HUDK 4050: CORE METHODS IN EDM

10/22/19 1:56 PM

In the news

Forbes

The Classroom Connectivity Gap Is Closed. How Did That Happen?

'The online test preparation market will be the fastest-growing within Ed-tech'

THE TIMES OF INDIA

Forbes

Computer Vision Can Transform Education

A bipartisan group of senators wants to help you leave Facebook

CNBC

stackoverflow

Make Stack Overflow More Welcoming

We launched a new Code of Conduct in August that reinforces our commitment to mutual respect and kindness.

Read the Code of Conduct

We improved flagging so it's simpler to report comments that are abusive or unkind. We also introduced the New Contributor Indicator to make it easier to identify and respond to new users.

Events

Title	Date - Time	Location
Where on Earth is AI Headed?	10/25 - 2:00pm	Davis Auditorium
Columbia Nano Day	10/29	Schapiro CEPSR
Race After Technology	10/30 - 6:00pm	Online
Robotics to Retrain & Restore Human Movements	11/1 - 12:00pm	NWB Rm 1406
The Color of Surveillance	11/7	Georgetown University
All Tech is Human: NYC	11/9	ThoughtWorks
Science Communication Workshop	11/20 - 9:30am	Low Library

Anonymous Check In

bit.ly/HUDK4050-Checkin

Assessment

- One assignment due per week for the rest of the semester
- Ask question on Stack Overflow (pull requisition to /so-question)
- Final group project will be a video
- Rate project videos

Class 27 - Work Session: Assignment 8, Group Project (12/5/19)

Class 28 - Work Session: Assignment 8, Group Project (12/10/19)

Due: Assignment 7 - Diagnostic Metrics

Class 29 - Rate video presentations (12/12/19)

Class 30 - Rate video presentations (12/17/19)

EVERYTHING DUE - 12/19/19

Adaptive Systems

Adaptive

- Originally = assistive
- ~1990s = sequential estimate of aptitude (IRT)
- ~2012 = a system that adapts the educational environment according to students' learning needs
- Distinct from Intelligent Tutors in terms of methods employed

Adaptive Systems

NETFLIX

amazon.com

PANDORA

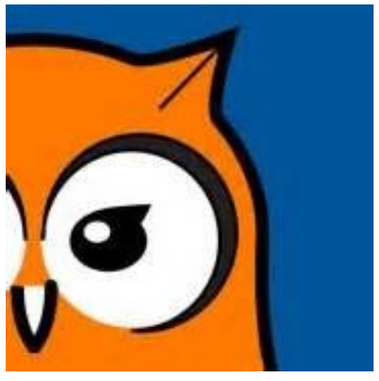
last.fm

hulu

LinkedIn

Adaptive Engines

KNEWTON



Grockit

prepme

KAPLAN

adapt courseware

gymglish

Recommender Systems

Collaborative filter: build a model from a user's past behavior + similar decisions made by other users

Content filter: utilize a series of discrete characteristics of an item in order to recommend additional items with similar properties

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

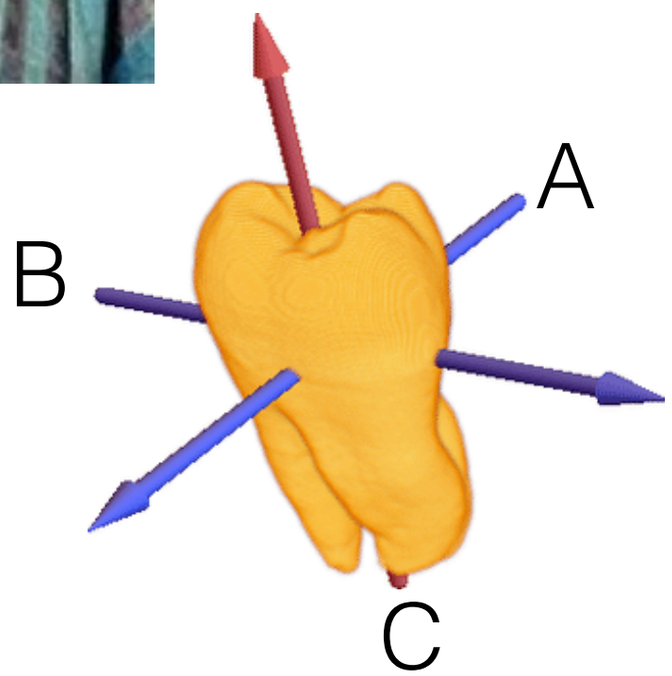
By People

ID	Var1	Var2	Var3
A			
C			

ID	Var1	Var2	Var3
B			
D			

History

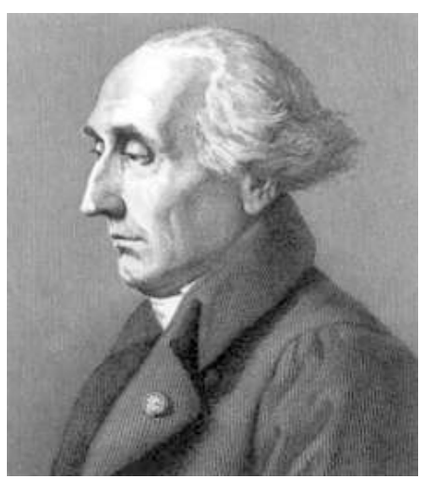
- Part of a set of issues called “Eigen Problems”
- Arose as a subset of phenomena related to differential equations (Your old buddy Euler, c.1750)
- Principal Axes



History

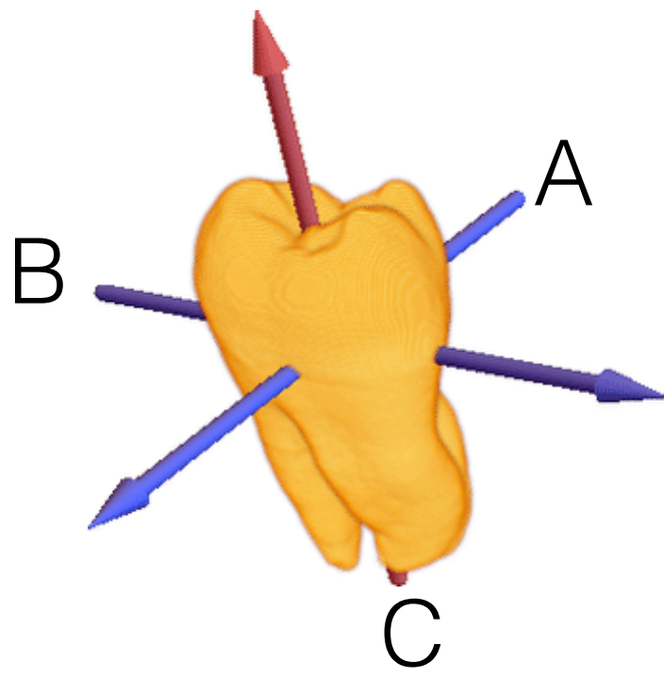
- Describe inertia as a matrix of measurements from the center of an object as it moves

(Joseph-Louis Lagrange)



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

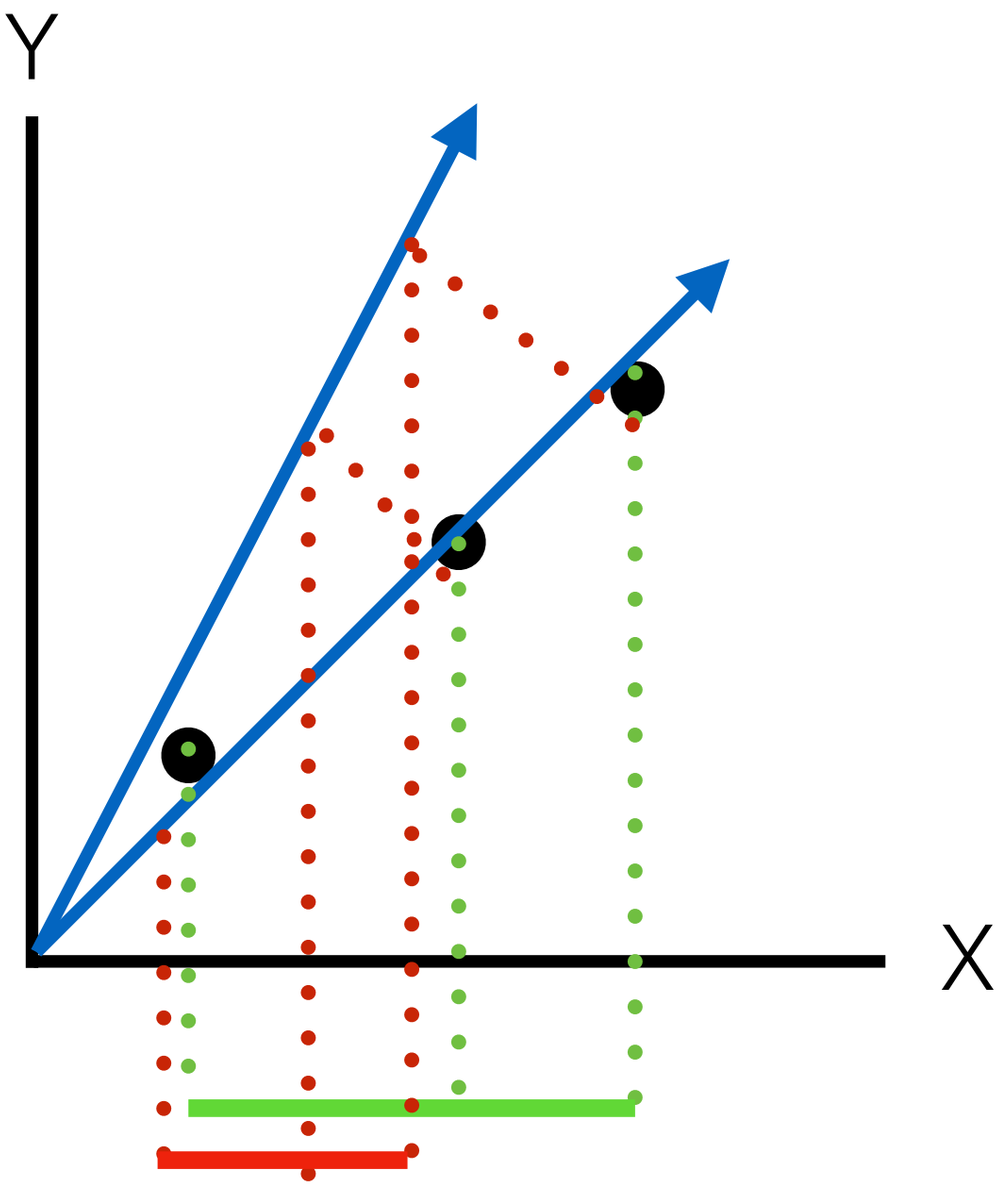
- Principal axes = the lines through which you can describe the object, while maximizing the amount of variation maintained

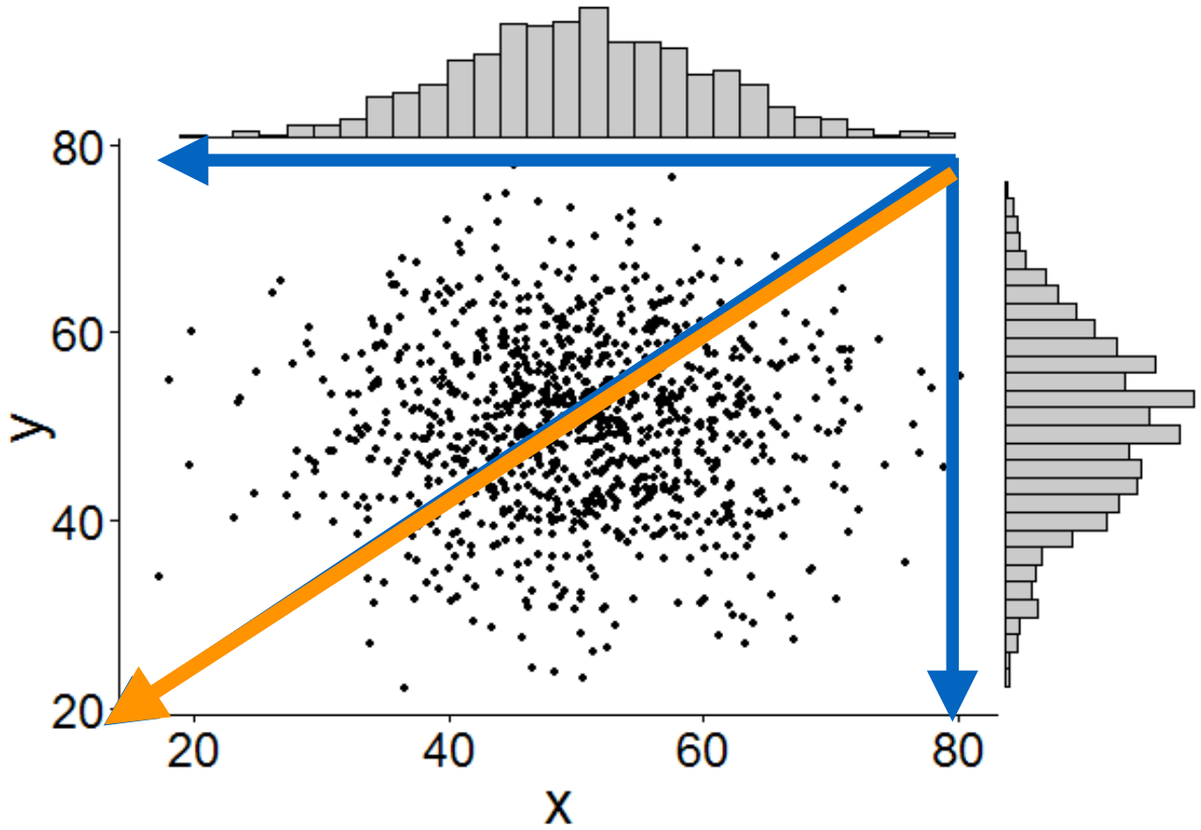


Yada, yada, yada...

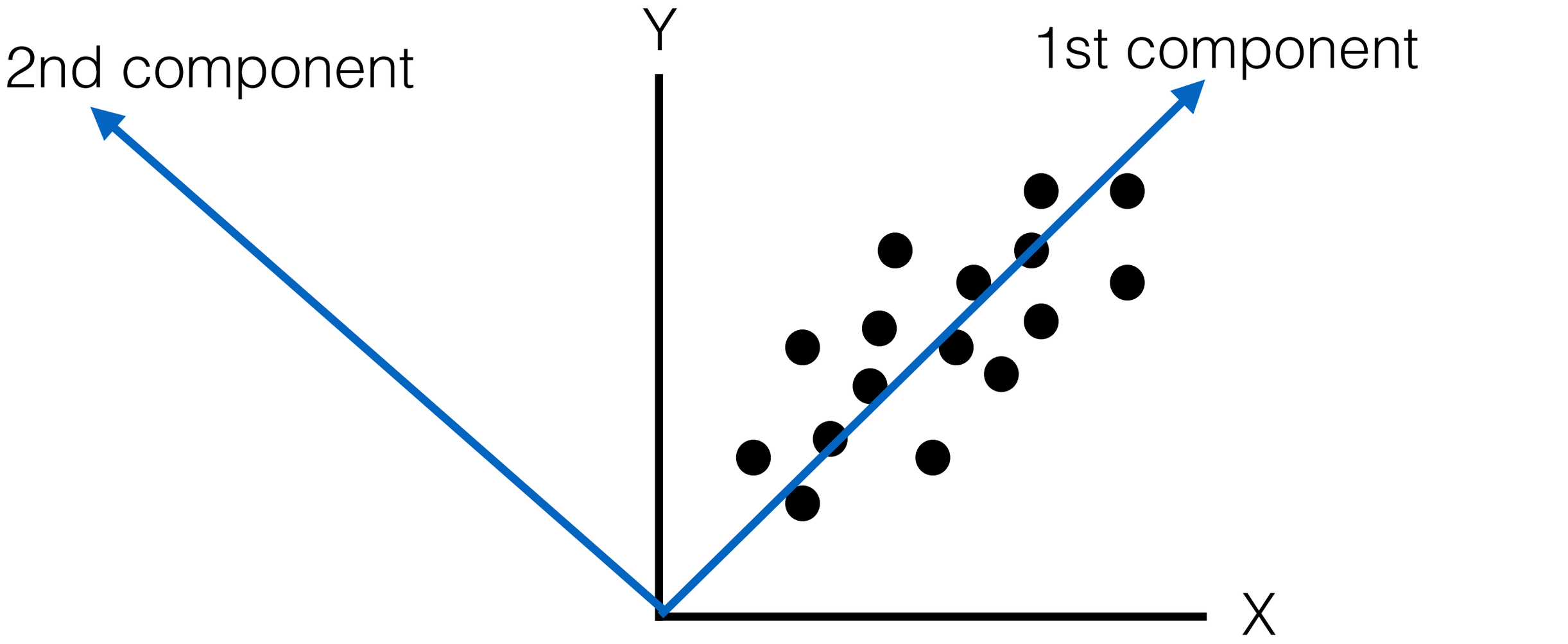
Google

PCA is about Finding the Direction of Maximal Variance

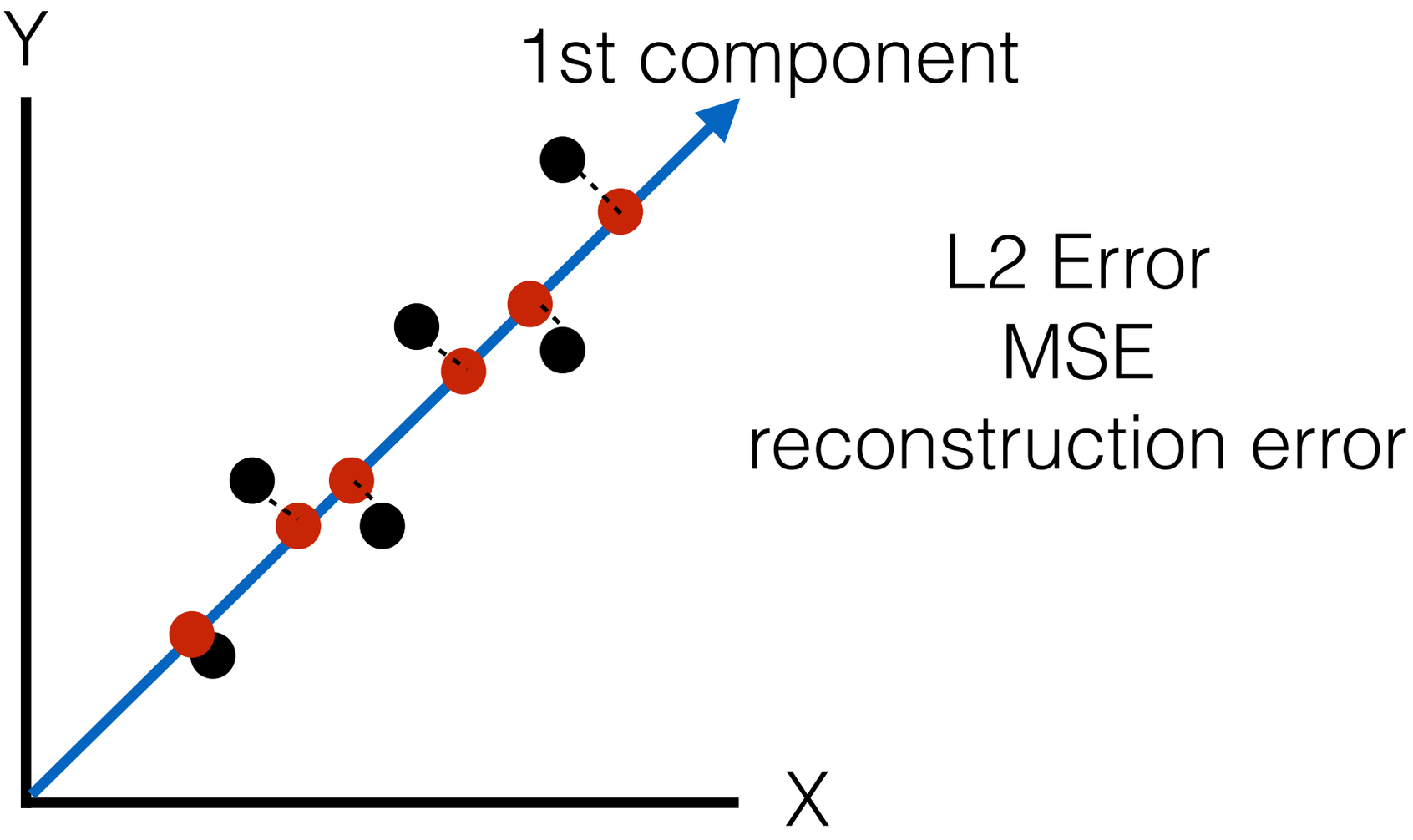




Global Constraint: Orthogonal Components



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



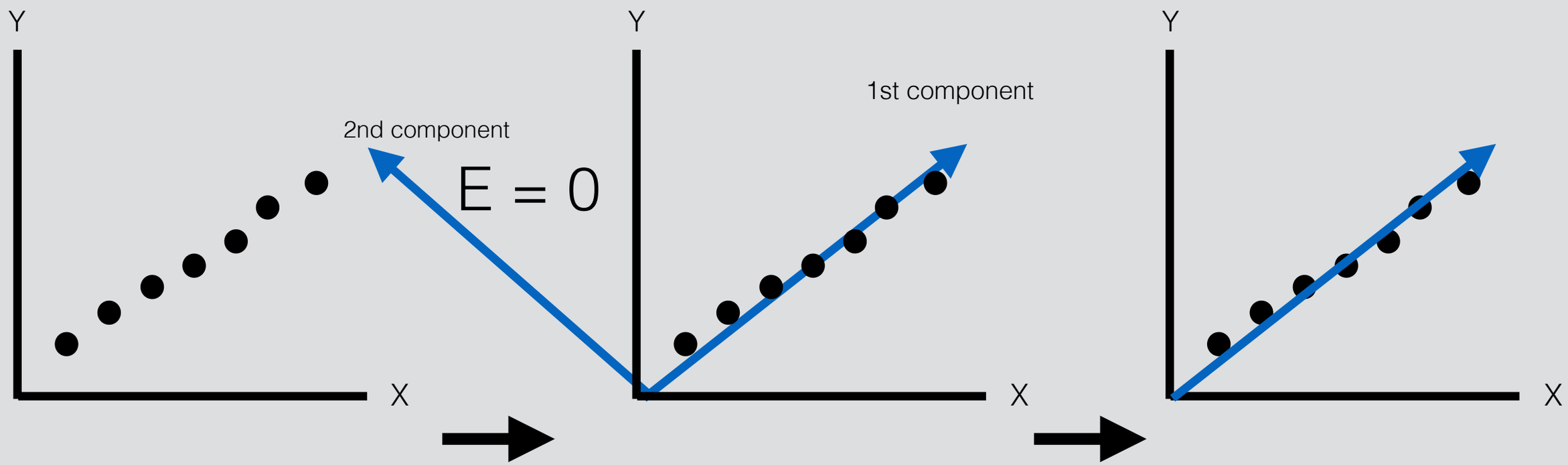
Component is a description of X & Y

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

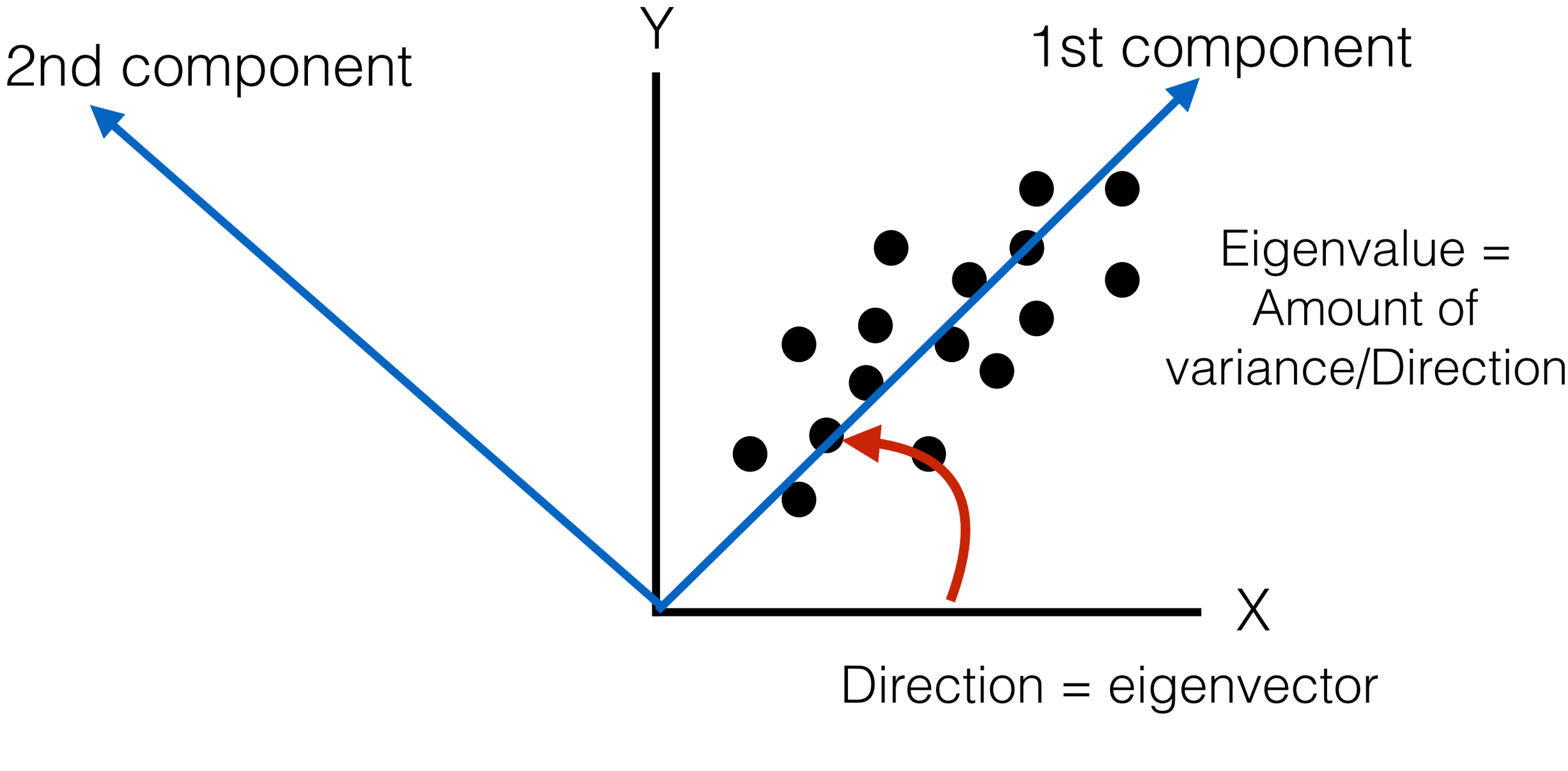
- If a component has an eigenvalue of zero = non-informative (will not effect reconstruction error)
- Therefore, we can delete it = reduce features



Questions?

http://setosa.io/ev/principal-component-analysis/

Orthogonal Components



Eigenvectors

pca$rotation

	PC1	PC2
V1	0.34	-1.6
V2	0.13	-0.07
V3	0.01	0.6
V4	0.02	1.5

Eigenvectors

Creating Composites

Because the eigenvectors represent the shift of each dimension, accounting for max variance, we can use these numbers to weight the construction of a composite.

Composite1(PC1) =
(V1 x E1) + (V2 x E2) + (V3 x E3) + (V4 x E4)

HOWEVER: You must make substantive sense of the component!

Gotchas

- Data needs to be scaled
- Often centered so that the direction goes through zero
- Outliers have an outsized impact on your results
- Continuous variables (or binary but be careful)
- Linear relationships between variables (sometimes impractical)
- Better with larger samples (no real way to test though)
- Components will be uncorrelated!

http://bit.ly/

HUDK405019PCA