PRACTICAL

MACHINE LEARNING

RECOMMENDER ENGINE AND ANOMALY DETECTION

Seth Juarez

sethj@devexpress.com

@sethjuarez

Analytics Program Manager

DevExpress

Titanium Sponsors

AJi

VALOREM CONSULTING
WWW.VALOREMCONSULTING.COM
REGISTER TO WIN AN XBOX!

PAIGE TECHNOLOGIES
INTELLIGENT PAIRING. PERPETUAL SUCCESS.

Platinum Sponsors

ADAPTIVE SOLUTIONS GROUP

DST SYSTEMS

Balance Innovations

Sprint

epiq SYSTEMS

DevExpress

jack henry & ASSOCIATES INC.

O'REILLY

Kauffman LABS for Enterprise Creation

FREIGHTQUOTE

Gold Sponsors

Cerner

CENTRIQ TRAINING

Bradford & Galt CONSULTING SERVICES

ADVANTAGE TECH

KEYHOLE SOFTWARE

ComponentOne a division of GrapeCity

dsi

LRS Consulting Services

GARMIN

JetBRAINS

Microsoft

MULTI SERVICE INNOVATION WHERE IT MATTERS.

KU EDWARDS CAMPUS The University of Kansas

NETCHEMIA Transforming the way education works

OAKWOOD

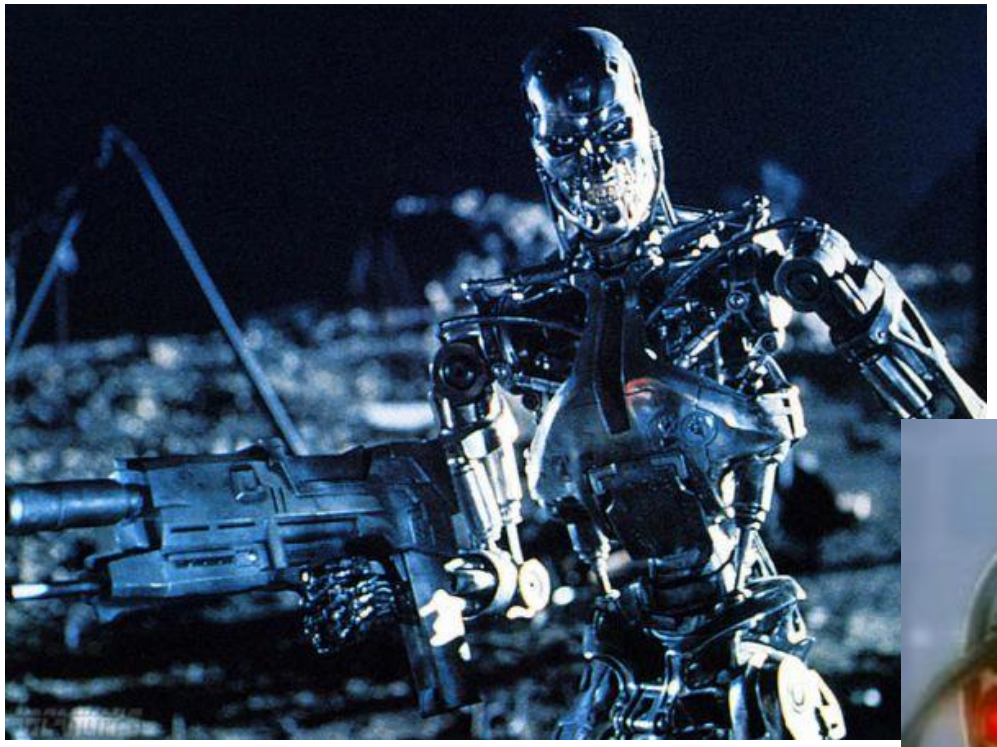
stackify

perceptivesoftware from Lexmark

TEKsystems

twilio CLOUD COMMUNICATIONS

UNITEDLEX



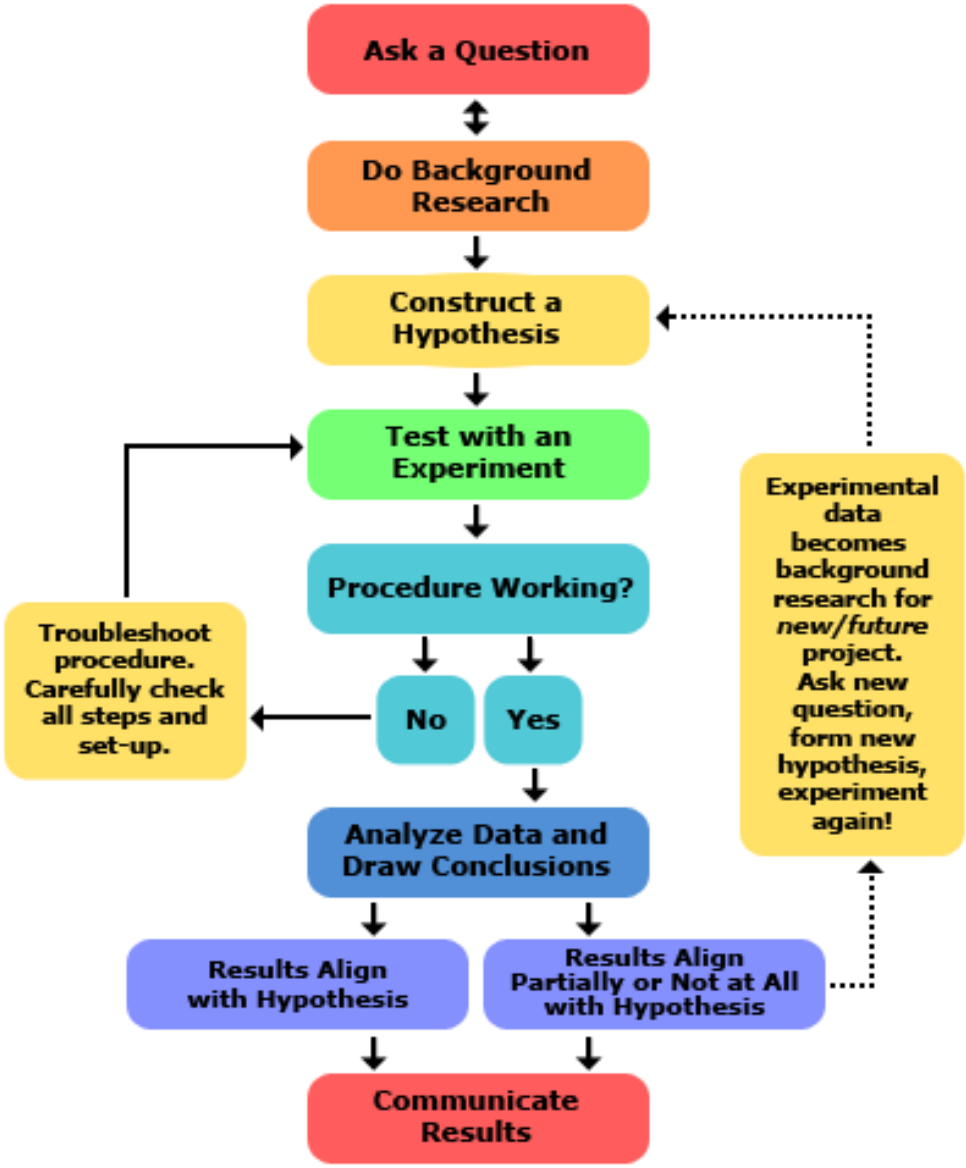
DevExpress™

agenda

- a word about data science
- what is machine learning?
- recommender systems
- unsupervised learning – organization
  - k-means
  - hierarchical clustering
- anomaly detection (motivation)

data science

- key word: science
- try stuff
- it (might not | won't) work the first time



DevExpress™

http://www.sciencebuddies.org/science-fair-projects/project_scientific_method.shtml

machine learning

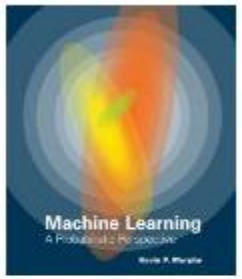
- finding (and exploiting) patterns in data
- replacing “human writing code” with “human supplying data”
  - system figures out what the person wants based on examples
  - need to abstract from “training” examples to “test” examples
  - most central issue in ML: generalization

machine learning

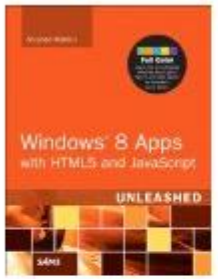
- split into two (ish) areas
  - **supervised learning**
    - predicting the future
    - learn from past examples to predict future
  - **unsupervised learning**
    - understanding the past
    - making sense of data
    - learning structure of data
    - compressing data for consumption

neat applications

Recommendations for You in Books

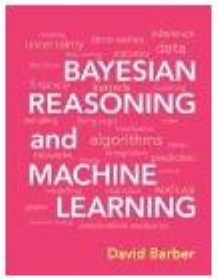


Machine Learning: A Probabilistic...
› Kevin P. Murphy
Hardcover
★★★★½ (16)
~~$90.00~~ $81.00
Why recommended?

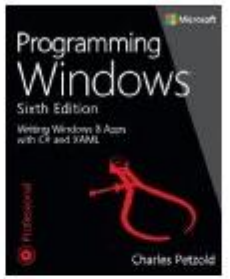


Windows 8 Apps with HTML5 and...
› Stephen Walther
Paperback
★★★★½ (10)
~~$39.99~~ $26.62
Why recommended?

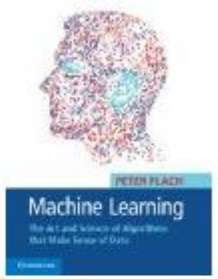
Boosting: Foundations and Algorithms
Robert E. Schapire, Yoav Freund
Hardcover
★★★★☆ (5)
~~$50.00~~ $41.42
Why recommended?



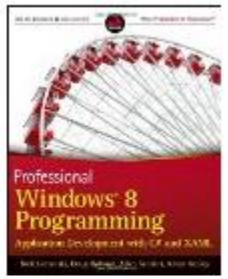
Bayesian Reasoning and Machine Learning
› David Barber
Hardcover
★★★★★ (6)
~~$90.00~~ $81.00
Why recommended?



Programming Windows: Writing Windows...
Charles Petzold
Paperback
★★★★½ (8)
~~$59.99~~ $41.78
Why recommended?



Machine Learning: The Art and Science...
Peter Flach
Paperback
★★★★½ (8)
~~$60.00~~ $54.00
Why recommended?

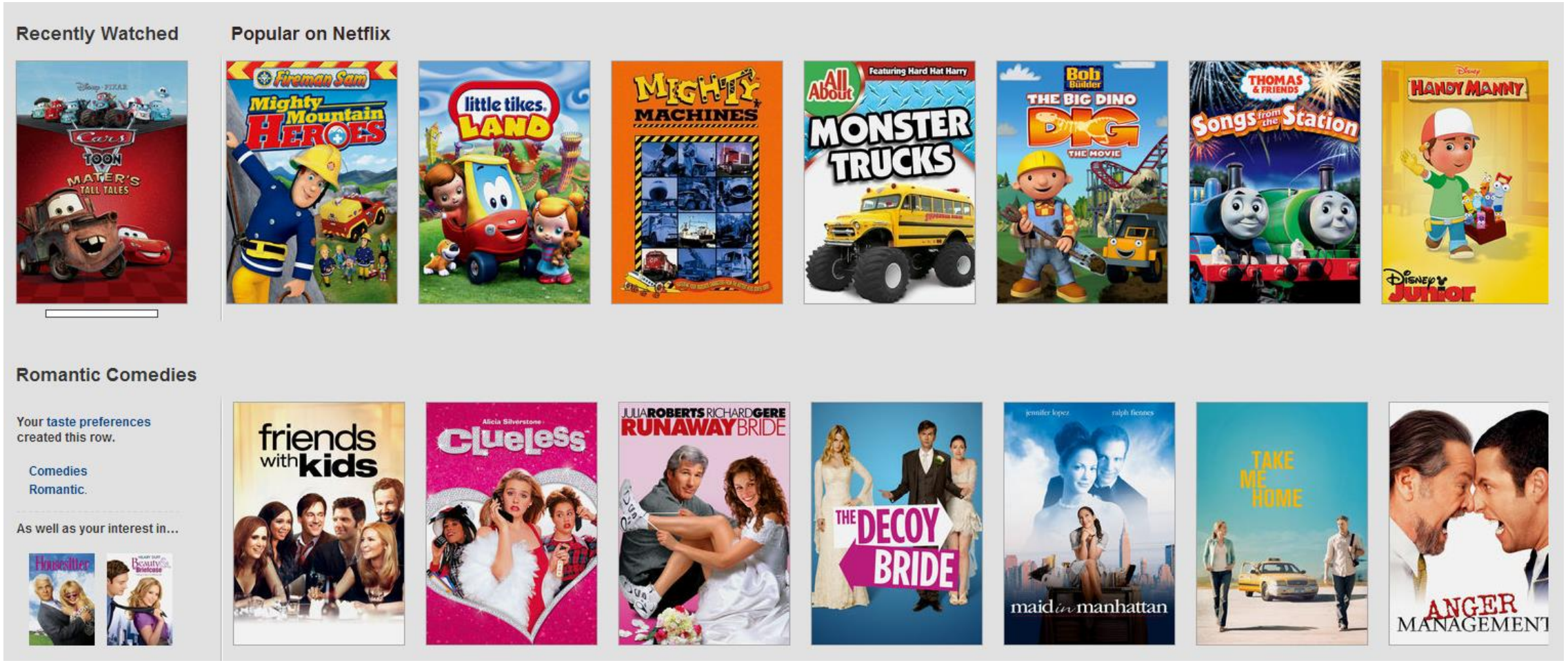


Professional Windows 8 Programming...
Nick Lecrenski, Doug Holland, ...
Paperback
★★★★½ (6)
~~$44.99~~ $27.71
Why recommended?

› See more recommendations

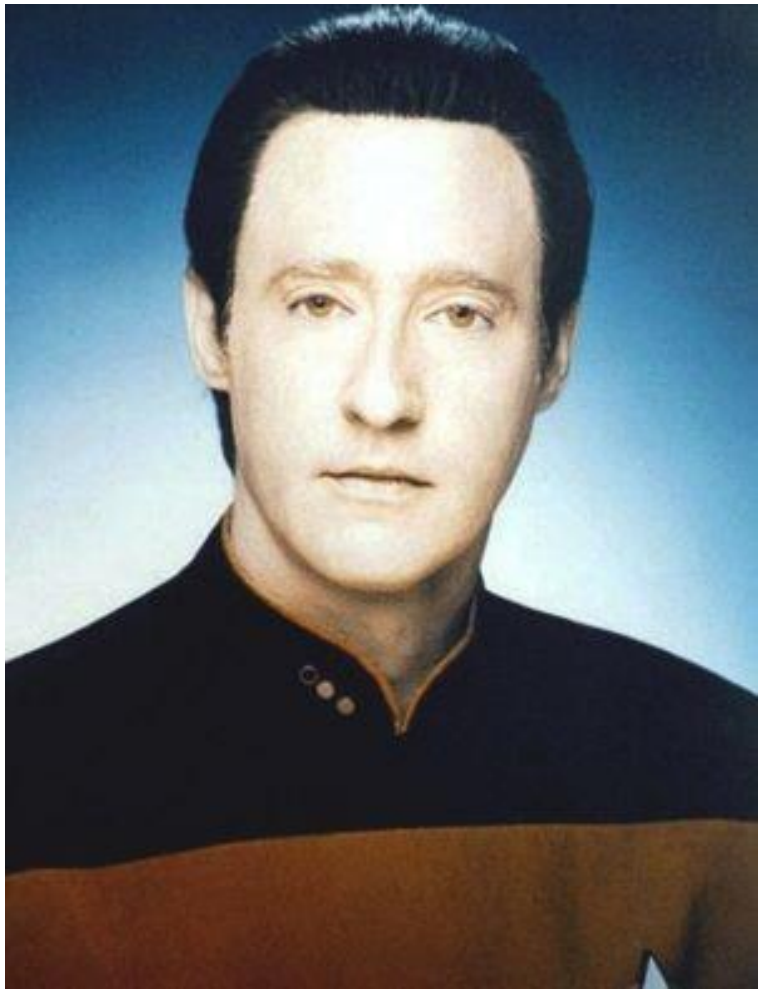
DevExpress™

neat applications



DevExpress™

neat applications



DevExpress™

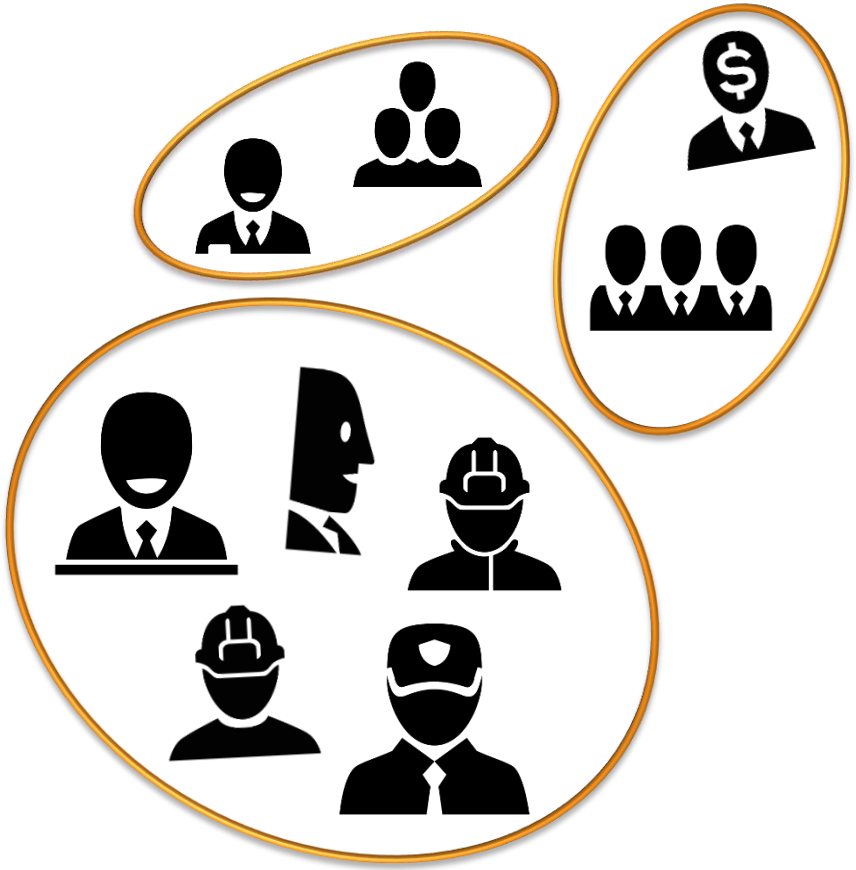
neat applications

- spam catchers
- ocr (optical character recognition)
- natural language processing
- machine translation
- biology
- medicine
- robotics (Autonomous Systems)
- etc…

RECOMMENDER SYSTEM

recommender systems

- what do people like?



how does knowing people similarity help when recommending something?

recommender systems

- how are things alike?

DevExpress™

how does knowing item similarity
help when recommending something?

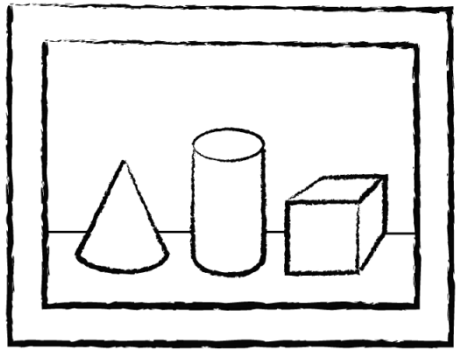
UNSUPERVISED LEARNING

figuring out similarity (among other things)

HOW DOES IT WORK?

pattern

1. data

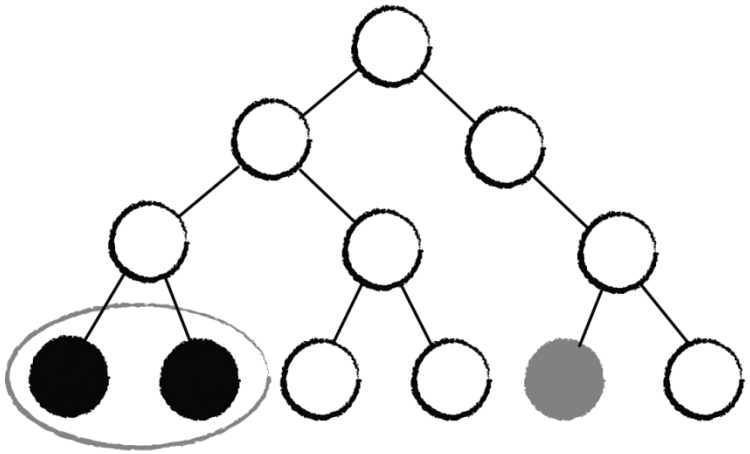


2. maths

$$\begin{bmatrix} 4 & 3 & 1 & 2 & 6 \\ 4 & 3 & 2 & 3 & 8 \\ 5 & 3 & 1 & 2 & 7 \\ 2 & 1 & 5 & 8 & 0 \\ 1 & 0 & 8 & 6 & 4 \end{bmatrix}$$

$$8^{6}5^{6}7^{8}7^{1}$$

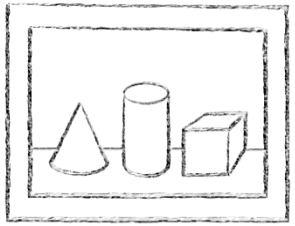
$$\begin{bmatrix} X + Y = Z \\ A + B = C \\ X + A = Z \\ B + Y = ? \end{bmatrix}$$



3. model

DevExpress™

data – example



Grade	GPA	Age	Tall	Friends
A	3.5	16	Yes	12
C	2.0	12	No	3
F	2.1	13	Yes	1
B	3.5	17	Yes	6
D	2.0	18	No	4
A	3.8	15	No	6
D	2.3	14	No	4
B	3.3	17	Yes	8

features

gpa, age, tall, friends

values (x)

[A, 3.5, 16, Yes, 12]

- which students are most similar? how should they be grouped? given a new student, where does she belong?

DO SOMETHING

code

DISTANCE AND SIMILARITY

math

distance (metric)

- $d : X \times X \to \mathbb{R}$
- Must follow these rules:
  1. $d(x, y) \geq 0$
  2. $d(x, y) = 0 \iff x = y$
  3. $d(x, y) = d(y, x)$
  4. $d(x, z) \leq d(x, y) + d(y, z)$
- Main idea: if I have a Φ and a Ψ how far away are they from each other?
- **closer** – **similar**, **farther** – **dissimilar**

distance

- euclidian distance
- manhattan distance
- cosine distance
- hamming distance

k-means

- how it works
  - initialize K centers
  - find closest (distance) points to K centers
  - set each center to the average of the closest points
  - rinse and repeat until convergence

DEMONSTRATION

k-means

strings?!?!

1. the red dog →
2. cat eats dog →
3. dog eats food →
4. red cat eats →

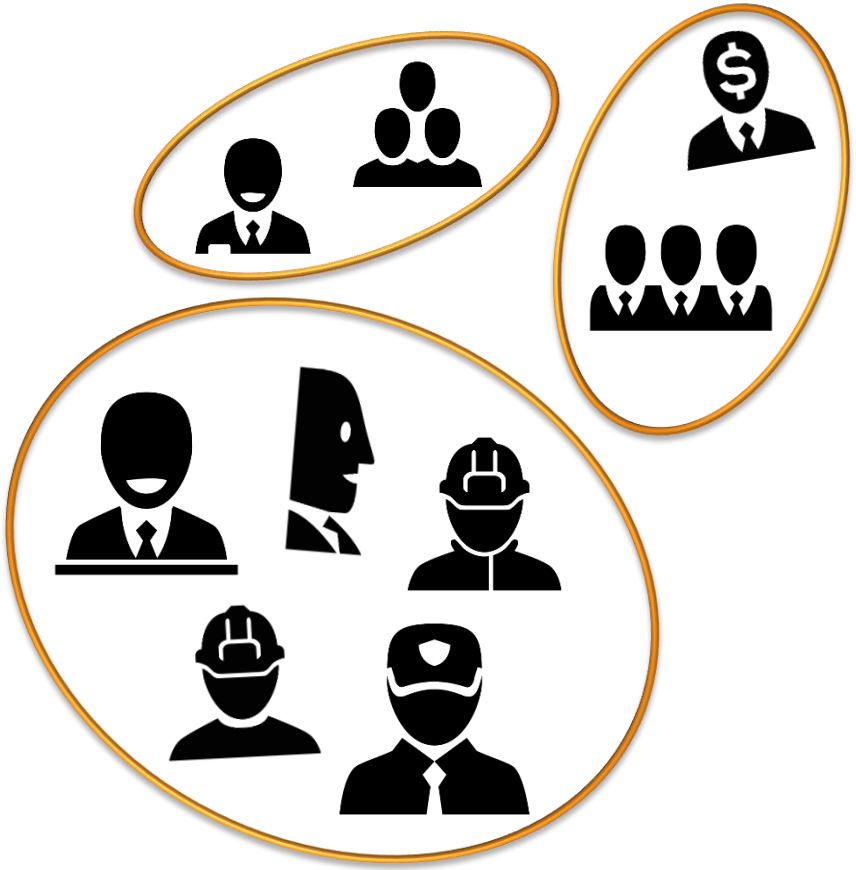
the	red	dog	cat	eats	food
1	1	1	0	0	0
0	0	1	1	1	0
0	0	1	0	1	1
0	1	0	1	1	0

RECOMMENDING

collaborative filtering / content-based filtering

collaborative filtering

- what people who are similar to me like?



DevExpress™

content-based filtering

- what items are similar to the one I chose?

DevExpress™

DEMONSTRATION

collaborative filtering / content-based filtering

DevExpress™

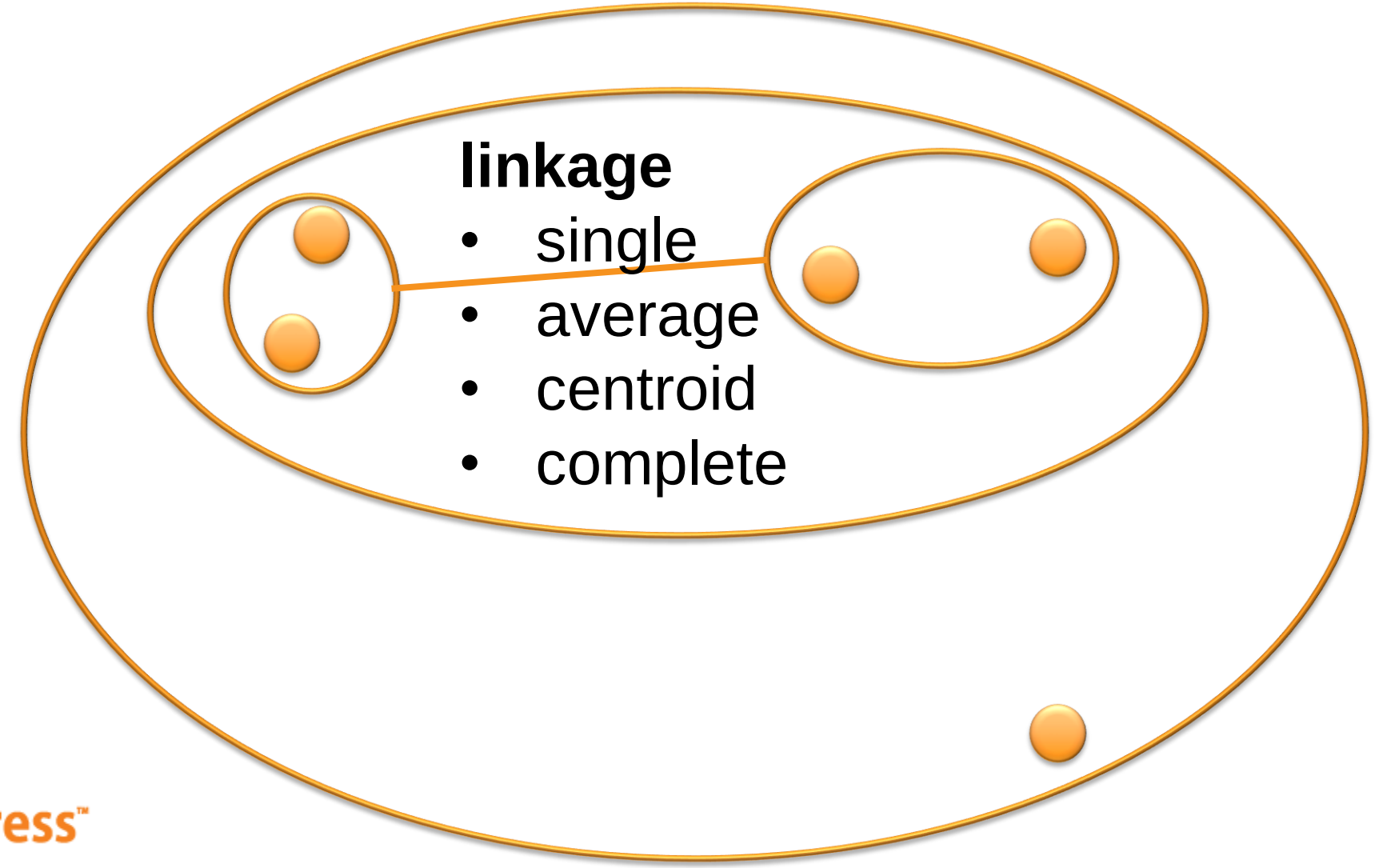
SIMILARITY

other ways of measuring similarity

HIERARCHICAL CLUSTERING

models

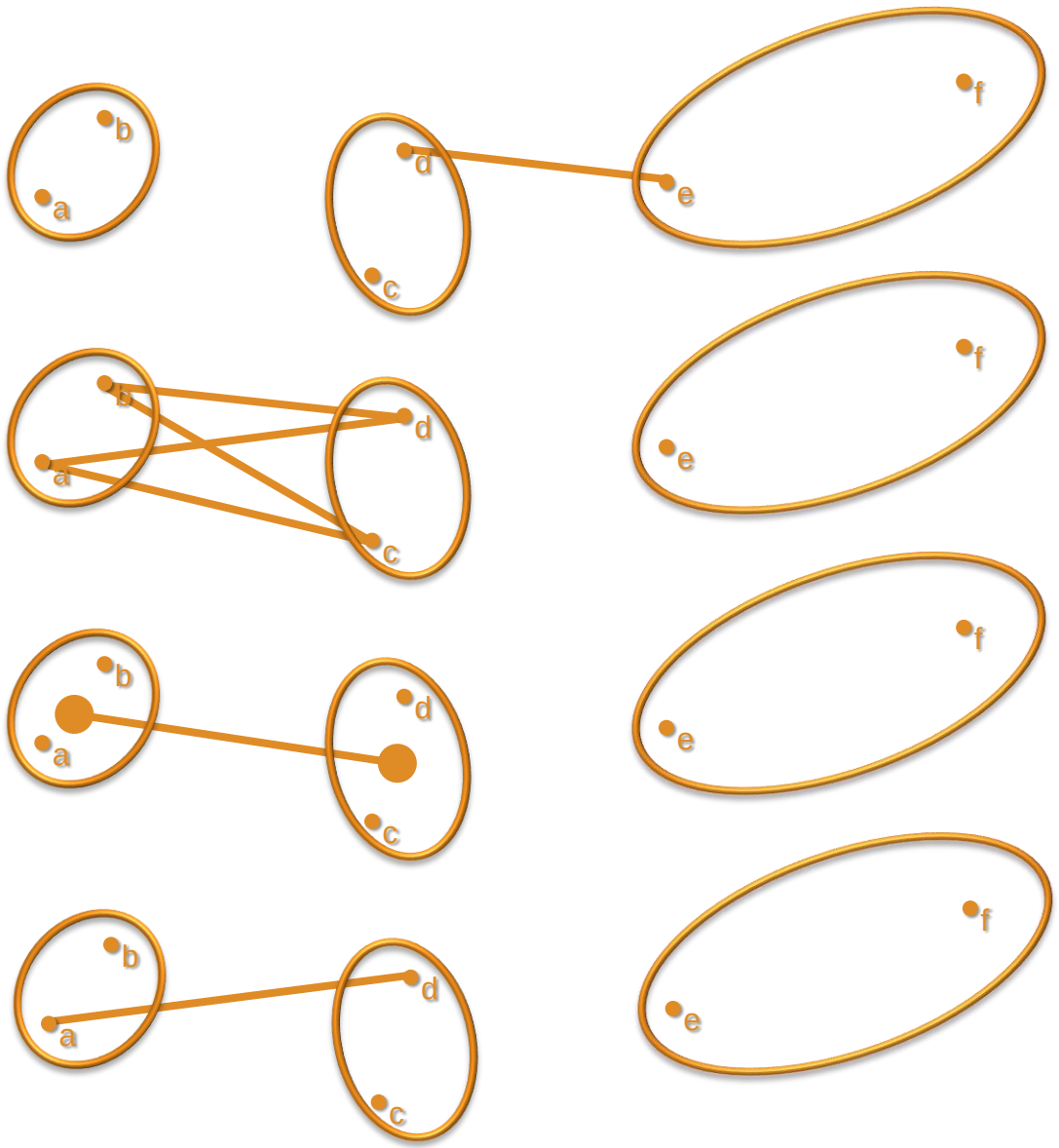
clustering



DevExpress™

linkage

- single
- average
- centroid
- complete



DevExpress™

DEMONSTRATION

hierarchical clustering

DevExpress™

ANOMALY DETECTION

brainstorm / motivation

anomaly detection

- how could we use KMeans to detect anomalies?
  - step 1: get last **n** things and convert them to vectors (matrix)
  - step 2: get thing in question and add it to the bunch
  - step 3: run KMeans
  - step 4: measure distance of [new thing] from all centers
  - step 5: return min/avg/mode – whatever
  - step 6: tune (what is an acceptable threshold?)

anomaly detection

- how could we use Hierarchical Clustering to detect anomalies?

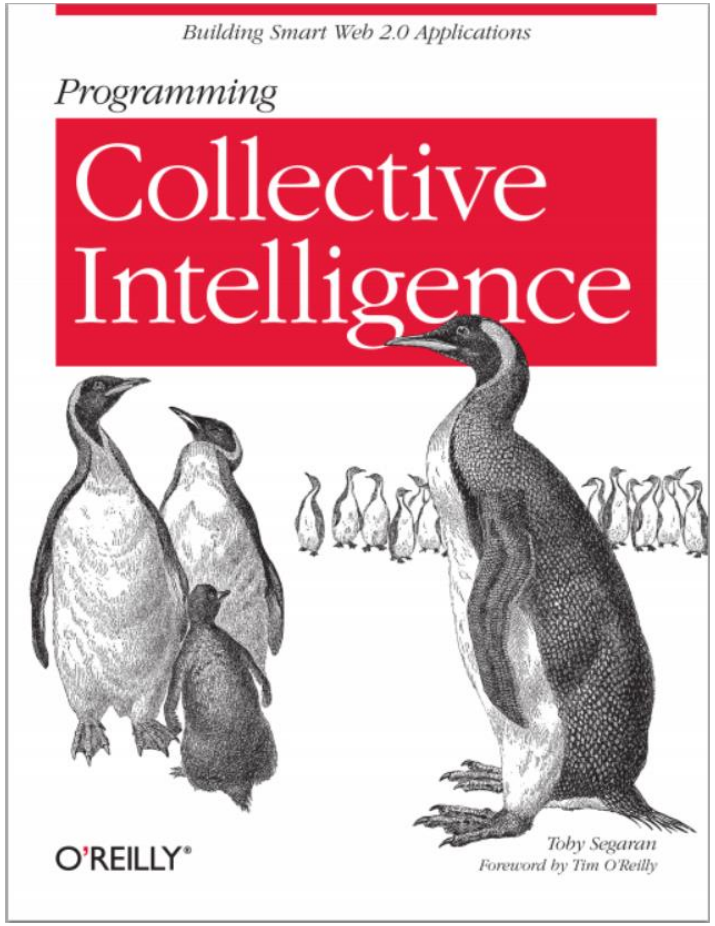
recap

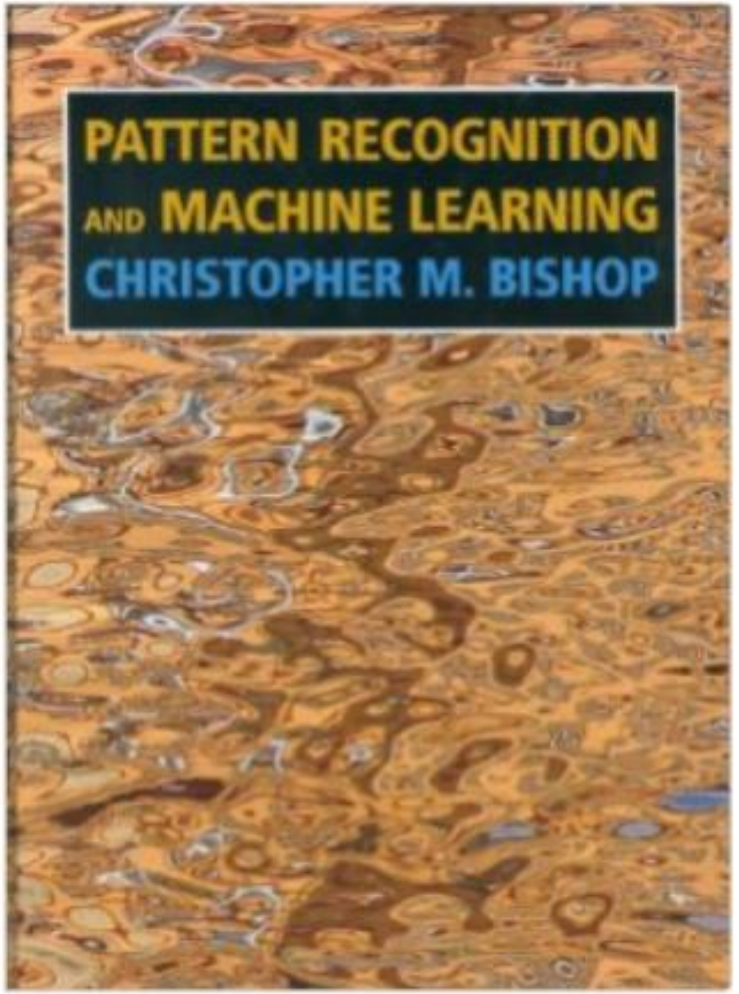
- a word about data science
- what is machine learning?
  - supervised
  - unsupervised
- recommender systems, anomaly detection
- k-means, hierarchical clustering
- nuML – http://numl.net

planned

- unsupervised learning:
  - gaussian mixture models
  - latent semantic analysis (better pca for text)
  - self organizing maps
  - smarter interfaces (API to deal with reduction chaining)
  - multi-processor, gpgpu

some reading





DevExpress™

QUESTIONS?

Seth Juarez

sethj@devexpress.com
@sethjuarez
Analytics Program Manager
DevExpress