

LCLS Analysis and Automation Advances: Bi-Cross Validation of the Inverted Laplacian for Cluster Number and Generalized Hyper-Parameter Estimation

S. Zohar

*Photon Data and Controls Systems,
Linac Coherent Light Source,
SLAC National Accelerator Laboratory
2575 Sand Hill Rd, Menlo Park, CA 94025
(Dated: March 12, 2019)*

Robust automation of analysis procedures capable of handling diverse data sets is critical for high data throughput experiments at the Linac Coherent Light Source (LCLS). A challenge encountered in this process is determining the number of clusters required for the execution of conventional clustering algorithms. Here, it is demonstrated that bi-cross validation of the inverted and regularized Laplacian used to embed data in a higher dimensional representation provides a convex surface that a minimum at the expected number of clusters and radial basis function kernel gamma parameter. These results indicate that the process of estimating the number of clusters cannot be divorced from the process of estimating other hyper-parameters. An application of this approach to LCLS x-ray scattering data demonstrates the ability to identify clusters of dropped shots without manual assigning of cut ranges and provides a path towards identification of rare events.