HUDK 4050:CORE METHODS IN EDM

10/25/16 8:29 AM

In the news

The Washington Post

Google Has Quietly Dropped Ban on Personally Identifiable Web Tracking

PRO PUBLICA

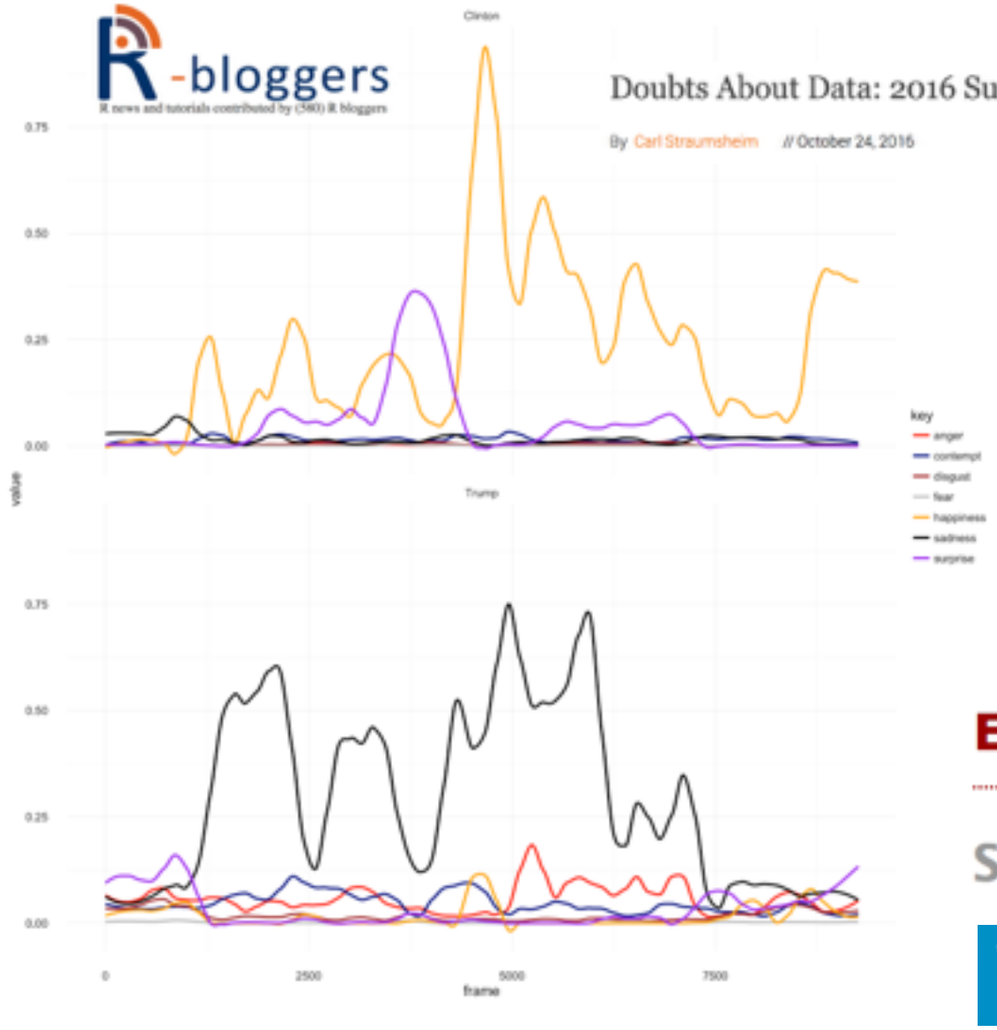
Asia & Pacific

China's plan to organize its society relies on 'big data' to rate everyone

https://support.google.com/websearch/answer/465

R-bloggers

R news and tutorials contributed by (580) R bloggers



Doubts About Data: 2016 Survey of Faculty Attitudes on Technology

By Carl Straumsheim // October 24, 2016

INSIDE HIGHER ED

EDUCATION WEEK

Parents Bullish on Ed Tech, Skeptical About Its Implementation, Survey Says

By Benjamin Herold on October 20, 2016 1:53 PM

Tune into Tableau Conference Live—starting November 8

EdLab Seminar: CourseLab

Start: 10/26/2016 – 12:00pm

TABLEAU CONFERENCE 2016

VR In Education

Meetup

https://www.rdocumentation.org/

Anonymous Mid-Semester Check In

http://bit.ly/2f2qxgQ

Assessment

- Meet with me once
- Github contains all assignments
- Zotero contains notes on readings & assignments
- Ask question on Stack Overflow

Principal Component Analysis

Grouping stuff

By Variables

ID	Var1	Var2	Var3
A			
B			
C			
D			

ID	Var2
A	
B	
C	
D	

Selection

ID	Var2+3
A	
B	
C	
D	

Extraction

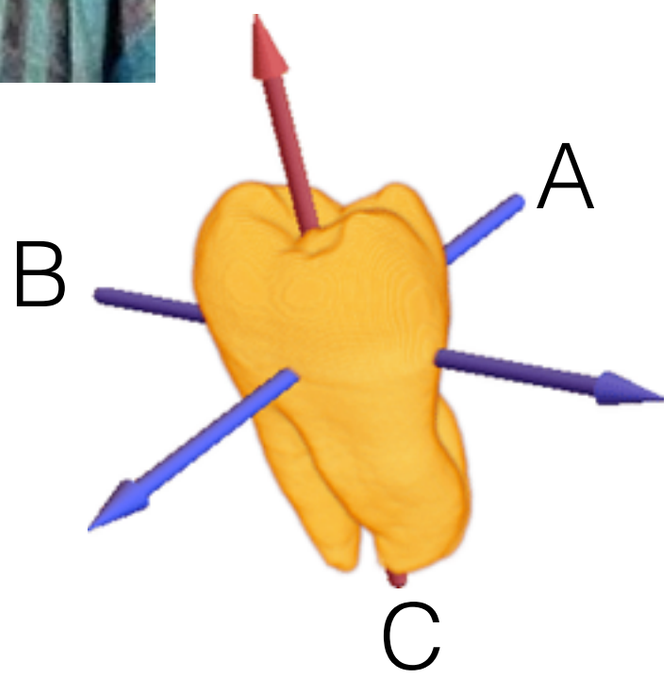
By People

ID	Var1	Var2	Var3
A			
C			

ID	Var1	Var2	Var3
B			
D			

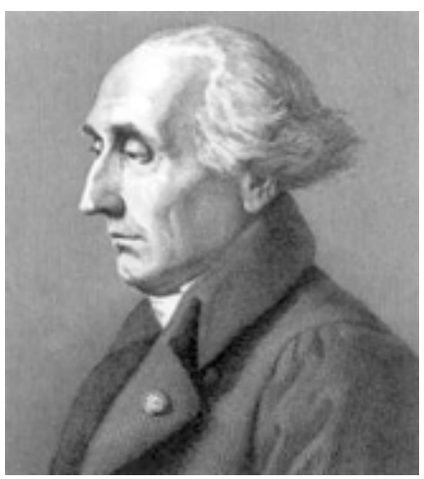
History

- Part of a set of issues called “Eigen Problems”
- Arose as a subset of phenomena related to differential equations (Your old buddy Euler, c.1750)
- Principal Axes



History

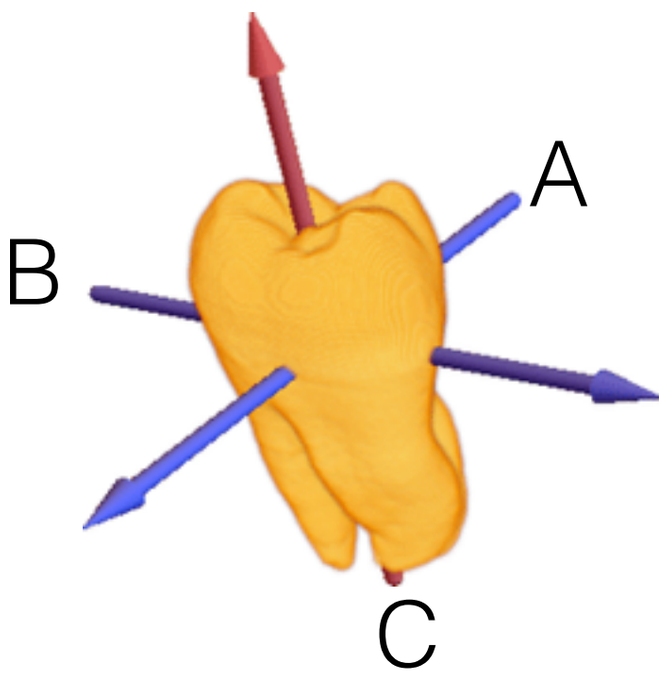
- Describe inertia as a matrix of measurements from the center of an object as it moves



$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$$

- Principal axes = the lines through which you can describe the object, while maximizing the amount of variation maintained

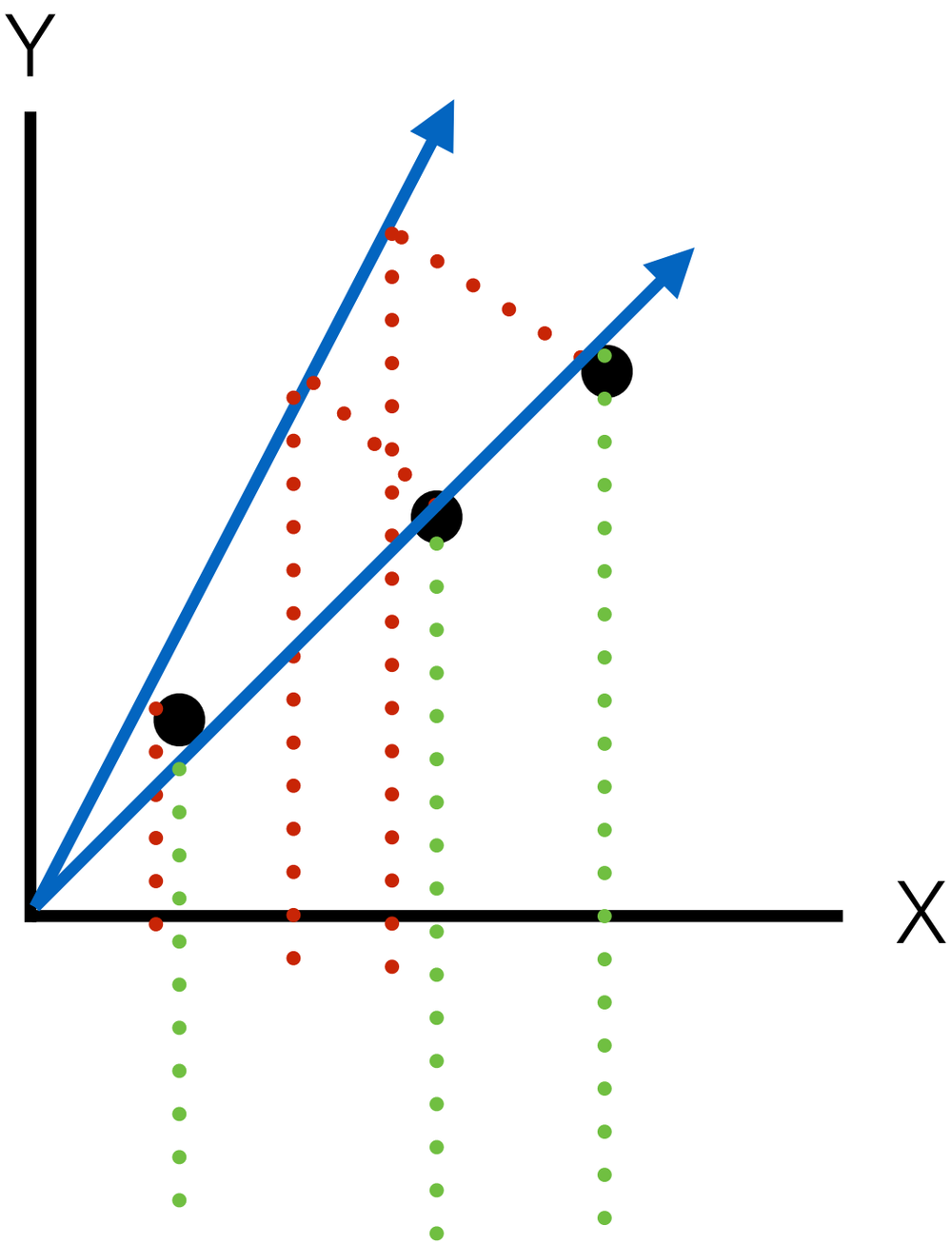
(Joseph-Louis Lagrange)

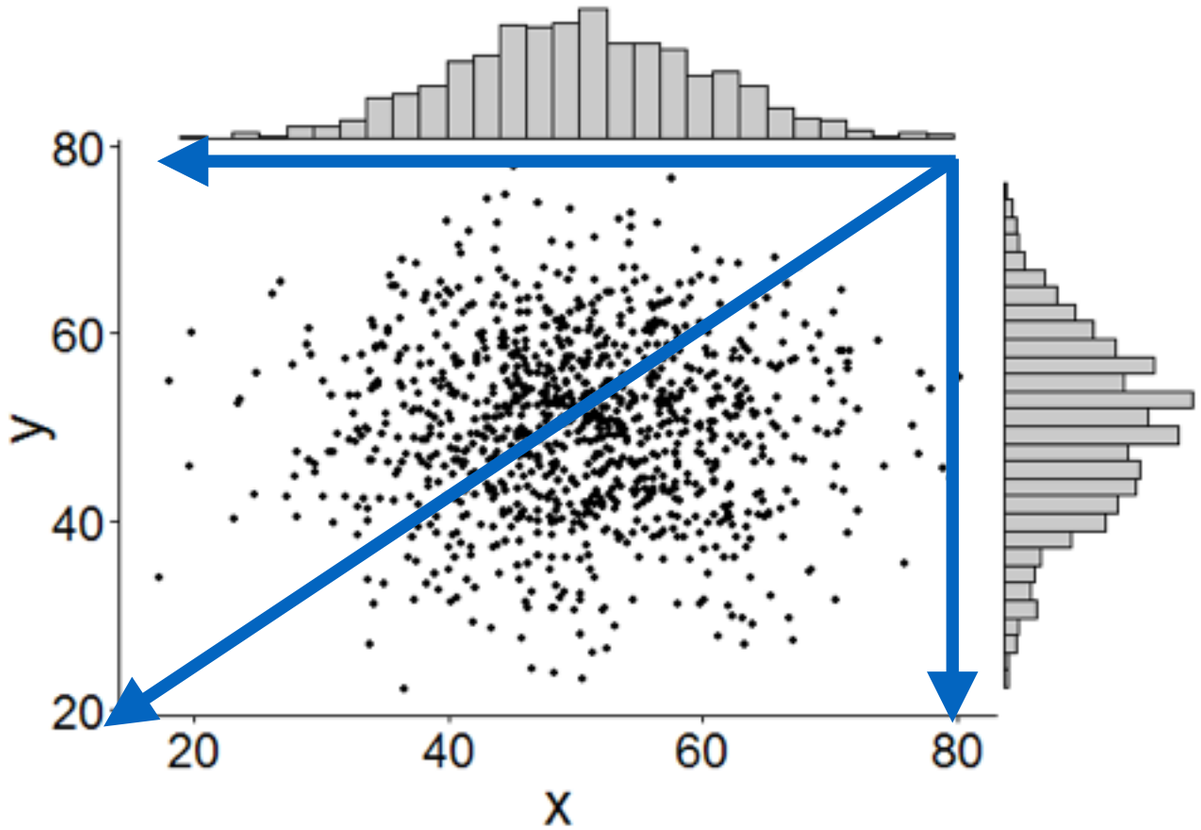


Yada, yada, yada...

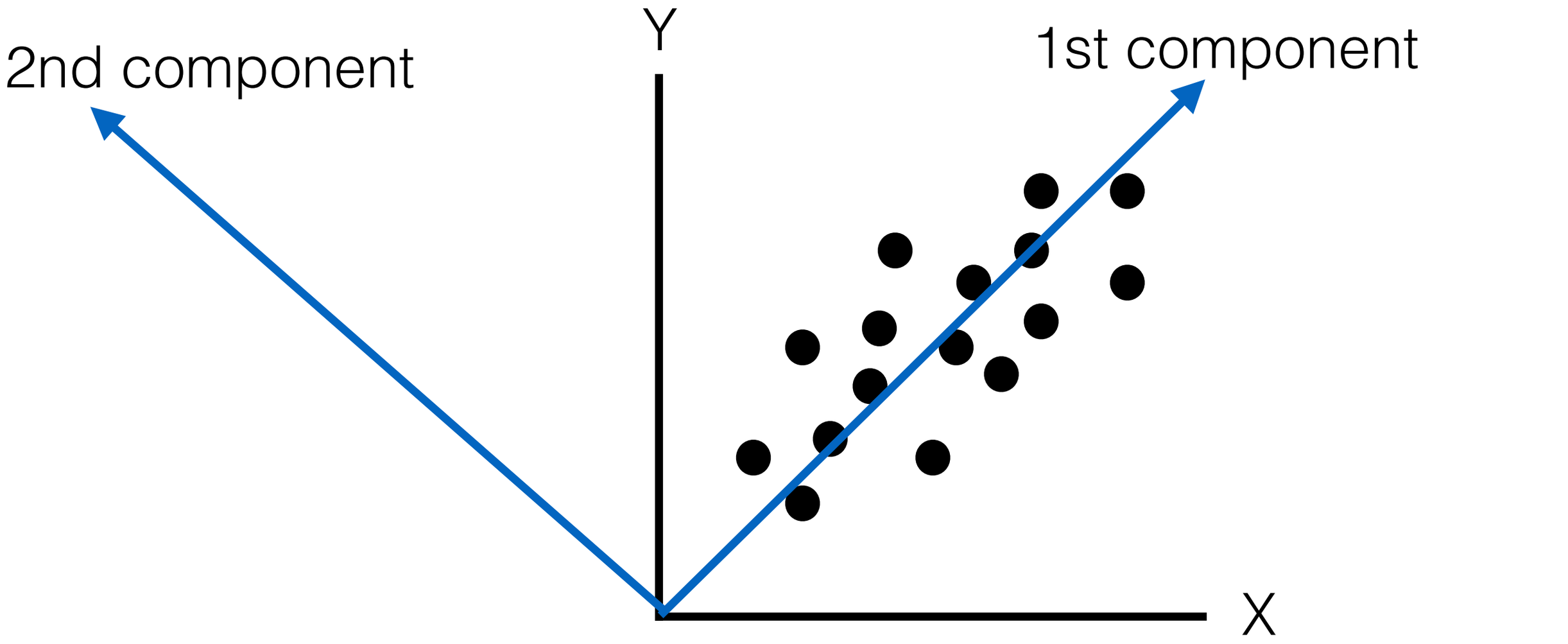
Google

PCA is about Finding the Direction of Maximal Variance

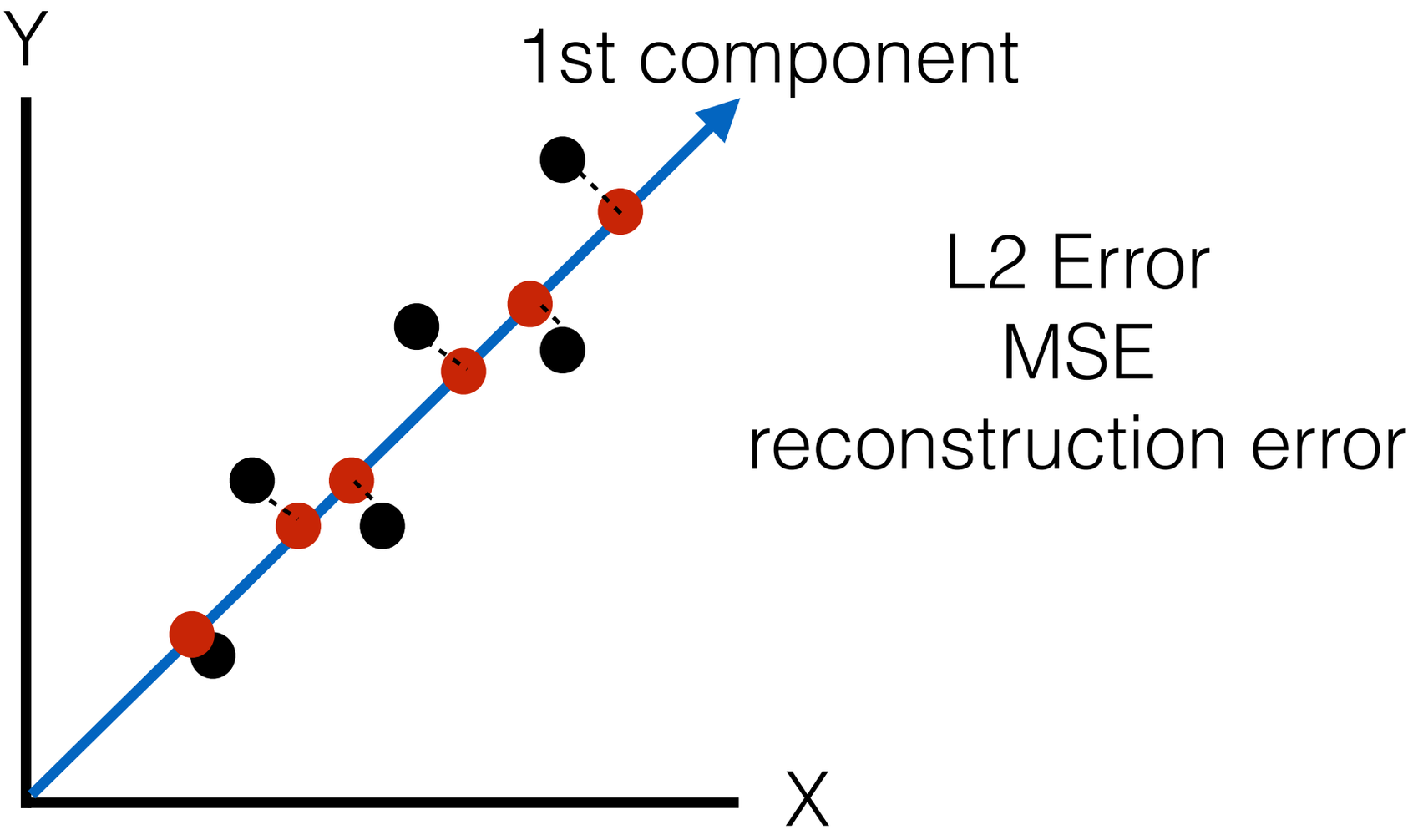




Global Constraint: Orthogonal Components



- “Best” reconstruction of the data (because not really doing anything)
- But also true for linear reconstruction of the data



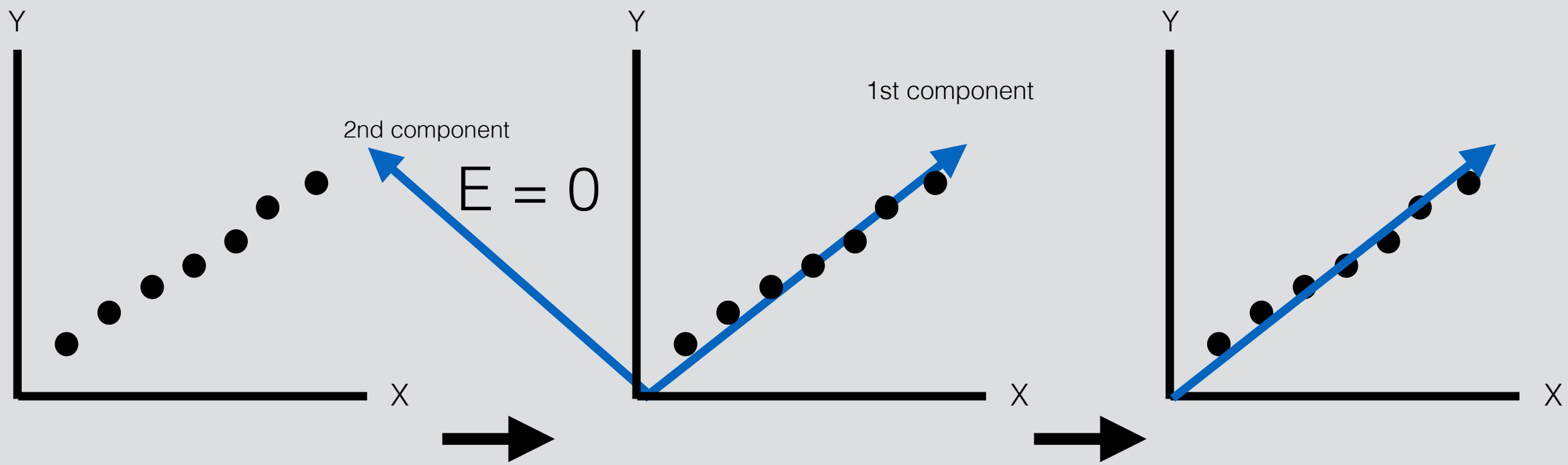
Component is a description of X & Y

Eigenvalues

- Every component has an associated eigenvalue
- Eigen- = “characteristic”
- Created when linear transformations are applied to a matrix
- Take away: the size of the eigenvalue is relative to how well the component maximizes variance

Feature Selection

- If a component has an eigenvalue of zero = non-informative (will not effect reconstruction error)
- Therefore, we can delete it = reduce features



Questions?

Assignment 4 Part A