

Week 8 Day 2

- Introduction to machine learning
- Supervised versus unsupervised learning
- How this all fits into EDA and Data Science
- Ethical mini case studies in machine learning and data science

THIS IS YOUR MACHINE LEARNING SYSTEM?

YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.

WHAT IF THE ANSWERS ARE WRONG?

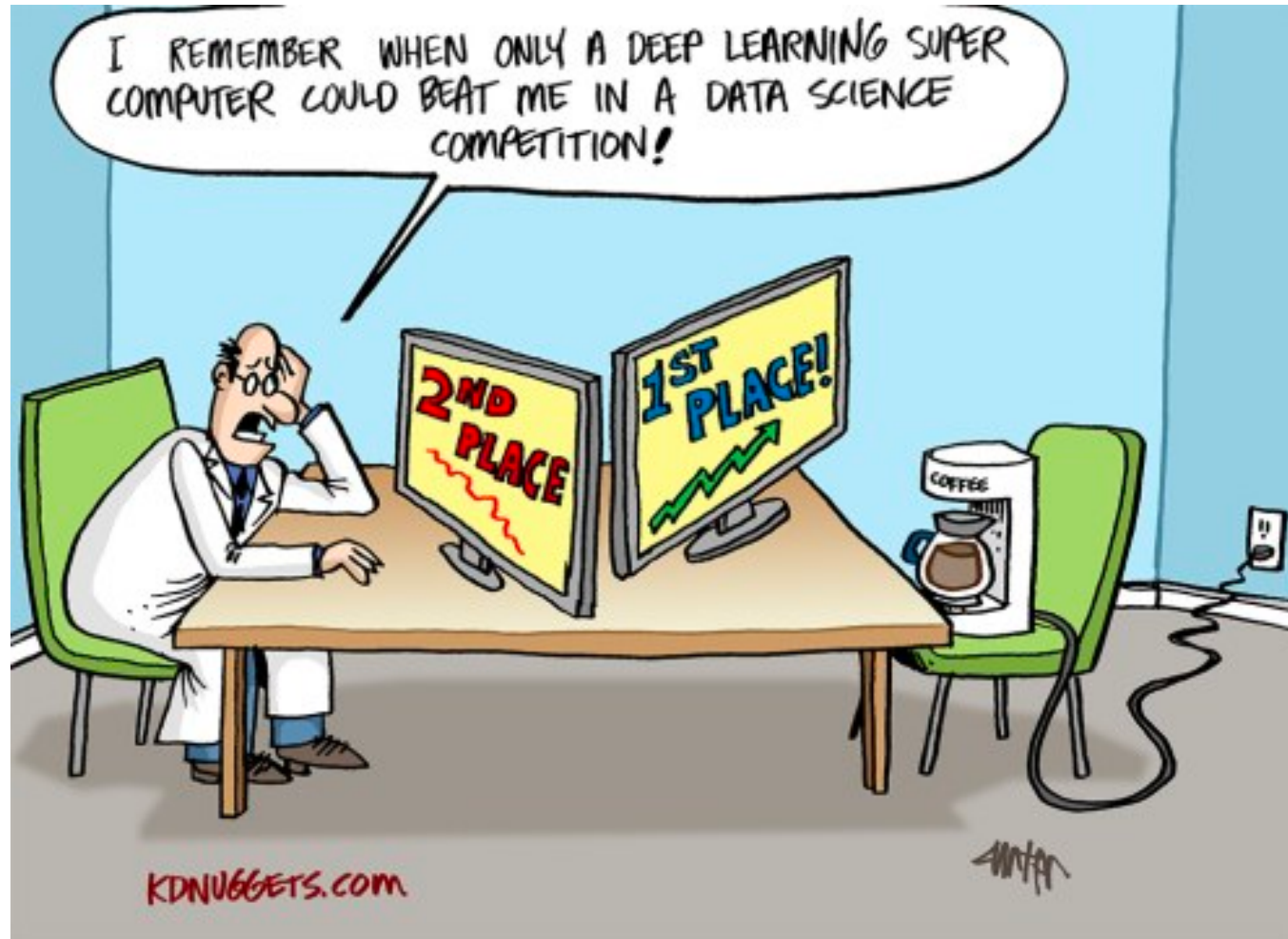
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.





SORRY, KID, OUR MACHINE LEARNING CRM WITH PREDICTIVE ANALYTICS SAYS YOU'RE GETTING COAL THIS YEAR.

TOM FISH BURNE



KDNUGGETS.COM

MILLER

© 2014 Ted Goff
KDnuggets Cartoon



“The machine learning algorithm wants to know if we’d like a dozen wireless mice to feed the Python book we just bought.”

TOPHAT Questions

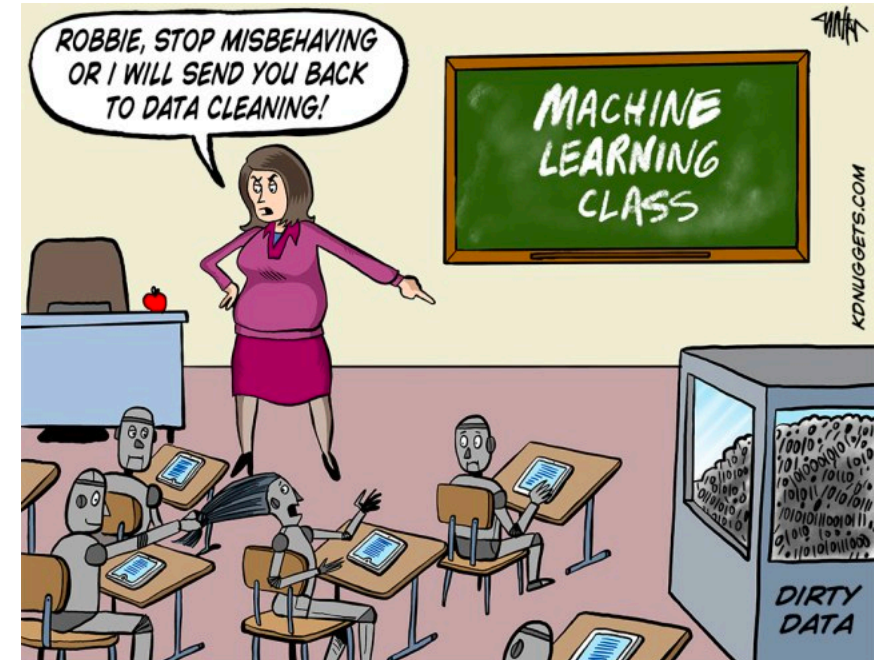


Machine Learning is a huge topic

Basic definition (that you hear a lot):

[Machine Learning is the] field of study that gives computers the ability to learn without being explicitly programmed.

- Arthur Samuel, 1959



Machine learning

Machine learning algorithms use computational methods to learn information from data in order to make predictions, decisions, or complete tasks.

Spotify

Collaborative
Filtering models

Natural Language
Processing (NLP)

Audio models



dave horwitz
@Dave_Horwitz



It's scary how well @Spotify Discover Weekly playlists know me. Like former-lover-who-lived-through-a-near-death experience-with-me well.

♡ 281 2:09 PM - Oct 27, 2015



💬 193 people are talking about this



Amanda Whitbred
@amandawhitbred



At this point @Spotify's discover weekly knows me so well that if it proposed I'd say yes

♡ 80 4:36 PM - Aug 18, 2016



💬 28 people are talking about this



Hello Barbie

- Listens and responds to a child
- Microphone records what is said and transmits it to the server at Toy Talk
- There, the recording is analyzed to determine the appropriate response from 8,000 lines of dialogue
- Servers transmit the correct response back to Barbie in under a second so she can respond to the child.



Hello Barbie

Microphone, speaker and tri-color LED lights embedded in necklace.

Turn the doll on with the power button on her belt.

Press and hold down belt buckle to activate speech recognition.
Note: Speech Recognition is Not "On" Unless Pushed.

Doll cannot stand alone.

Flat feet for charging

ONE TIME APP DOWNLOAD AND WIFI CONNECTION REQUIRED FOR 2-WAY CONVERSATION
Disclaimer: Compatible smart device required.

PARENT CONSENT REQUIRED

CHARGING STAND INCLUDED
Note: Playtime on the battery life is about an hour.

Get Started

HELLO HELLO HELLO

HELLO HELLO HELLO

HELLO HELLO HELLO

HELLO HELLO HELLO

HELLO HELLO HELLO

HELLO HELLO HELLO

Deep Face

- Researchers use photos “from the wild”
- Once it has an original image of a face it turns this into a 3-D model
- Once this is done it can then use its neural network to look for high-level similarities between different photos of the same person.

03.29.14

How Facebook's Machines Got So Good At Recognizing Your Face

In DeepFace, Facebook has some of the best facial recognition technology in the world. Here's why it's a big deal.



ML is essential to many fields

- Natural Language Processing
- Computational Biology
- Advertising and Marketing
- Astronomy
- Social Science
- Robotics
- *We will not cover it all, this is just a soft introduction as it is relevant to exploratory data analysis!*

Learning Algorithms

Supervised

labeled data or
desired output
(or a clear DV)

algorithm/model

output

Unsupervised

No clear
labeled system
or DV

algorithm/model

output

Reinforcement

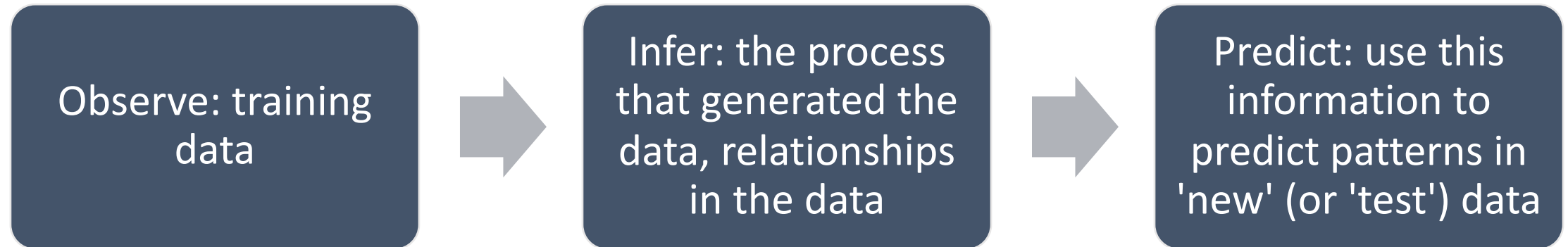
Input state

Decision making
function = action

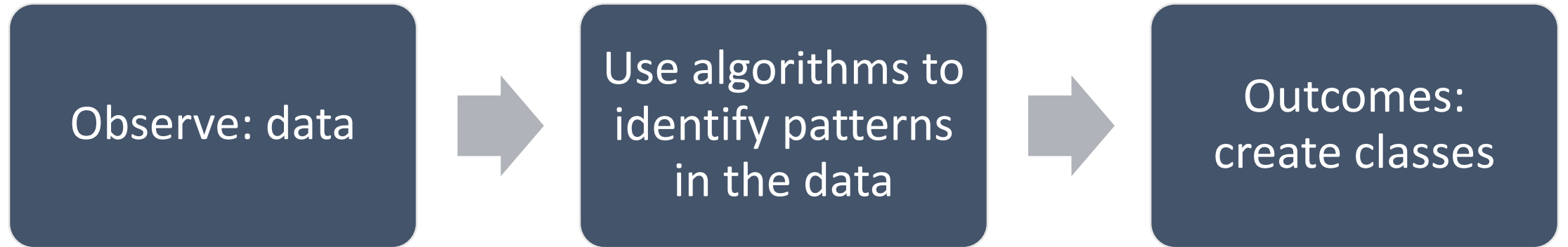
agent receives reward
or reinforcement

state-action pair
information about the
reward is stored

Supervised learning in practice



Unsupervised learning in practice



Supervised v. Unsupervised

#1. Method

Supervised Learning



Input variables and output variables will be given.

Unsupervised Learning



Only input data will be given.

#2. Goal

Supervised Learning



Supervised learning goal is to determine the function so well that when new input data set given, can predict the output.

Unsupervised Learning



Unsupervised learning goal is to model the hidden patterns or underlying structure in the given input data in order to learn about the data.

Modeling Approaches

Supervised

- Classification
- Regression
- Support vector machine
- Nearest neighbor
- Decision Trees
- Neural Networks

Unsupervised:

- Clustering
- Association rules
- Component analysis

How does this fit in to EDA?

Overview: the Data Science Pipeline
+
all the terms you keep hearing

DATA



EXPLORE



MODEL



Conclusions (with caveats)

DATA

DGP collection sample design ethics scraping sourcing
 domain expertise industry bias selection

EXPLORE

clean wrangle process visualize patterns
 reduce features reduce complexity

MODEL

$y = f(x)$ prediction v. inference selected features
 test v. training data

Conclusions (with caveats)

Where would you write 'unsupervised'
and 'supervised' learning?

DATA

DGP collection sample design ethics scraping sourcing
 domain expertise industry bias selection

EXPLORE

clean wrangle process visualize patterns
 reduce features reduce complexity

MODEL

$y = f(x)$ prediction v. inference selected features
 test v. training data

Conclusions (with caveats)

DATA

DGP collection sample design ethics scraping sourcing
 domain expertise industry bias selection

EXPLORE

 clean wrangle process visualize patterns
reduce features reduce complexity **unsupervised learning**

MODEL

$y = f(x)$ prediction v. inference selected features
 test v. training data **supervised learning**

Conclusions (with caveats)

EDA: unsupervised learning

Clustering

Refers to broad set of techniques for finding subgroups, or clusters, in a dataset.

Principal Component Analysis

When faced with a large set of correlated variables, principal components allow us to summarize this set with a smaller number of representative variables

ML types + ethics Activity Handout