PHYS4038/MLiS and AS1/MPAGS

Scientific Programming in

python™

mpags-python.github.io

Steven Bamford

University of Nottingham
UK | CHINA | MALAYSIA

An introduction to scientific programming with

python™

Session 9:
Bayesian Inference and Machine Learning

Student Evaluation of Module

Students at Nottingham:

Please visit bluecastle-uk-surveys.nottingham.ac.uk

MSc students should see PHYS4038 module listed

PhD students will need to enter a "PIN number": V6Q6S8

(Interpret "Moodle" as "course webpage")

PhD students at other institutions:

Please use this anonymous form

MCMC in Python

- **Bayesian inference**
  - Given some data and a parameterised model
  - Model gives likelihood of the data for particular parameters
  - Assuming "prior" probability distributions on the parameters
  - Bayes' Theorem gives the "posterior" probability of the model
  - Sample this probability distribution in parameter space

  → Parameter estimation
  - Integrate likelihood over parameter space: "evidence"

  → Model selection

MCMC in Python

- PyMC3
  - main python module for MCMC and related tasks
- emcee
  - alternative methods, write own likelihood functions
- PyStan
  - uses own probabilistic programming language
- PyMultiNest
  - nested sampling , write own likelihood functions
  - good for model selection
- …

emcee

- "The MCMC hammer"

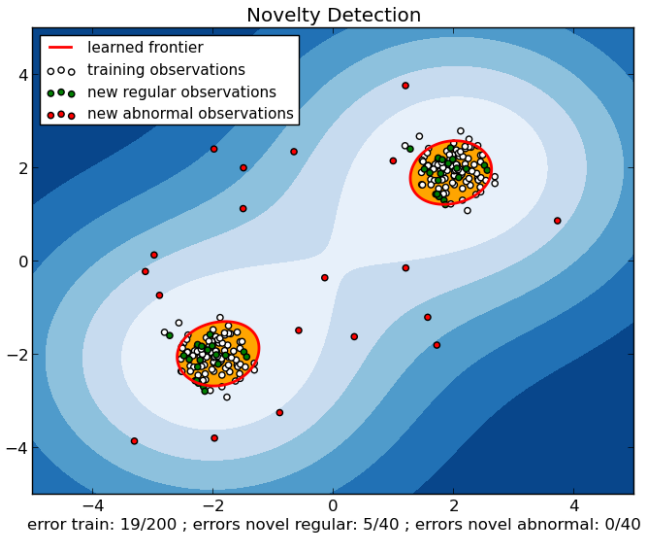
- Affine-invariant sampler
- Parallel tempering
- Easy to use
- Highly effective
- Written and advocated by NYU hipsters

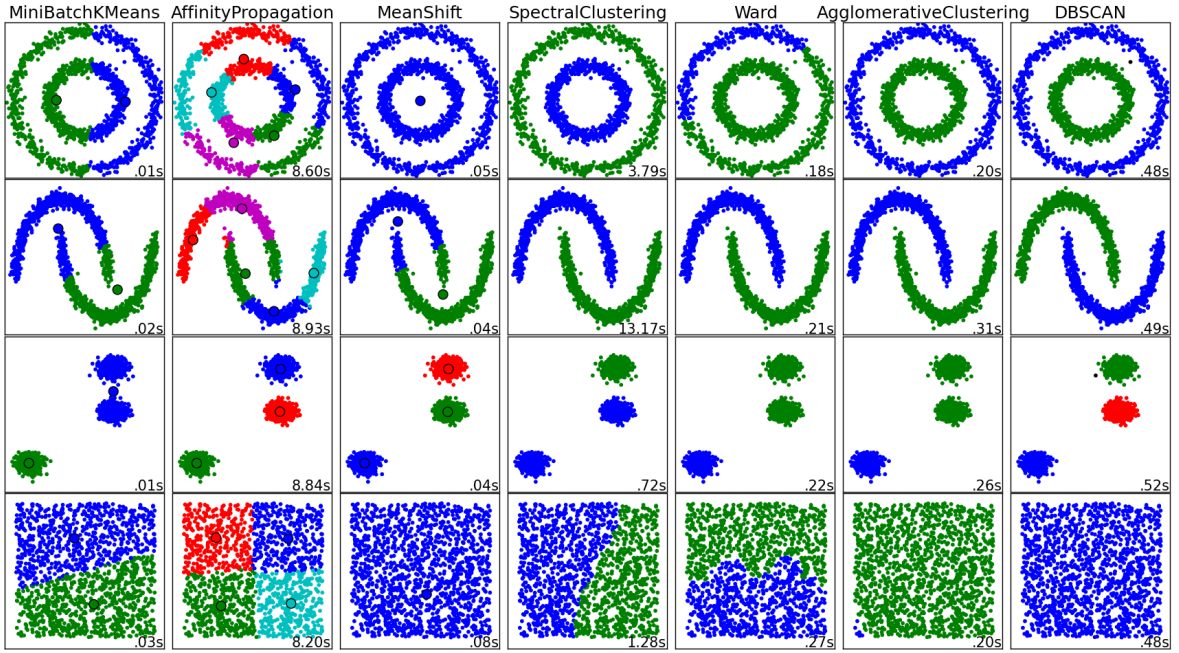
emcee notebook example

[link to online notebook]

Machine learning: scikit-learn

- http://scikit-learn.org/
- Machine learning tools for data mining and analysis
  - Classification, regression, clustering, PCA, model selection, etc.



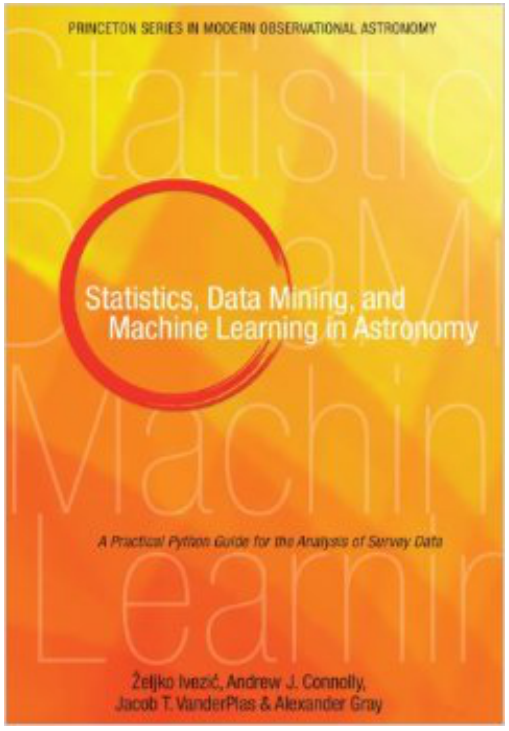


- Also see Statsmodels
  - http://statsmodels.sourceforge.net

Machine learning: AstroML

- Machine Learning and Data Mining for Astronomy
- http://www.astroml.org
- Accompanied by a book (but open-source software):
  - '*Statistics, Data Mining, and Machine Learning in Astronomy*'
    - by Zeljko Ivezic, Andrew Connolly, Jacob VanderPlas, and Alex Gray

- Routines for: dealing with survey data, density estimation, clustering, regression, classification, extreme deconvolution, two-point correlation functions, luminosity functions, etc.



Machine learning: others

Neural networks

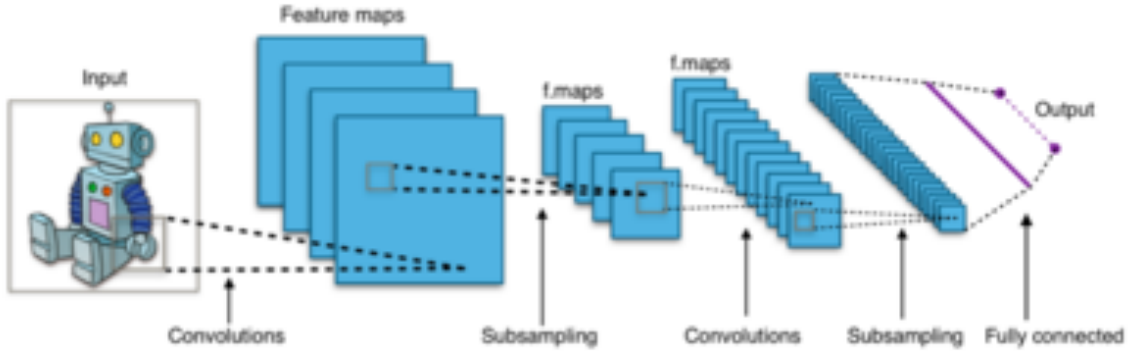
- TensorFlow
  - including keras higher-level interface
- PyTorch, …

Boosted trees

- XGBoost, …

Clustering

- HDBSCAN, …



keras

Keras MNIST example

[link to online notebook]