

ensemble

Authors

Alexander Jung*, Department of Computer Science, Aalto University, Espoo, Finland

*Corresponding author: alex.jung@aalto.fi

Abstract

An ensemble method combines several machine learning (ML) methods, each referred to as a base learner, into one predictor. The predictions of the base learners are aggregated by averaging in regression or by a majority vote in classification. The combination often predicts more reliably than any single base learner. Ensemble methods differ in how they construct the base learners: bootstrap aggregating (bagging) trains each on a resampled copy of the training set, boosting runs them sequentially so that each corrects the errors of its predecessors, and stacking trains different models on the same training set.

Definition

For each incoming email, three different spam filters cast a vote, and the message is moved to the spam folder when at least two of them say spam. Such a combination of machine learning (ML) methods is an ensemble method: each combined method is a base learner, and the combination often predicts more reliably than any single base learner (Hastie et al., 2009, Ch. 16). The base learners can be empirical risk minimization (ERM)-based, using different choices for the loss, model, and training set. The aggregation of their predictions can amount to averaging (in regression) or to a majority vote (in classification), as in the spam example (Fig. 1).

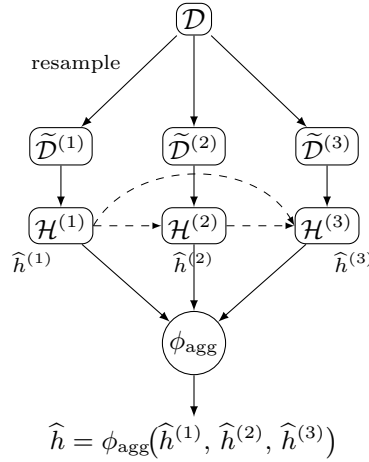


Figure 1: Generic ensemble with three base learners: base learner $\ell \in \{1, 2, 3\}$ uses ERM to learn $\hat{h}^{(\ell)} \in \mathcal{H}^{(\ell)}$ from the perturbed training set $\tilde{\mathcal{D}}^{(\ell)}$. A base learner might also use the output of other base learners (dashed). The final hypothesis $\hat{h}$ is obtained by aggregating the hypotheses of the base learners via an aggregation rule ϕ_{agg}

Different ensemble methods use different constructions for the base learners. For example, bootstrap aggregating (bagging) methods (such as a random forest) use random sampling to construct slightly different training sets for each base learner. On the other hand, boosting methods run the base learners sequentially, i.e., each base learner tries to correct the prediction errors of the previous ones. A third family of ensemble methods is stacking, where base learners are trained on the same training set but with different models.

Cross References

bagging, boosting, stacking, base learner, random forest

References

1. Hastie et al. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed., Springer Science+Business Media, New York, NY, USA.