

Special Information Science and Technology Seminar Speaker



Jake Vanderplas
University of Washington

AstroML: Machine Learning and Reproducible Research in Astronomy

Thursday, August 1, 2013

9:30 - 10:30 AM

TA-3, Bldg. 1698, Room A103 (MSL Auditorium)

Abstract: As astronomical data sets grow in size and complexity, automated machine learning and data mining methods are becoming an increasingly fundamental component of research in the field. The astroML project (<http://astroML.github.com>), first released in fall 2012, provides a common repository for practical examples of the data mining and machine learning tools used and developed by astronomical researchers, implemented in the Python programming language. The astroML module offers a host of general data analysis and machine learning routines, loaders for openly-available astronomical datasets, and fast implementations of specific computational methods often used in astronomy and astrophysics. The associated website features hundreds of examples of these routines in action, using real datasets. In this talk I'll go over some of the highlights of the astroML code and examples, discuss in detail some of the machine learning and data mining approaches which have opened up new areas of astronomical research, and share how open source Python tools and packages like AstroML are being used in the field to enhance astronomical education and research.

Biography: Jake Vanderplas is an NSF Postdoctoral fellow at the University of Washington, working jointly between the Astronomy department's Survey Science Group, and the Computer Science department's Database Research Group. He received his PhD in Astronomy from the University of Washington in 2012 for work in the areas of Cosmology, Weak Gravitational Lensing, and the application of automated machine learning and data mining approaches to large astronomical datasets in various contexts. This work led him to co-author an upcoming book, "Statistics, Data Mining, and Machine Learning in Astronomy" to be published in late 2013. Along with his research interests, Jake is an active user, developer, proponent, and teacher of the Python language for open, reproducible science, and will argue vehemently with anyone who says R is better.

For more information contact the technical host Curt Canada, cvc@lanl.gov, 665-7453 or James Ahrens, ahrens@lanl.gov, 667-5797.