Statistical Learning - Vienna, Fall2017

Homework problem 9

Boosting with other loss functions

In class we showed a Boosting implementation which uses the current residuals as the response for building each tree, is in fact performing gradient descent for regression with a squared error loss function $L(y, \hat{y}) = (y - \hat{y})^2$. In this problem, we will discuss applying the same idea to regression with an absolute loss function: $L(y, \hat{y}) = |y - \hat{y}|$.

1. For the b^{th} iteration, write the loss function in terms of the current residual $r_i = y_i - \hat{f}_{b-1}(x_i)$.

2. Calculate its derivative, and explain what would need to change in the boosting algorithm so it would implement gradient descent in this loss function.

3. The code `boost.r` available from the class homepage, implements regular gradient boosting with squared error loss on the famous Boston Housing dataset. Implement the absolute loss version as well. Using appropriate choices of depth, ϵ and B (you can try several), compare prediction performance of both boosting models on the test set. Use both squared error loss and absolute loss in comparing the test performance. Which method does better on which performance task? Is this as you would expect?
   Hints: You need to change only one line in the main loop to implement the change of loss function, think carefully which one... You may find the R function `sign()` useful.