Clustering Techniques and Hierarchical Matrix Formats for Kernel Ridge Regression

Xiaoye S. Li [1]

We present memory-efficient and scalable algorithms for kernel methods used in machine learning. Using hierarchical matrix approximations for the kernel matrix the memory requirements, the number of floating point operations, and the execution time are drastically reduced compared to standard dense linear algebra routines. We consider both the general H-matrix hierarchical format as well as Hierarchically Semi-Separable (HSS) matrices. Furthermore, we investigate the impact of several preprocessing and clustering techniques on the hierarchical matrix compression. Effective clustering of the input leads to a ten-fold increase in efficiency of the compression. The algorithms are implemented using the STRUMPACK solver library. These results confirm that classification using kernel ridge regression with the compressed matrix does not lose prediction accuracy compared to the exact kernel matrix and that our approach can be extended to O(1M) datasets, for which computation with the full kernel matrix becomes prohibitively expensive. We present numerical experiments in a distributed memory environment up to 1,024 processors of the NERSC's Cori supercomputer using well-known datasets to the machine learning community that range from dimension 8 up to 784.

---

[1] Lawrence Berkeley Lab.