

Machine learning algorithms

PCA, #2

2020-10-5

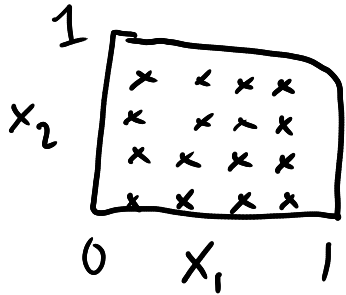
CSCI 471 / 571, Fall 2020

Kameron Decker Harris

Recommended talk!

- Nathan Kutz, 3 p.m. on Thursday
- Western Washington Data-driven Discovery Seminar Series
- *Machine Learning for Science: Data-Driven Discovery Methods for Governing equations, Coordinates and Sensors*

Homework questions?



total # = n_{test}

16, 64, 144

4x4, 8x8, 12x12

for x_1 in `np.linspace(0, 1, 4)`

for x_2 . . .

`vec_x = np.array([x1, x2])`

Problem II

$f(\vec{x})$

`true_fun(x)`

`pred_fun(x, beta)`

$\underbrace{\text{length}} \leftrightarrow \text{degree}$

`beta = ...`

`pred_wrap(x) = lambda x: pred_fun(x, beta)`

Problem 6 $\|X\vec{\beta} - \vec{y}\|^2 = \vec{u}^T A \vec{u} + \vec{v}^T B \vec{u} + C$

$X, \vec{\beta}, \vec{y}$

will depend
on one of the
least square terms

$\|\vec{v}\|^2 =$ inner product ... rules of transposes, etc.

Problem 10

$\vec{x} = \begin{pmatrix} a \\ b \end{pmatrix}$

$\begin{bmatrix} a & b \end{bmatrix}$

$p=3$

$\begin{bmatrix} 1 & \overbrace{a}^{p=1} & b & \overbrace{a^2 \ b^2}^{p=2} & \dots \end{bmatrix}$

degree - 0

10.2 : take in \sum
output F

Problem 7(7.1) write in $f(\vec{u}, \vec{v})$

use $\#u$ to get
form
gradient

(7.2) $\text{grad} = 0$

$M\vec{\beta} = \vec{c}$

Problem 4

$A \in \mathbb{R}^{m \times n}$

same as

"A is $m \times n$ matrix"

$\vec{x} \in \mathbb{R}^n$

" \vec{x} is n -dim vector (column)"

PCA, SVD Explained variance

$$X = U \Sigma V^T$$

Σ singular values, largest \rightarrow smallest

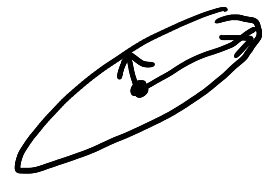
σ_i^2 = variance captured by i^{th} component

$$\frac{\sum_{i=1}^3 \sigma_i^2}{\sum_{j=1}^{\text{rank}(X)} \sigma_j^2} = 80\% \quad (\text{for example})$$

variance explained
by top 3 components

$$X \rightarrow X'$$

Projection



Dimensionality reduction to r dimensions $< d$

$$V_r = \begin{bmatrix} | & & | \\ \vec{v}_1 & \dots & \vec{v}_r \\ | & & | \end{bmatrix} \leftarrow \vec{v}_i \in \mathbb{R}^d$$

$$(r=1, 2, 3)$$

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix}, \vec{x}^T = [x_1 \dots x_d]$$

$$\underset{(n \times d)}{X} \underset{(d \times r)}{V_r} = \underset{(n \times r)}{X'} = \begin{bmatrix} -\vec{x}_1'^T- \\ \vdots \\ -\vec{x}_n'^T- \end{bmatrix}$$

$$\vec{x}_i' \in \mathbb{R}^r$$

$$i=1, \dots, n$$

$$X' = U \underbrace{S V^T}_{V_r}$$

$$X = \begin{bmatrix} -\vec{x}_1'^T- \\ \vdots \\ -\vec{x}_n'^T- \end{bmatrix}$$

Examples

- Super popular 3.6 M
 - Search “PCA”, “principal components” on Scholar
 - Credit for some of slides: Bing Brunton

The Shapes of Stories

by Kurt Vonnegut

Kurt Vonnegut gained worldwide fame and adoration through the publication of his novels, including *Slaughterhouse-Five*, *Cat's Cradle*, *Breakfast of Champions*, and more.

But it was his rejected master's thesis in anthropology that he called his prettiest contribution to his culture.

The basic idea of his thesis was that a story's main character has ups and downs that can be graphed to reveal the story's shape.

The shape of a society's stories, he said, is at least as interesting as the shape of its pots or spearheads. Let's have a look.

Designer: Maya Eitam, www.mayaaitam.com
Sources: *A Man without a Country* and *Palm Sunday* by Kurt Vonnegut



Man in Hole



The main character gets into trouble then gets out of it again and ends up better off for the experience.

Arsenic and Old Lace
 Harold & Kumar Go To White Castle

Boy Meets Girl



The main character comes across something wonderful, gets it, loses it, then gets it back forever.

Jane Eyre
 Eternal Sunshine of the Spotless Mind

From Bad to Worse



The main character starts off poorly then gets continually worse with no hope for improvement.

The Metamorphosis
 The Twilight Zone

Which Way Is Up?



The story has a lifelike ambiguity that keeps us from knowing if new developments are good or bad.

Hamlet
 The Sopranos

Creation Story



In many cultures' creation stories, humankind receives incremental gifts from a deity. First major staples like the earth and sky, then smaller things like sparrows and cell phones. Not a common shape for Western stories, however.

Old Testament



Humankind receives incremental gifts from a deity, but is suddenly ousted from good standing in a fall of enormous proportions.

Great Expectations

New Testament



Humankind receives incremental gifts from a deity, but is suddenly ousted from good standing, but then receives off-the-charts bliss.

Great Expectations with Dickens' alternate ending

Cinderella



It was the similarity between the shapes of Cinderella and the New Testament that thrilled Vonnegut for the first time in 1947 and then over the course of his life as he continued to write essays and give lectures on the shapes of stories.

Man in Hole



The main character gets into trouble then gets out of it again and ends up better off for the experience.



Arsenic and Old Lace



Harold & Kumar Go To White Castle

Boy Meets Girl



The main character comes across something wonderful, gets it, loses it, then gets it back forever.



Jane Eyre



Eternal Sunshine of the Spotless Mind

From Bad to Worse



The main character starts off poorly then gets continually worse with no hope for improvement.



The Metamorphosis



The Twilight Zone

Which Way Is Up?



The story has a lifelike ambiguity that keeps us from knowing if new developments are good or bad.



Hamlet



The Sopranos

sentiment
vs. progress

X = books



"Fairy Tale Castle," by Mikalojus Konstantinas Čiurlionis (1909)

Wikimedia

The Six Main Arcs in Storytelling, as Identified by an A.I.

A machine mapped the most frequently used emotional trajectories in fiction, and compared them with the ones readers like best.

ADRIENNE LAFRANCE | JUL 12, 2016 | TECHNOLOGY

<https://www.theatlantic.com/technology/archive/2016/07/the-six-main-arcs-in-storytelling-identified-by-a-computer/490733/>

The emotional arcs of stories are dominated by six basic shapes

Andrew J Reagan , Lewis Mitchell, Dilan Kiley, Christopher M Danforth and Peter Sheridan Dodds

EPJ Data Science 2016 5:31 | DOI: 10.1140/epjds/s13688-016-0093-1 | © Reagan et al. 2016

Received: 2 July 2016 | Accepted: 26 October 2016 | Published: 4 November 2016

Harry Potter and the Deathly Hallows by J.K. Rowling

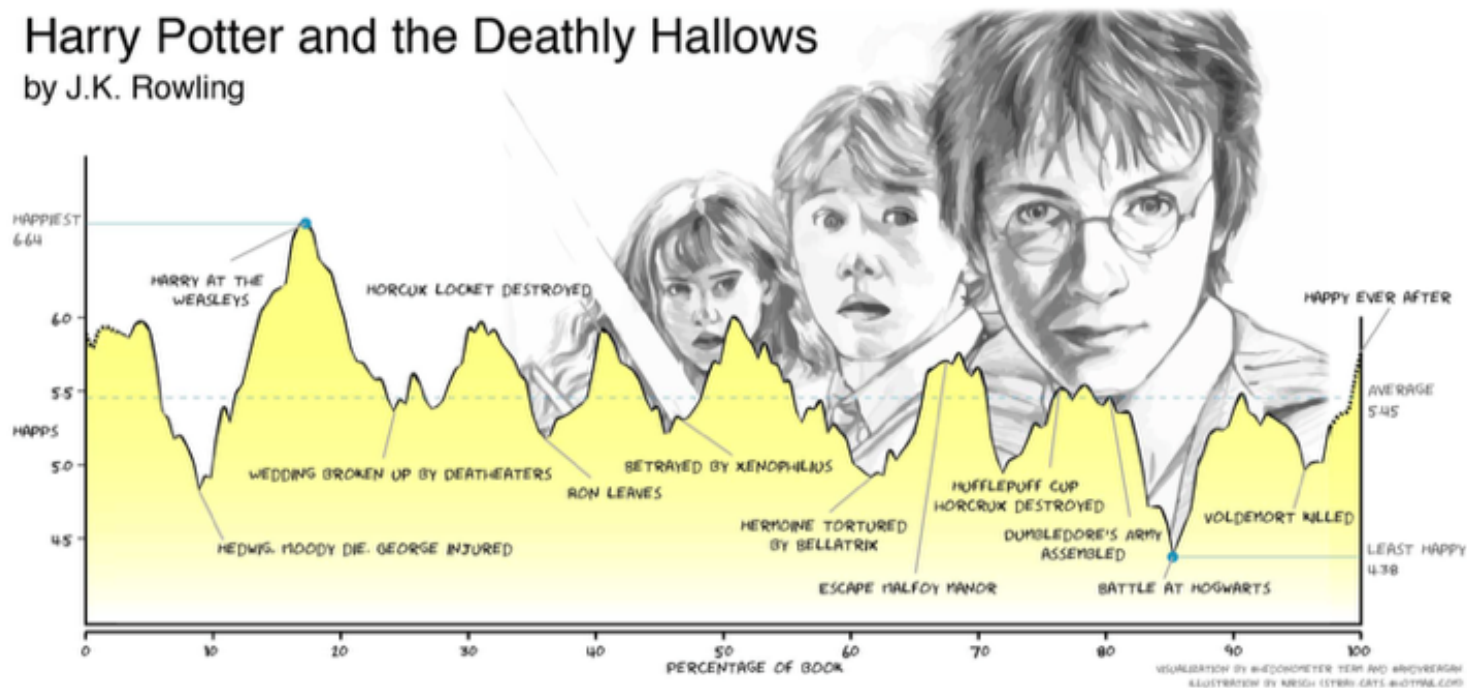


Fig 2

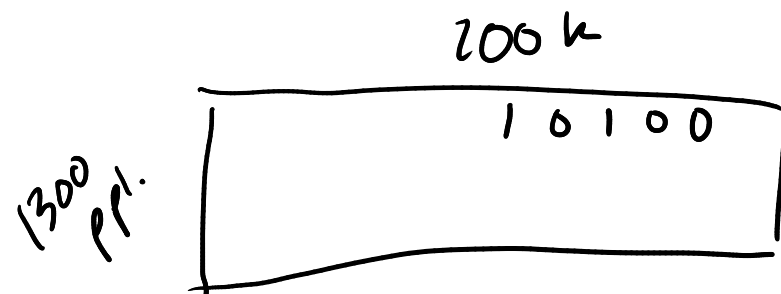
LETTERS

Genes mirror geography within Europe

John Novembre^{1,2}, Toby Johnson^{4,5,6}, Katarzyna Bryc⁷, Zoltán Kutalik^{4,6}, Adam R. Boyko⁷, Adam Auton⁷, Amit Indap⁷, Karen S. King⁸, Sven Bergmann^{4,6}, Matthew R. Nelson⁸, Matthew Stephens^{2,3} & Carlos D. Bustamante⁷

3,192 European individuals, PORES project
each genotyped at 500,568 SNP loci

after stringent selection criteria,
1,387 individuals
197,145 loci



$m = 197,145$ measurements

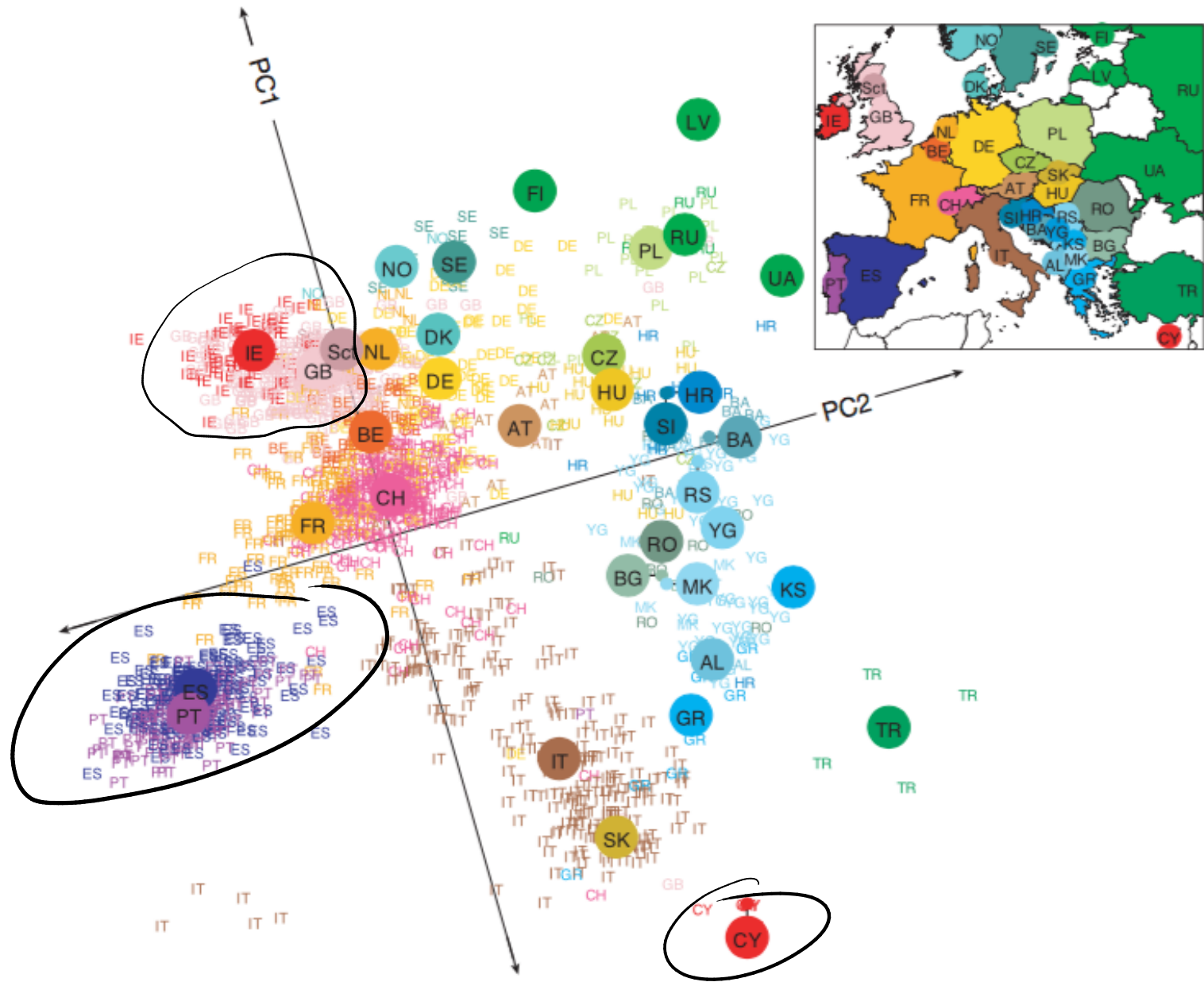
$n = 1,387$ people



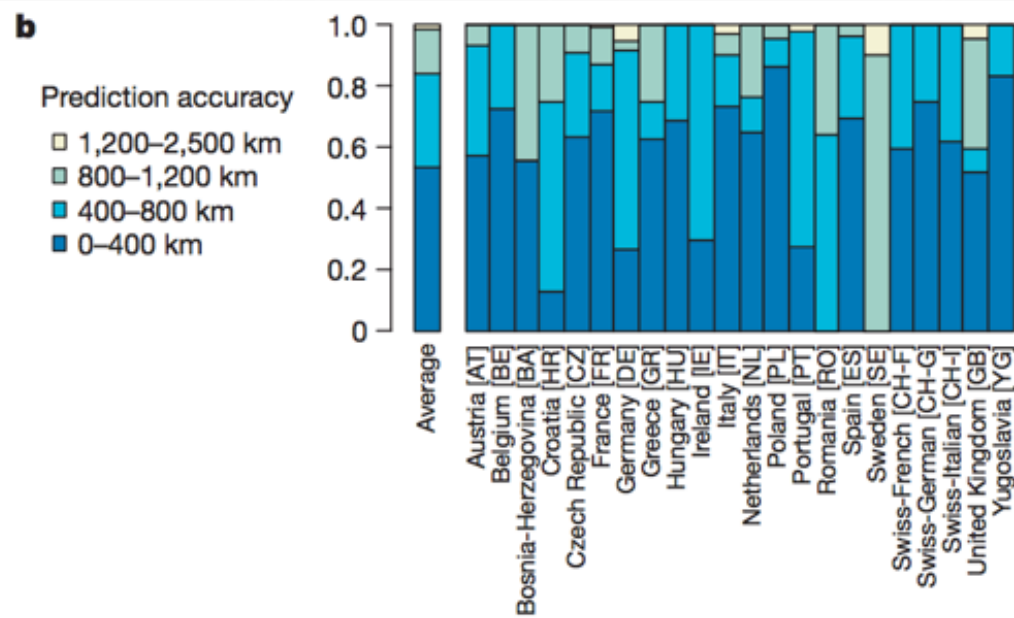
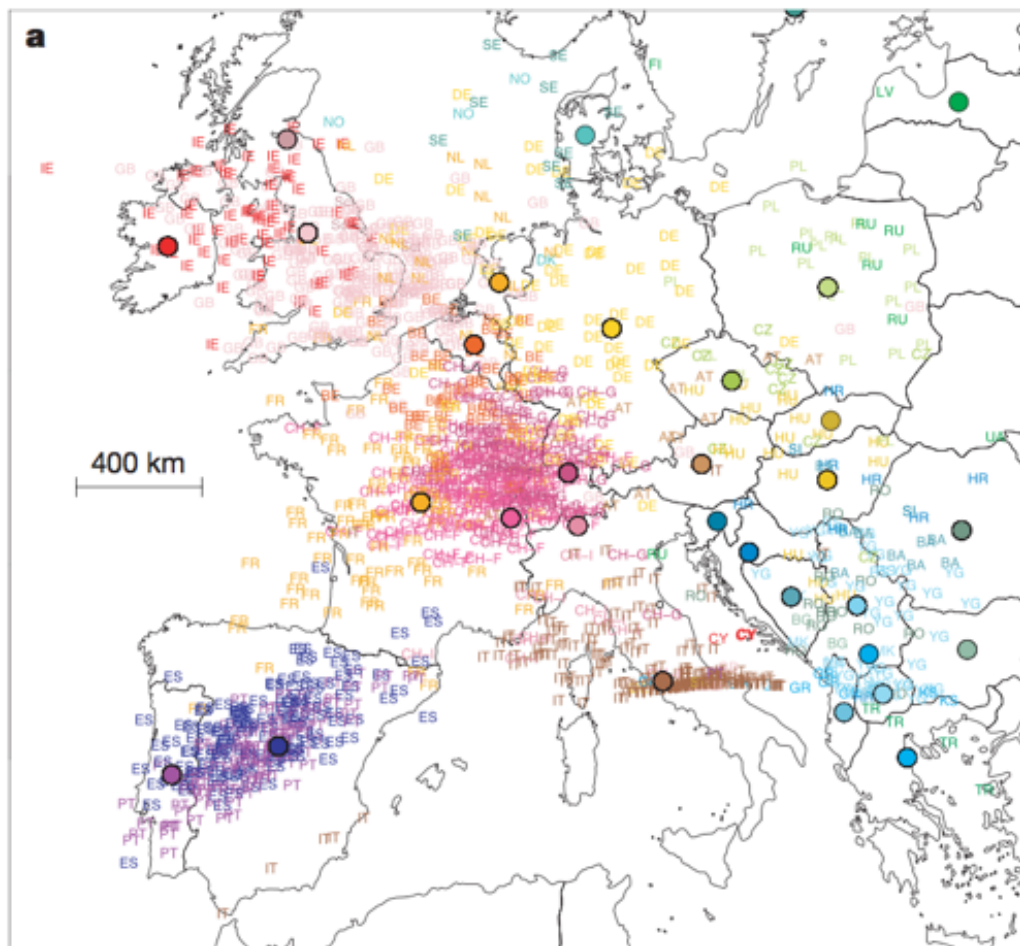
Data Matrix

that's a huge amount of high-dimensional data!
we can't even look at that sensibly...

a



Cross-validated Accuracy



Data from this paper:

https://github.com/NovembreLab/Novembre_etal_2008_misc

E-ARTICLE

Radiating despite a Lack of Character: Ecological Divergence among Closely Related, Morphologically Similar Honeyeaters (Aves: Meliphagidae) Co-occurring in Arid Australian Environments

Eliot T. Miller,^{1,2,3,4,*} Sarah K. Wagner,^{4,5} Luke J. Harmon,^{3,6} and Robert E. Ricklefs¹

1. Harris World Ecology Center and Department of Biology, University of Missouri, St. Louis, Missouri 63121; 2. Department of Biological Sciences, Macquarie University, Sydney, New South Wales 2109, Australia; 3. Department of Biological Sciences, University of Idaho, Moscow, Idaho 83844; 4. Cornell Lab of Ornithology, Cornell University, Ithaca, New York 14950; 5. Ecology and Evolutionary Biology Department, University of Colorado, Boulder, Colorado 80309; 6. Institute for Bioinformatics and Evolutionary Studies (IBEST), University of Idaho, Moscow, Idaho 83844

Submitted December 11, 2015; Accepted September 14, 2016; Electronically published December 20, 2016

Online enhancements: appendix, supplemental PDF. *Dryad data:* <http://dx.doi.org/10.5061/dryad.r82jg>.

