

Foundations of Machine Learning

Brett Bernstein

August 22, 2018

Boosting: Concept Check

Boosting Learning Objectives

- Compare learning a linear model on a fixed set of basis functions on the input space, and an "adaptive basis function model" where the basis functions are learned.
 - In particular, explain the "recipe" for an adaptive basis function model in terms of the base hypothesis space, and combined hypothesis space.
 - Give psuedo-code for forward stagewise additive modeling (FSAM).
 - Give the ingredients for gradient boosting machines; in particular, be able to explain why we need a [sub]differentiable loss function w.r.t. the prediction.
 - Explain how gradient boosting uses "functional" gradient descent - i.e. learning the basis function (i.e. function in the base hypothesis space) that is closest to the negative gradient step direction given the current prediction function.
 - Explain options for step sizes (line search and shrinkage parameter/learning rate).
 - Explain variations on gradient boosting (stochastic gradient boosting, and column subsampling).
1. (★) Show the exponential margin loss is a convex upper bound for the $0 - 1$ loss.
 2. Show how to perform gradient boosting with the hinge loss.
 3. Suppose we are using gradient boosting. On each step we can do a better job of fitting the pseudoresiduals if we allow for deeper trees. Why might deep trees be discouraged while gradient boosting?