

A Central Limit Theorem for Centered Purely Random Forests using U-Statistic Theory

Wednesday, 3 August 2022 18:00 (20 minutes)

Random forests are a popular method in supervised learning and can be used for regression and classification problems. For a regression problem a random forest averages the results of several randomized decision trees that are constructed on different subsamples of the dataset. In practice random forests appear to be very successful and are therefore a commonly used algorithm. Contrary to this there is little known about the mathematical properties of classic random forests that use data dependent partitions. Most results in the literature cover simpler versions of random forests often with partitions that are independent of the dataset. One example of these simpler algorithms are centered purely random forests. Moreover the majority of the results in the literature are consistency theorems and there are noticeably less central limit theorems. In our work we prove a central limit theorem for centered purely random forests. The proof uses results by Peng et al. (2022) which are based on an interpretation of random forests as generalized U-Statistics.

References

Wei Peng, Tim Coleman, and Lucas Mentch. Rates of convergence for random forests via generalized u-statistics. *Electronic Journal of Statistics*, 16(1):232–292, 2022.

Primary author: RABE, Jan (Universität Hamburg)

Session Classification: Session B2 Statistics

Track Classification: Statistics