Osprey: Hyperparameter Optimization for Machine Learning

Robert T. McGibbon[1], Carlos X. Hernández[2], Matthew P. Harrigan[3], Steven Kearnes[4], Mohammad M. Sultan[5], Stanislaw Jastrzebski[6], Brooke E. Husic[7], and Vijay S. Pande[8]

[1]D.E. Shaw Research
[2]Stanford University
[3]Stanford University
[4]Stanford University
[5]Stanford University
[6]Jagiellonian University
[7]Stanford University
[8]Stanford University

1 July 2016

Paper DOI: http://dx.doi.org/
Software Repository:
Software Archive:

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 functionals, 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 and supports many different search strategies for choosing the next set of parameters with which to evaluate your model (Pedregosa et al. 2011; GPy 2012; Yamins, Tax, and Bergstra 2013). Its simple command-line interface makes Osprey instances easy to submit on high-performance computing environments.

Osprey is actively being developed by researchers at Stanford University with primary application areas in computational protein dynamics and drug design. The source code for *Osprey* is hosted on GitHub and has been archived to Zenodo (McGibbon et al. 2016).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., Carlos Hernández, Matthew Harrigan, Steven Kearnes, Muneeb Sultan, Stanislaw

Jastrzebski, Brooke 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.