STAT 202A Homework 7

Xiaohan Wang

405033965

(1) Write R code for ridge regression and linear spline, using sweep operator written in Rcpp. Explore the effect of regularization on the smoothness of the fitted curve (Output curve in one plot) and the estimator error in terms of the L2 difference between the true curve and the learned curve. (plot error by different regularize)

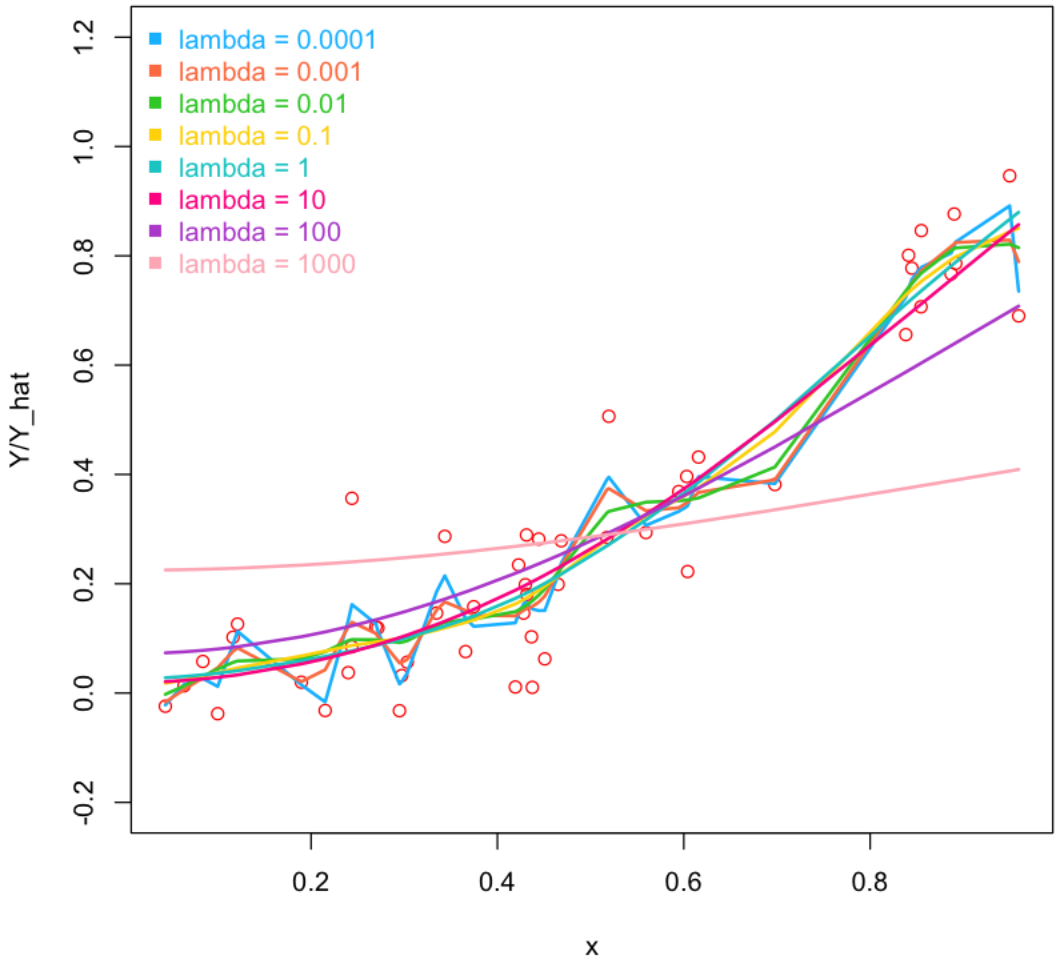

Figure 1. linear spline curve with different regularization

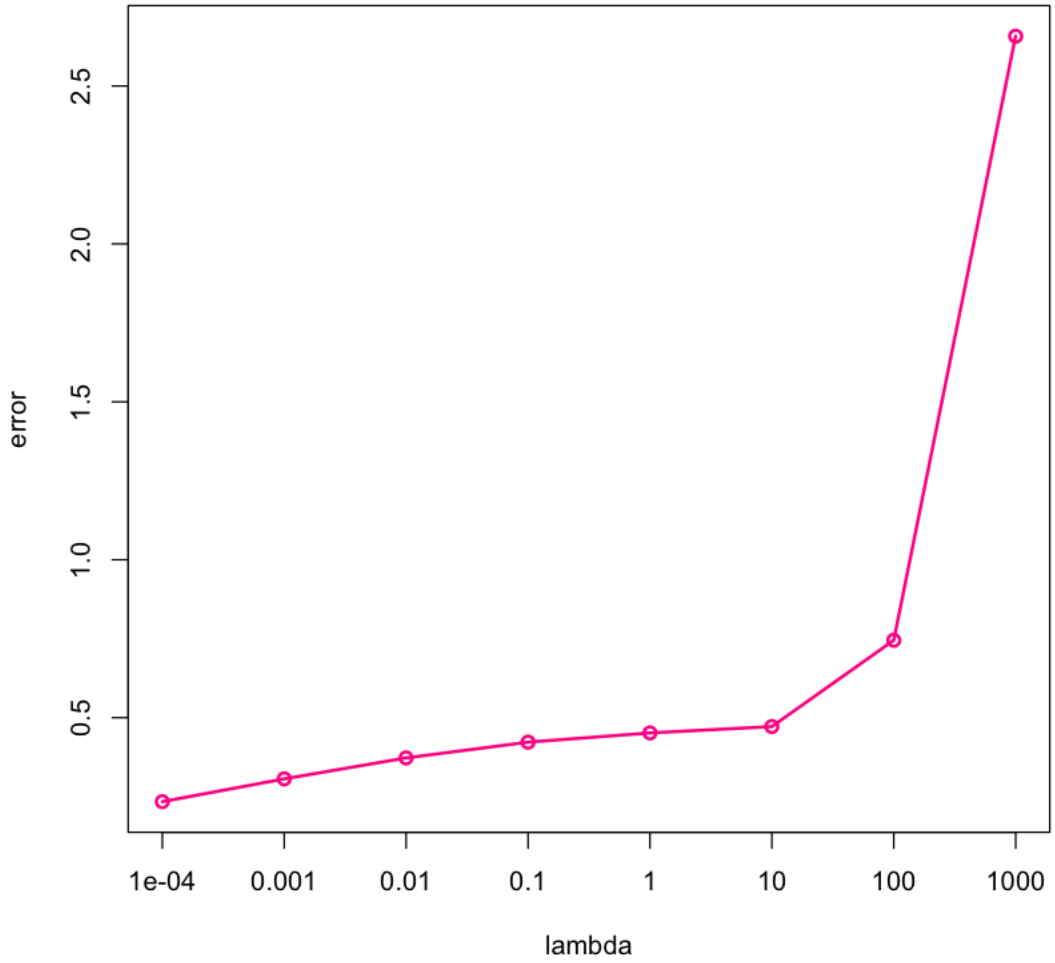

Figure 2. error by different regularization

From the above two figures, we can find that when lambda is small, the error is small. But the curve is very zigzag and seems to be overfitted. As lambda increases, the curve becomes smooth, and the error increases as well. On the whole, with the trade-off of smoothness of the curve and fitted error, we can choose lambda of 1 or 10.

(2) Write Rcpp code for coordinate descent algorithm for computing the solution path of Lasso. Write Rcpp code for epsilon-boosting and compare its solution path with the solution path of Lasso. (Plot solution path for both and comment)

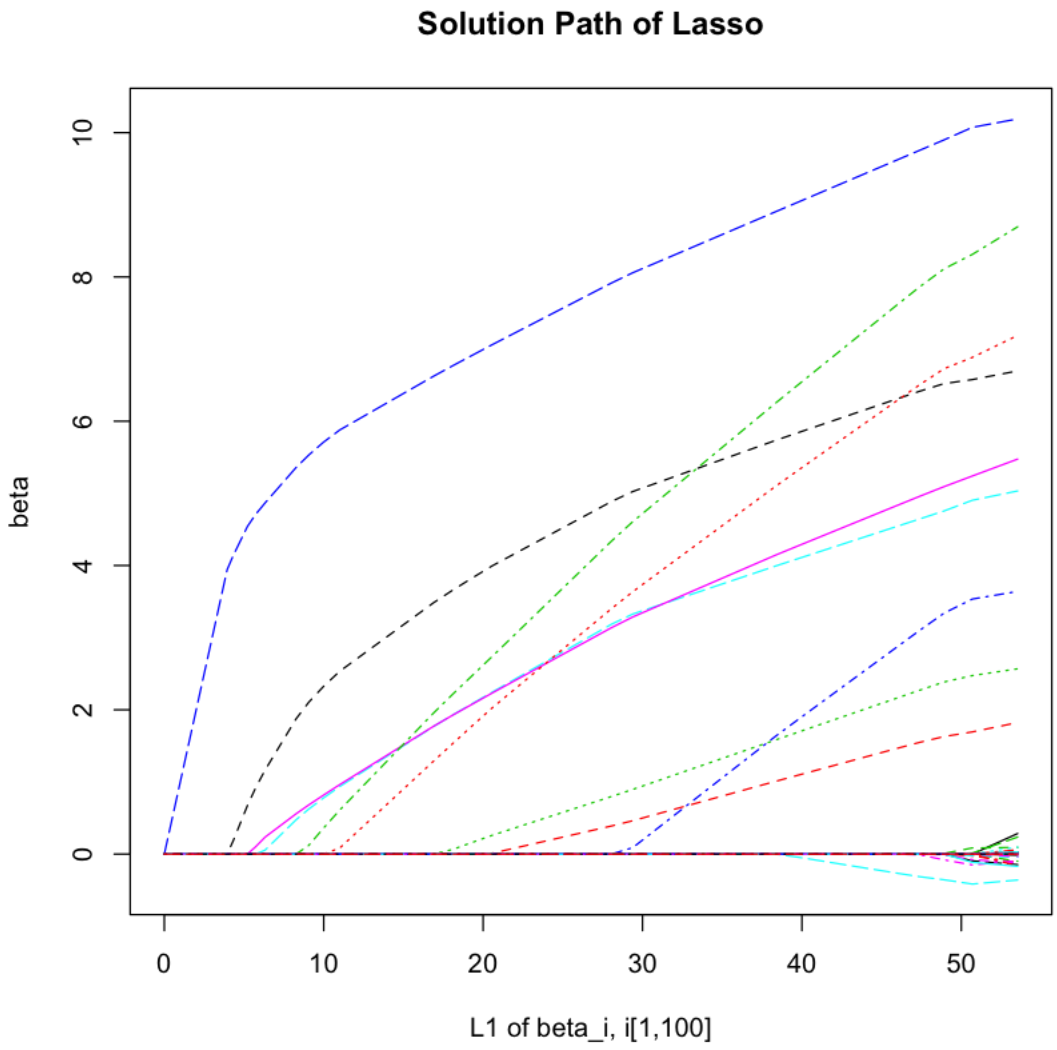

Figure 3. solution path of Lasso

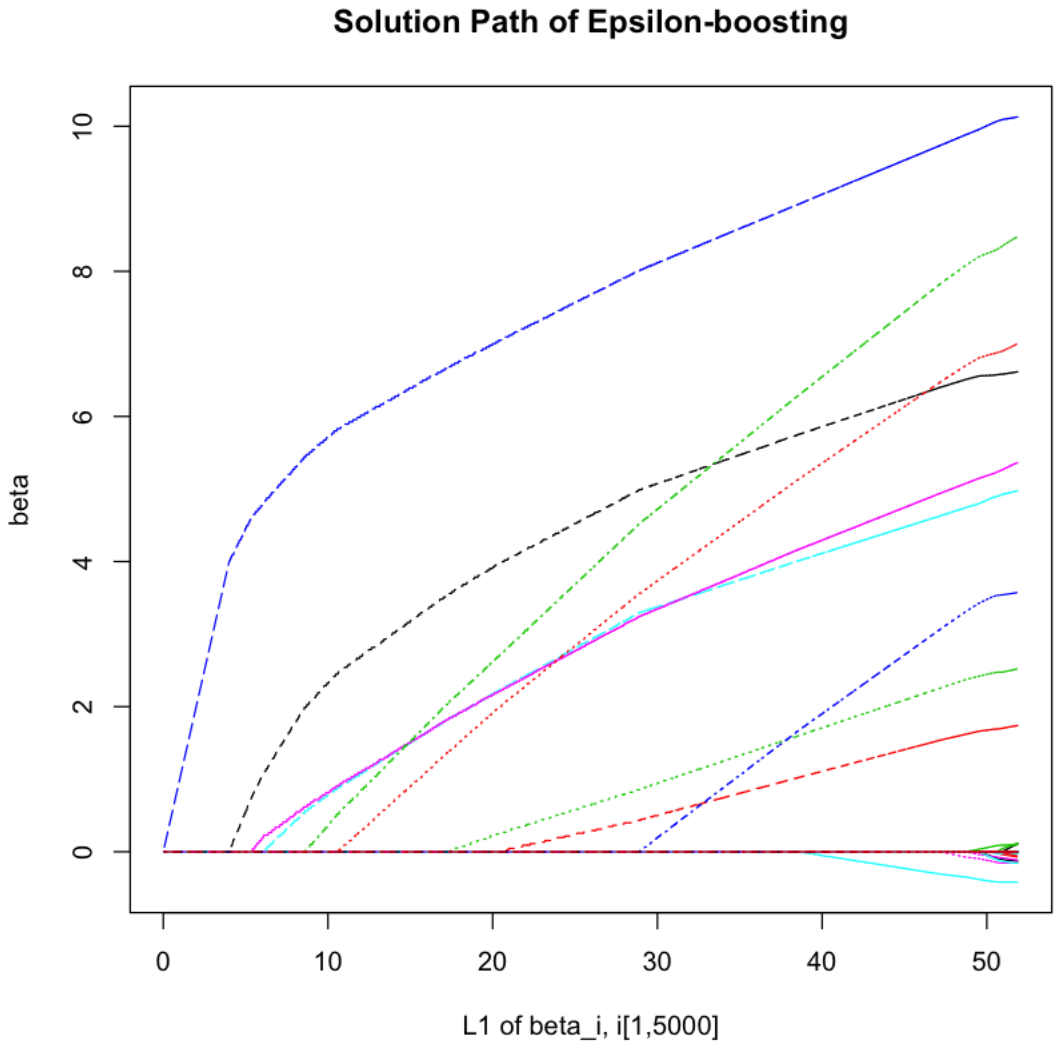

Figure 4. solution path of Epsilon-boosting

From the above two figures, we can see that on the one hand, these two figures are very similar. All components of beta are zero at first (lambda is big), then they increase or decrease and gradually become stable. On the other hand, although they have same range of x-axis, Lasso takes 100*10 rounds, while epsilon-boosting takes 5000 rounds. If we use the same rounds, epsilon-boosting is much less selectable (components of beta are less separable) than Lasso, and the L1 of beta_i of epsilon-boosting is also smaller than that of Lasso.