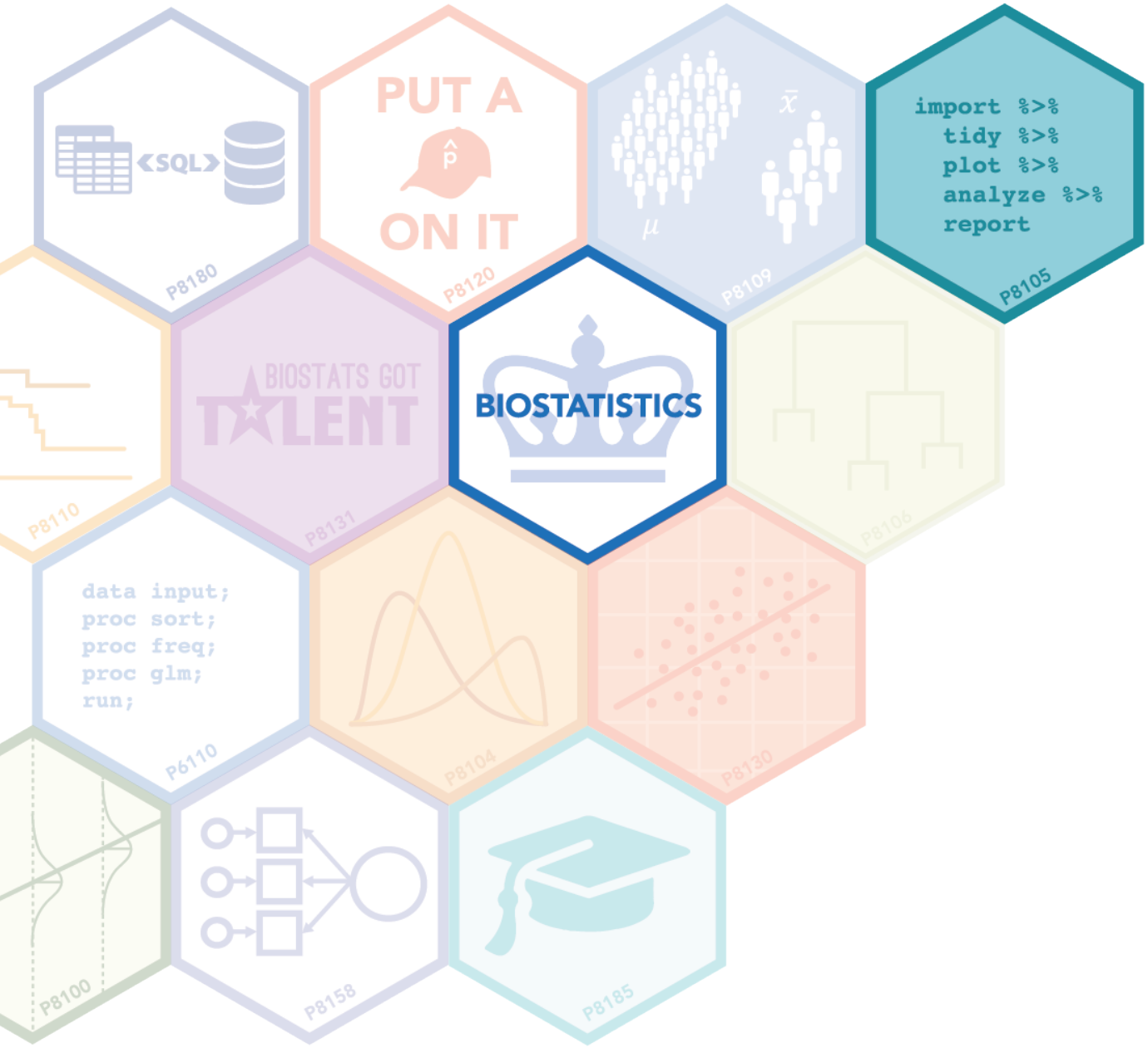


STATISTICAL LEARNING

Jeff Goldsmith, PhD
Department of Biostatistics

Statistical learning

- "Data science" is often associated with statistical learning
  - AKA machine learning, sometimes "AI"

- Becoming very popular…

Statistical learning

- "Data science" is often associated with statistical learning
  - AKA machine learning, sometimes "AI"

- Becoming very popular…

February 9, 2016

Machine Learning and the Profession of Medicine

Alison M. Darcy, PhD[1]; Alan K. Louie, MD[1]; Laura Weiss Roberts, MD, MA[1]

» Author Affiliations | Article Information

JAMA. 2016;315(6):551-552. doi:10.1001/jama.2015.18421

Statistical learning

- "Data sciences" is often associated with statistical learning
  - Al
- Beco

Previously unimaginable opportunities to apply machine learning to the care of individual patients in medical practice are currently within grasp.

Machine Learning and the Profession of Medicine

Alison M. Darcy, PhD[1]; Alan K. Louie, MD[1]; Laura Weiss Roberts, MD, MA[1]

» Author Affiliations | Article Information

JAMA. 2016;315(6):551-552. doi:10.1001/jama.2015.18421

Statistical learning

- "Data sciences" is often associated with statistical learning
  - Al
- Beco

Previously unimaginable opportunities to apply machine learning to the care of individual patients in medical practice are currently within grasp.

Machine

cine

Alison M. Darcy, PhI

» Author Affiliation

JAMA. 2016;315(6):

medical practice are currently within grasp. Machine learning algorithms can accommodate different configurations of raw data, assign context weighting, and calculate the predictive power of every combination of variables available to assess diagnostic and prognostic elements. The applications of machine learning

Statistical learning vs statistics

- Helpful to view statistical learning as part of a spectrum of tools

Statistical learning vs statistics

- Helpful to view statistical learning as part of a spectrum of tools

April 3, 2018

Big Data and Machine Learning in Health Care

Andrew L. Beam, PhD[1]; Isaac S. Kohane, MD, PhD[1]

» Author Affiliations | Article Information

JAMA. 2018;319(13):1317-1318. doi:10.1001/jama.2017.18391

Statistical learning vs statistics

- Helpful to view statistical learning as part of a spectrum of tools

April 3, 2018

Big Data and Machine Learning in Health Care

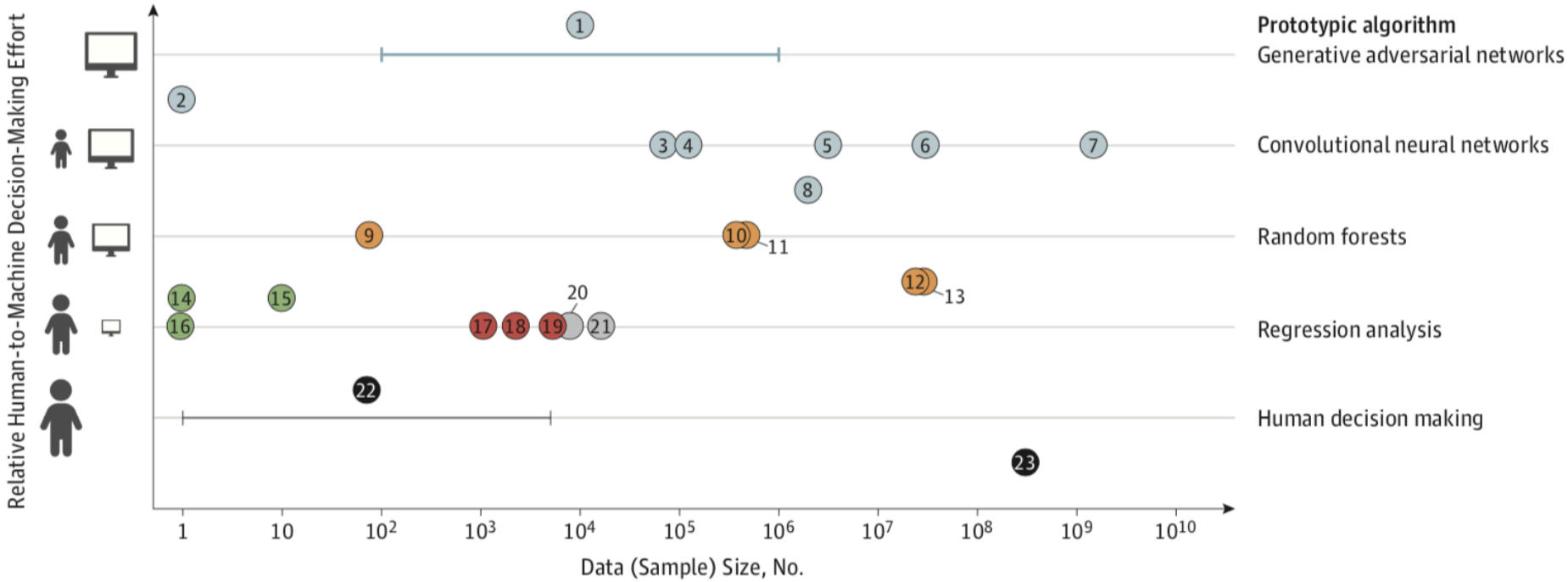
Andrew L. Beam, PhD[1]; Isaac S. Kohane, MD, PhD[1]

» Author Affiliations | Article Information

JAMA. 2018;319(13):1317-1318. doi:10.1001/jama.2017.18391

that perform on par with human physicians.[1] Though machine learning and big data may seem mysterious at first, they are in fact deeply related to traditional statistical models that are recognizable to most clinicians. It is our hope

Statistical learning spectrum



Risk calculators

17 CHA_2DS_2-VASc Score for atrial fibrillation stroke risk (2017)

18 MELD end-stage liver disease risk score (2001)

19 Framingham CV risk score (1998)

Randomized Clinical Trials

20 Celecoxib vs nonsteroidal anti-inflammatory drugs for osteoarthritis and rheumatoid arthritis (2002)

21 Use of estrogen plus progestin in healthy postmenopausal women (2002)

Other

22 Clinical wisdom

23 Mortality rate estimates from US Census (2010)

Beam and Kohane, 2018

Learning from data

- Supervised learning
  - There's an outcome you care about, and what you learn depends on that outcome
  - Regression, lasso / elastic net, regression trees, support vector machines …

- Unsupervised learning
  - You just have data and want to learn stuff – probably find patterns or identify subgroups
  - Clustering, principal components, factor analysis …

Regression

- Regression (linear, logistic, etc) is interested in the conditional distribution of an outcome Y given some predictors x

- Common form (continuous outcome):

$$E(Y|x) = b_0 + b_1x$$

- Regression has a lot of benefits, including:
  - Common understanding
  - Interpretable coefficients
  - Inference / p-values

Regression → Lasso

- One drawback of regression is lack of scalability
  - When you have some covariates, you have model-building options
  - When you have a lot of covariates, you have fewer options

- Lasso is useful when you have a lot of coefficients and few strong hypotheses
  - Goal is a regression-like model that "automatically" selects variables

Regression → Lasso

- Regression is estimated using the data likelihood:

$$\min_{\mathbf{b}} [RSS(\mathbf{b})]$$

- Lasso adds a penalty on the sum of all coefficients

$$\min_{\mathbf{b}} [RSS(\mathbf{b}) + \lambda p(\mathbf{b})]$$

- Estimation is now a balance between overall fit and coefficient size
  - Roughly the same is true in other regression models

Lasso

$$\min_{\mathbf{b}} \left[RSS(\mathbf{b}) + \lambda p(\mathbf{b}) \right]$$

- Penalized estimation forces some coefficients to be 0, which effectively removes some covariates from the model
- Result has a similar form to regression
  - Can get predicted values based on covariates

Lasso

- There are also some drawbacks:
  - No inference / p-values
  - Very different interpretation (if any)
  - Have to choose the tuning parameter (to maximize prediction accuracy)
  - Coefficients for included covariates is not the same as in a regression using only those covariates

These drawbacks are roughly similar across statistical learning methods

Tuning parameter selection

$$\min_{\mathbf{b}} \left[RSS(\mathbf{b}) + \lambda p(\mathbf{b}) \right]$$

- For any tuning parameter value, Lasso returns coefficient estimates
- These can be used to produce predicted values based on covariates

- Tuning parameters are frequently chosen using cross validation
  - Split the data into training and testing sets
  - Fit Lasso for a fixed tuning parameter using training data
  - Compare observations to predictions using testing data
  - Repeat for many possible tuning parameter values
  - Pick the tuning parameter that gives the best predictions for "held out" testing data

Clustering

- Broad collection of techniques that try to find data-driven subgroups
  - Subgroups are non-overlapping, and every data point is in one subgroup
  - Data points in the same subgroup are more similar to each other than to points in another subgroup

- Have to define “similarity” …
- You can usually tell if clustering worked if it looks right

- Lots of methods; we’ll look at k-means

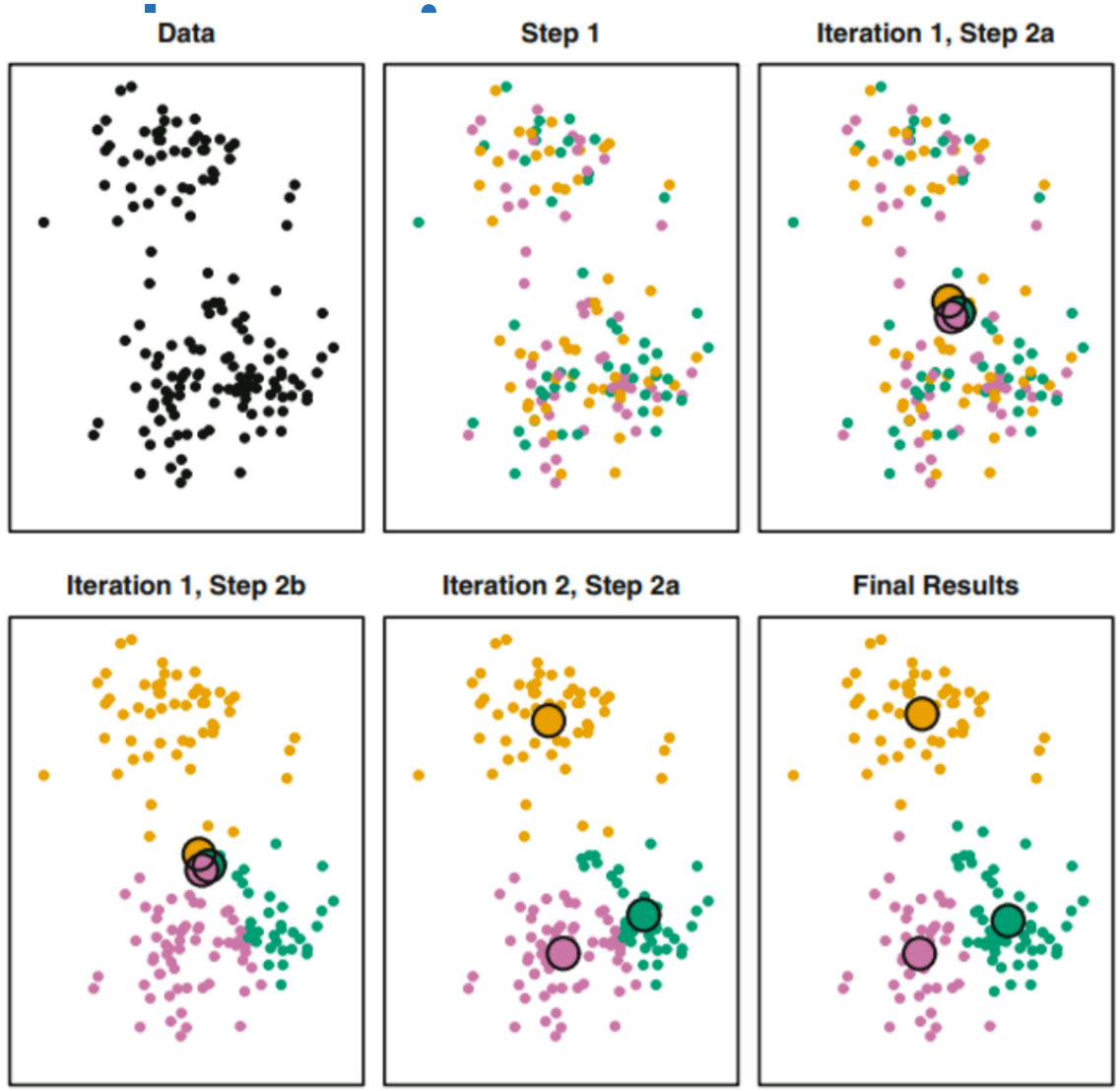
K-means clustering

- In a nutshell:
  - Assume there are k groups, each with it's own mean ("centroid")
  - Put all data points in a group at random
  - Alternate between two steps:
    - Recompute group mean
    - Reassign points to the cluster with the closest centroid
  - Stop when things stop

- Not a lot of guarantees here...

K-mean

- In a nutshe
  - Assume
  - Put all
  - Alterna
    - Rec
    - Reas
  - Stop wh
- Not a lot o



an (“centroid”)

centroid

ISLR Ch 10

K-mean

- In a nutshe
  - Assume an (“centroid”)
  - Put all c
  - Alterna
    - Rec
    - Reas centroid
  - Stop wh
- Not a lot o

ISLR Ch 10