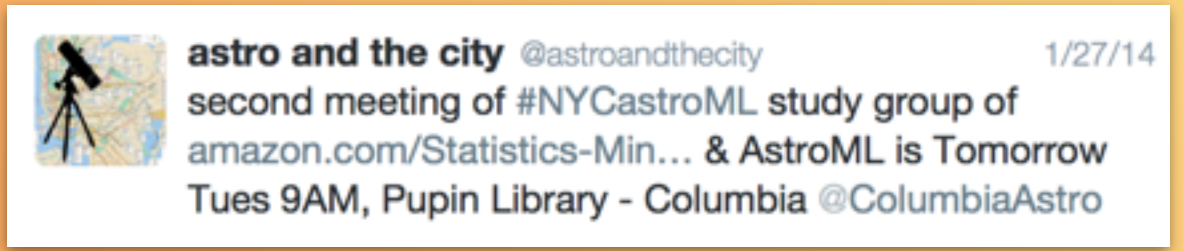
YastroML

Statistics, Data Mining & Machine Learning in Astronomy

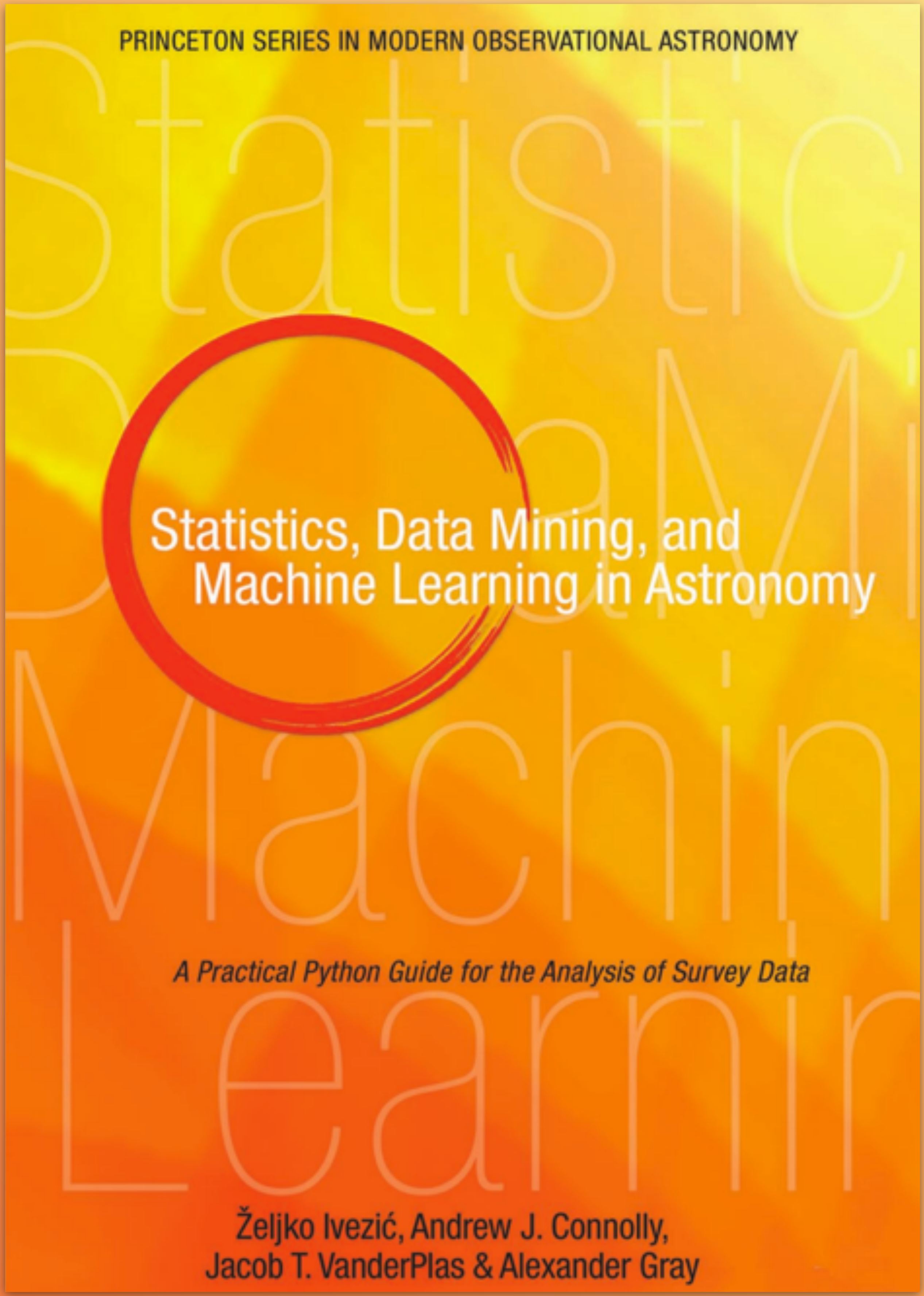
Discussion Group

Matt Giguere



astro and the city @astroandthecity 1/27/14

second meeting of #NYCastroML study group of amazon.com/Statistics-Min… & AstroML is Tomorrow Tues 9AM, Pupin Library - Columbia @ColumbiaAstro



PRINCETON SERIES IN MODERN OBSERVATIONAL ASTRONOMY

Statistics, Data Mining, and Machine Learning in Astronomy

A Practical Python Guide for the Analysis of Survey Data

Željko Ivezić, Andrew J. Connolly,
Jacob T. VanderPlas & Alexander Gray

Statistics

Data Mining

Machine Learning

Statistics Chapters 3-5

Data Mining

Machine Learning

Statistics Chapters 3-5

Data Mining Chapters 6-7

Machine Learning

Statistics	Chapters 3-5
Data Mining	Chapters 6-7
Machine Learning	Chapters 8-10

Statistics

Data Mining

Machine Learning

Statistics

Data Mining

Machine Learning

Statistics

March 28th

3. Probability & Statistical Distributions

- Brief Overview of Probability & Random Variables
- Descriptive Statistics
- Common Univariate Distribution Functions
- The Central Limit Theorem
- Bivariate & Multivariate Distribution Functions
- Correlation Coefficients
- Random Number Generation

Statistics

March 28th

3. Probability & Statistical Distributions

- Brief Overview of Probability & Random Variables
- Descriptive Statistics
- Common Univariate Distribution Functions
- The Central Limit Theorem
- Bivariate & Multivariate Distribution Functions
- Correlation Coefficients
- Random Number Generation

April 4th

4. Classical Statistical Inference

- Classical vs. Bayesian Inference
- Maximum Likelihood Estimation (MLE)
- Goodness of Fit & Model Selection
- ML Applied to Gaussian Mixtures
- Confidence Estimates: Bootstrap & Jackknife
- Hypothesis Testing
- Comparison of Distributions
- Nonparametric Modeling & Histograms
- Selection Effects & Luminosity Function Estimation

Statistics

March 28th

3. Probability & Statistical Distributions

- Brief Overview of Probability & Random Variables
- Descriptive Statistics
- Common Univariate Distribution Functions
- The Central Limit Theorem
- Bivariate & Multivariate Distribution Functions
- Correlation Coefficients
- Random Number Generation

April 4th

4. Classical Statistical Inference

- Classical vs. Bayesian Inference
- Maximum Likelihood Estimation (MLE)
- Goodness of Fit & Model Selection
- ML Applied to Gaussian Mixtures
- Confidence Estimates: Bootstrap & Jackknife
- Hypothesis Testing
- Comparison of Distributions
- Nonparametric Modeling & Histograms
- Selection Effects & Luminosity Function Estimation

April 11th

5. Bayesian Statistical Inference

- Intro to Bayesian Method
- Bayesian Priors
- Bayesian Parameter Uncertainty Quantification
- Bayesian Model Selection
- Nonuniform Priors: Edd, Malm, & LK Biases
- Example: Parametric Estimation
- Example: Model Selection
- Numerical Methods (MCMC)
- Frequentist vs Bayesian Comparison

Statistics

Data Mining

Machine Learning

Statistics

Data Mining

Machine Learning

Data Mining

Data Mining

AKA Knowledge Discovery

Data Mining

AKA Knowledge Discovery

What is it good for?

Data Mining

AKA Knowledge Discovery

What is it good for?

exploratory data analysis

Data Mining

AKA Knowledge Discovery

What is it good for?

exploratory data analysis

What qualitative features do my data possess?

Data Mining

AKA Knowledge Discovery

unsupervised learning

What is it good for?

exploratory data analysis

What qualitative features do my data possess?

Data Mining

April 18th

6. Searching for Structure in Point Data

- Density Estimation
- Clusters in Data
- Correlation Functions

Data Mining

April 25th

7. Dimensionatly and Its Reduction

- The Curse of Dimensionality
- Principal Component Analysis
- Nonnegative Matrix Factorization
- Manifold Learning
- Independent Component Analysis & Projection Pursuit
- ***Which technique to use?!***

April 18th

6. Searching for Structure in Point Data

- Density Estimation
- Clusters in Data
- Correlation Functions

Statistics

Data Mining

Machine Learning

Machine Learning

Machine Learning

prediction

Machine Learning

prediction

supervised learning

Machine Learning

May 2nd

8. Regression & Model Fitting

- Formulation of the problem
- Regression for Linear Models
- Regularization & Penalizing the Likelihood
- Principal Component Regression
- Kernel Regression
- Locally Linear Regression
- Nonlinear Regression
- Uncertainties in the Data
- Regression that is Robust to Outliers
- Gaussian Process Regression
- Overfitting, Underfitting, and Cross-validation
- ***Which method to choose?!***

Machine Learning

May 9th

9. Classification

- Assigning Categories
- Generative Classification
- K-Nearest-Neighbor Classifier
- Discriminative Classification
- Support Vector Machines
- Decision Trees
- Evaluating Classifiers: ROC Curves
- ***Which Classifier to use?!***

May 2nd

8. Regression & Model Fitting

- Formulation of the problem
- Regression for Linear Models
- Regularization & Penalizing the Likeliho
- Principal Component Regression
- Kernel Regression
- Locally Linear Regression
- Nonlinear Regression
- Uncertainties in the Data
- Regression that is Robust to O
- ussian Process Regressio
- ing, Underfitting, an
- ***to choo***

Machine Learning

May 16th
10. Time Series Analysis

- Main Concepts
- Modeling Toolkit
- Analysis of Periodic Time Series
- Temporally Localized Signals
- Analysis of Stochastic Processes
- ***Which Method to use?!***

May 9th
9. Classification

- Assigning Categories
- Generative Classification
- K-Nearest-Neighbor Classifier
- Discriminative Classification
- Support Vector Machines
- ecision Trees
- uating Classifiers
- ***Classifier t***

May 2nd
8. Regression & Model Fitting

- Formulation of the problem
- Regression for Linear Models
- Regularization & Penalizing the Likelihoo
- Principal Component Regression
- Kernel Regression
- Locally Linear Regression
- Nonlinear Regression
- Uncertainties in the Data
- Regression that is Robust to O
- ussian Process Regressio
- ting, Underfitting, an
- ***to choo***

YastroML

Introduction

- Group wiki: cod.al/yams
- Group repository: github.com/YastroML