In-class assignment #27

PHY-905-002, Computational Astrophysics and Astrostatistics
Fall 2018

Instructions: We're going to spend two days working through some of Jake VanderPlas' tutorials on Machine Learning. Jake is one of the authors of the website AstroML: Machine Learning and Data Mining for Astronomy as well as a very good textbook on the same subject, and also the author of the Python Data Science Handbook. These tutorials use the `scikit-learn` package, which is a set of open-source tools for data mining and data analysis.

Before we do anything else, download Jake's scikit-learn tutorial[1] and make sure that you have the correct packages installed by running the notebook `01-Preliminaries.ipynb`. You will probably have to install the Seaborn visualization library, which is based on matplotlib and useful for making pretty statistical graphs. You can do this by typing "`pip install seaborn`", and you will probably have to restart your Jupyter Notebook server after you do so to have access to Seaborn.

Useful machine learning resources:

- AstroML: Machine Learning and Data Mining for Astronomy website – lots of helpful resources and tutorials!
- `scikit-learn` website – broadly used machine learning toolkit; lots of resources.
- Book: Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data by Ivezic et al.
- Book: Python Data Science Handbook by Jake VanderPlas. (Also download the Jupyter Handbook version of the entire book, c/o Jake V.)
- A Coursera course on machine learning, taught by Andrew Ng of Stanford. This course was heavily recommended to me by my CMSE colleagues whose expertise is machine learning.
- Astronomical Machine Learning Tutorial from the 2015 Local Group Astrostatistics workshop at the University of Michigan. (This is less complete than the other references listed here, but has the advantage of brevity.)
- LSST Data Science Fellowship – A two-year fellowship that is open to grad students; you can apply for this next year! Even if you don't do the fellowship the training session materials are available on GitHub.

[1] `git clone https://github.com/jakevdp/sklearn_tutorial.git`