Meeting: Robust TFBoost

July 2, 2021

- Competitors:
  - Maronna and Yohai (2013): MM-estimator for functional linear model
  - Shin and Lee (2016): M-estimator for functional linear model (slope function in the RKHS, M estimator using MAD as the residual scale.) Compared to Maronna and Yohai (2013), they studied asymptotic properties of the estimator. Why want to assume RKHS?
  - Qingguo (2017): FPC scores + M-estimator (tuning constant selected to minimize a criterion) (seems not very hard to implement)
  - Huang et al. (2015): sieve M-estimators for semi-functional model. (have code)
  - Cai et al. (2020): a new Huber loss with data-driven tuning parameters. "enables us to reach better robustness and efficiency than other robust methods in the presence of outliers or heavy-tailed error distribution" (not a good journal?)
  - Boente et al. (2020): MM-estimator for semi-functional linear model. (have code)
  - Wang et al. (2017): functional sliced inverse regression (trimmed estimation of the mean and covariance) (different assumptions?)
  - Kalogridis and Van Aelst (2019): robust FPC scores + MM functional linear model.
- Other settings: high-leverage? scalar covariates?
- L2Boost + Robust base learner
  - Suggested by Lutz et al. (2008) (robust linear regression learners)
  - Robust trees
    - At each node, find the split to maximize the decrease in

$$\sum_{t=1}^{T} \sum_{i \in R_t} \rho\left(\frac{y_i - \theta_t}{\hat{\sigma}}\right),$$

      where θ_t is the location estimator at node t, and $\hat{\sigma}$ is the MAD of residuals calculated from a LADTree.

* Chambers et al. (2004) used to identify outliers. For the t-th node

$$WSSR_t = \sum_{i \in R_j}^{n} w_{i,t}(y_i - \theta_j)^2,$$

where $w_{i,t}$ are weights returned by rlm function (did not mention about scale)

$$w_{i,t} = \frac{\psi(y_i - \theta_j)}{y_i - \theta_j},$$

and

$$\theta_j = \frac{\sum_{i \in R_j} w_{i,t} y_i}{\sum_{i \in R_j} w_{i,t}}.$$

Identify outlier "as a case with an average weight over all node splits that are less than a specified threshold." Stop at a given number of leaf nodes.

* Yang et al. (2019) a rank-based impurity metric as the spliting criterion

- MM-tree: similar as RRBoost, but from top to bottom. First find splits that minimizes the S-scale, then use it as the robust scale. Then find splits to minimize the M-loss.
- Challenge: need more computation than RRBoost, since the robust fits are usually calculated iteratively itself.
- Maybe applied to random forest? ensemble horizontally. Each tree use a different set of projection indices.

References

Boente, G., Salibian-Barrera, M., and Vena, P. (2020). Robust estimation for semi-functional linear regression models. *Computational Statistics & Data Analysis*, 152:107041.

Cai, X., Xue, L., and Wang, Z. (2020). Robust estimation with modified huber's function for functional linear models. *Statistics*, pages 1–11.

Chambers, R., Hentges, A., and Zhao, X. (2004). Robust automatic methods for outlier and error detection. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 167(2):323–339.

Huang, L., Wang, H., Cui, H., and Wang, S. (2015). Sieve m-estimator for a semi-functional linear model. *Science China Mathematics*, 58(11):2421–2434.

Kalogridis, I. and Van Aelst, S. (2019). Robust functional regression based on principal components. *Journal of Multivariate Analysis*, 173:393–415.

Lutz, R. W., Kalisch, M., and Bühlmann, P. (2008). Robustified l2 boosting. *Computational Statistics & Data Analysis*, 52(7):3331–3341.

Maronna, R. A. and Yohai, V. J. (2013). Robust functional linear regression based on splines. *Computational Statistics & Data Analysis*, 65:46–55.

Qingguo, T. (2017). M-estimation for functional linear regression. *Communications in Statistics-Theory and Methods*, 46(8):3782–3800.

Shin, H. and Lee, S. (2016). An rkhs approach to robust functional linear regression. *Statistica Sinica*, pages 255–272.

Wang, G., Zhou, J., Wu, W., and Chen, M. (2017). Robust functional sliced inverse regression. *Statistical papers*, 58(1):227–245.

Yang, B.-B., Gao, W., and Li, M. (2019). On the robust splitting criterion of random forest. In *2019 IEEE International Conference on Data Mining (ICDM)*, pages 1420–1425. IEEE.