Osprey: Hyperparameter Optimization for Machine Learning

Robert T. McGibbon[1], Carlos X. Hernández[1], Matthew P. Harrigan[1], Steven Kearnes[1], Mohammad M. Sultan[1], Stanislaw Jastrzebski[2], Brooke E. Husic[1], and Vijay S. Pande[1]

[1]Stanford University
[2]Jagiellonian University

5 August 2016

Paper DOI: http://dx.doi.org/
Software Repository: https://github.com/msmbuilder/osprey
Software Archive: http://dx.doi.org/10.5281/zenodo.56251

Summary

Osprey is a tool for hyperparameter optimization of machine learning algorithms in Python. Hyperparameter optimization can often be an onerous process for researchers, due to time-consuming experimental replicates, non-convex objective functions, and constant tension between exploration of global parameter space and local optimization (Jones, Schonlau, and Welch 1998). We've designed *Osprey* to provide scientists with a practical, easy-to-use way of finding optimal model parameters. The software works seamlessly with `scikit-learn` estimators (Pedregosa et al. 2011) and supports many different search strategies for choosing the next set of parameters with which to evaluate a given model, including gaussian processes (GPy 2012), tree-structured Parzen estimators (Yamins, Tax, and Bergstra 2013), as well as random and grid search. As hyperparameter optimization is an embarrassingly parallel problem, *Osprey* can easily scale to hundreds of concurrent processes by executing a simple command-line program multiple times. This makes it easy to exploit large resources available in high-performance computing environments.

Osprey is actively maintained by researchers at Stanford University and other institutions around the world. While originally developed to analyze computational protein dynamics (McGibbon, Harrigan, et al. 2016), it is applicable to any `scikit-learn`-compatible pipeline. The source code for *Osprey* is hosted on GitHub and has been archived to Zenodo (McGibbon, Hernández, et al. 2016). Full documentation can be found at http://msmbuilder.org/osprey.

References

GPy. 2012. "GPy: A Gaussian Process Framework in Python." http://github.com/SheffieldML/GPy.

Jones, Donald R., Matthias Schonlau, and William J. Welch. 1998. "Efficient Global Optimization of Expensive Black-Box Functions." *Journal of Global Optimization* 13 (4): 455–92. doi:10.1023/A:1008306431147.

McGibbon, Robert T., Matthew Harrigan, Bharath Ramsundar, Kyle Beauchamp, Mohammad M. Sultan, Christian Schwantes, Carlos X. Hernández, et al. 2016. "Msmbuilder: MSMBuilder 3.4." Zenodo. doi:10.5281/zenodo.48545.

McGibbon, Robert T., Carlos X. Hernández, Matthew P. Harrigan, Steven Kearnes, Mohammad M.

Sultan, Stanislaw Jastrzebski, Brooke E. Husic, and Vijay Pande. 2016. "Osprey: Osprey 1.0.0." doi:10.5281/zenodo.56251.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12: 2825–30.

Yamins, Daniel, David Tax, and James S. Bergstra. 2013. "Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures." In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, edited by Sanjoy Dasgupta and David Mcallester, 28:115–23. 1. JMLR Workshop; Conference Proceedings. http://jmlr.csail.mit.edu/proceedings/papers/v28/bergstra13.pdf.