are best in accuracy among current learning algorithms.
- It is efficient on a large training dataset.
- It handles thousands of input variables.
- It estimates the importance of input variables.
- OOBs, which are easily calculated in the learning process, are validation errors.
- It has an effective method for estimating missing data as in decision tree building.

Disadvantages of random forest

- Unlike decision trees of moderate sizes, the predictions made by random forests are difficult for humans to interpret.
- As decision trees, if categorical variables with different number of values (levels) exist, those attributes with more levels are favored.