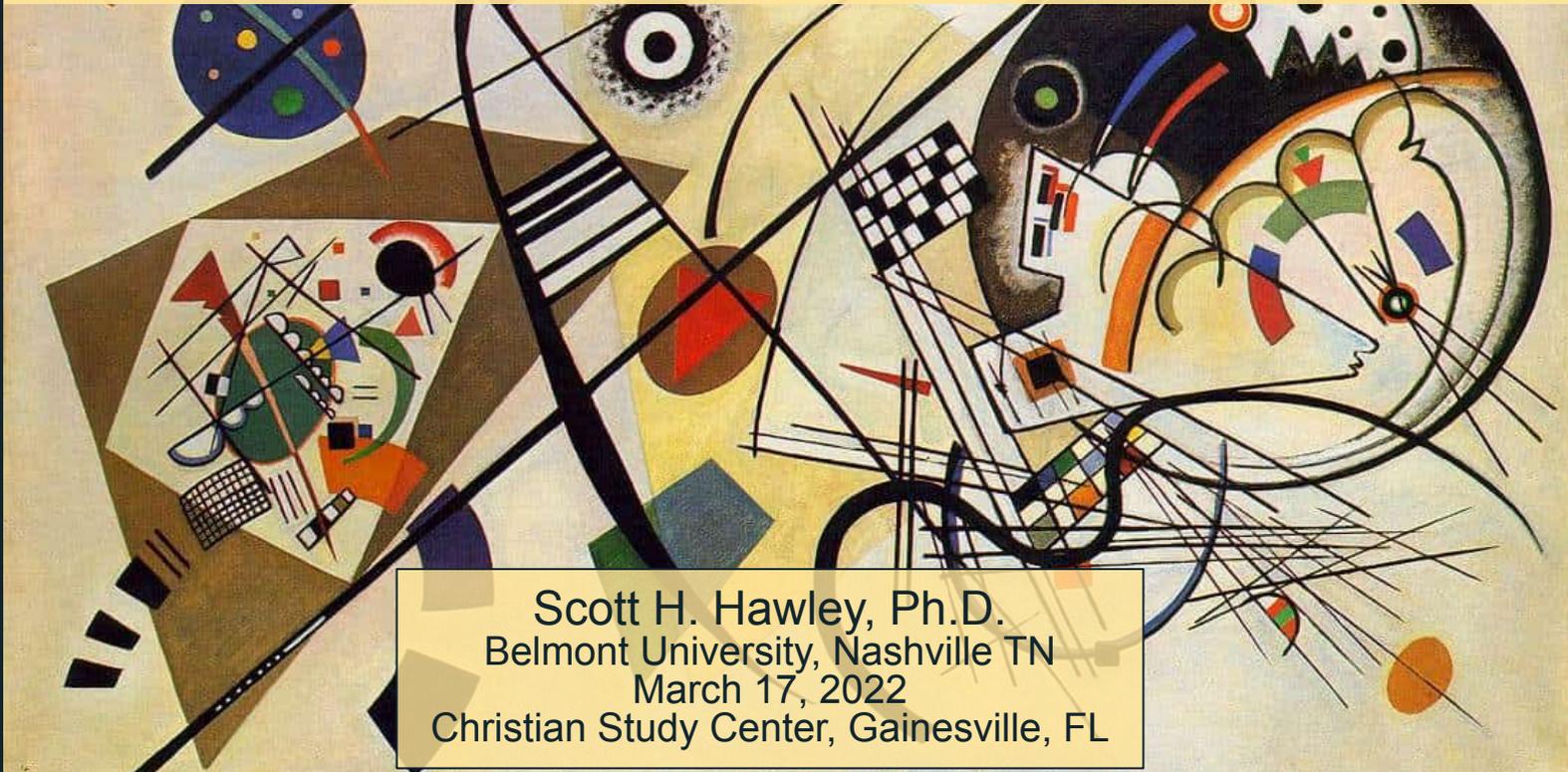


Curves & Categories:

Machine Learning, AI, and the Nature of Classification



Scott H. Hawley, Ph.D.
Belmont University, Nashville TN
March 17, 2022
Christian Study Center, Gainesville, FL

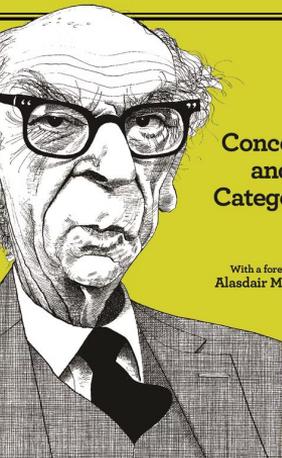


Curves & Categories:

Machine Learning, AI, and the Nature of Classification

Title is my
homage to...

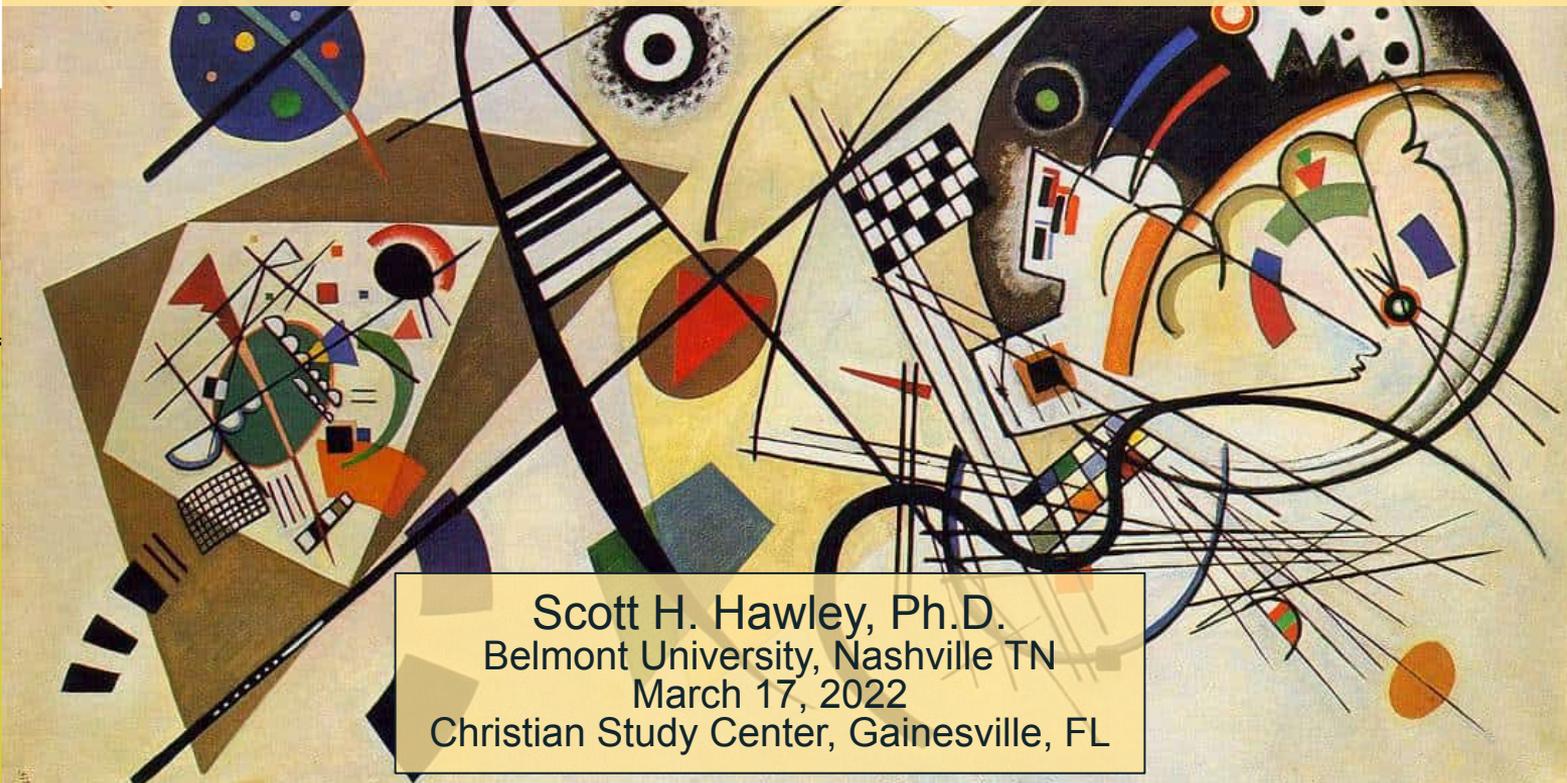
Isaiah Berlin



Concepts
and
Categories

With a foreword by
Alasdair MacIntyre

Edited by Henry Hardy



Scott H. Hawley, Ph.D.
Belmont University, Nashville TN
March 17, 2022
Christian Study Center, Gainesville, FL

Wassily Kandinsky, *Transverse Line*, 1923

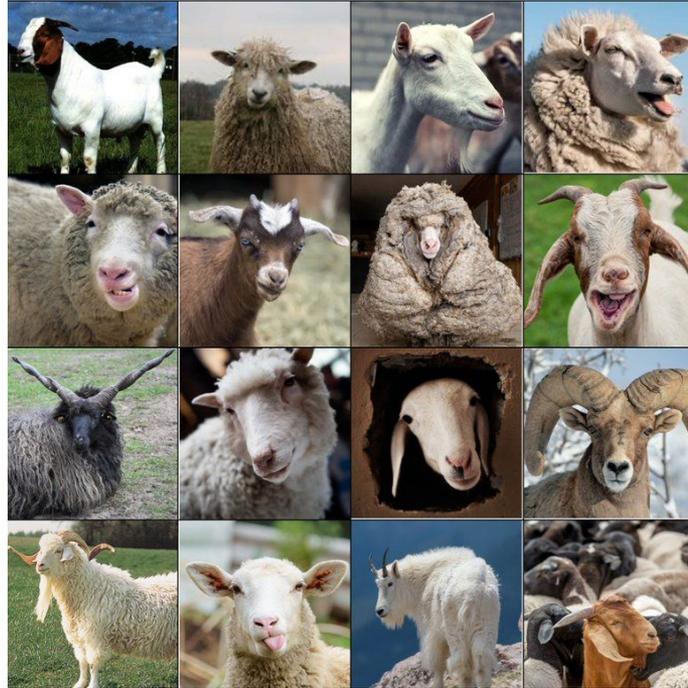
A Few Introductory Exercises

Exercise 1:

Select all squares with
sheep

If there are none, click skip.

Matthew 25:31-46



31 “When the Son of Man comes in his glory, and all the angels with him, he will sit on his glorious throne.
32 All the nations will be gathered before him, and he will separate the people one from another as a shepherd separates the sheep from the goats.
33 He will put the sheep on his right and the goats on his left.



Report a problem

SKIP

Select all squares with
sheep
If there are none, click skip.

Matthew 25:31-46



31 “When the Son of Man comes in his glory, and all the angels with him, he will sit on his glorious throne.
32 All the nations will be gathered before him, and he will separate the people one from another as a shepherd separates the sheep from the goats.
33 He will put the sheep on his right and the goats on his left.



Report a problem

SKIP

Select all squares with
sheep
If there are none, click skip.

Matthew 25:31-46



Goat vs Sheep: 6 Key Differences Explained

<https://a-z-animals.com/blog/goat-vs-sheep-6-key-differences-explained/>

	Sheep	Goat
Coat type	Wool	Hair
Tail	Points down	Points up
Upper lip	Has a distinct groove in the center (philtrum)	No groove
Foraging behavior	Grazers	Browsers
Temperament	Distant and aloof and have a strong flocking instinct	Curious and independent
Horns	Most don't have horns, but those that do usually curl round the sides of their head	Most usually have horns and they point upwards and slightly back



Report a problem

SKIP

Exercise 2: "Prove that you *are* a computer"

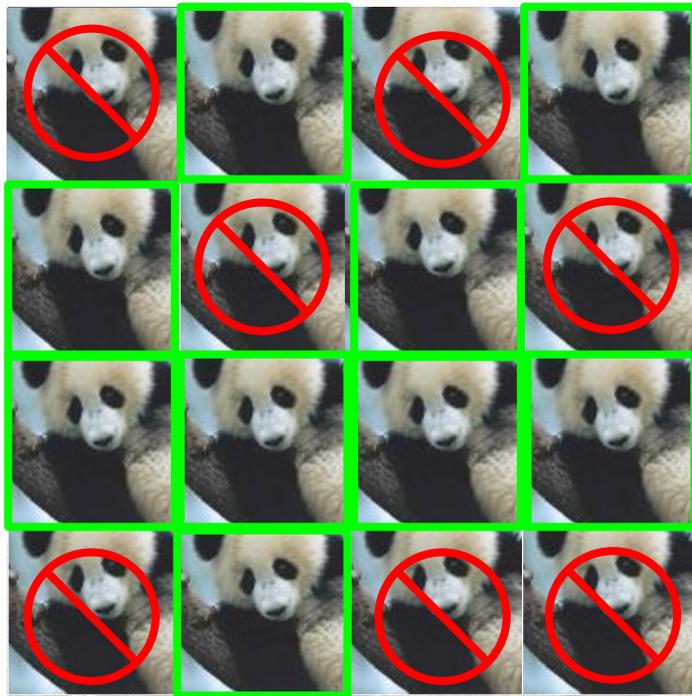
Select all squares with
pandas
If there are none, click skip.



Report a problem

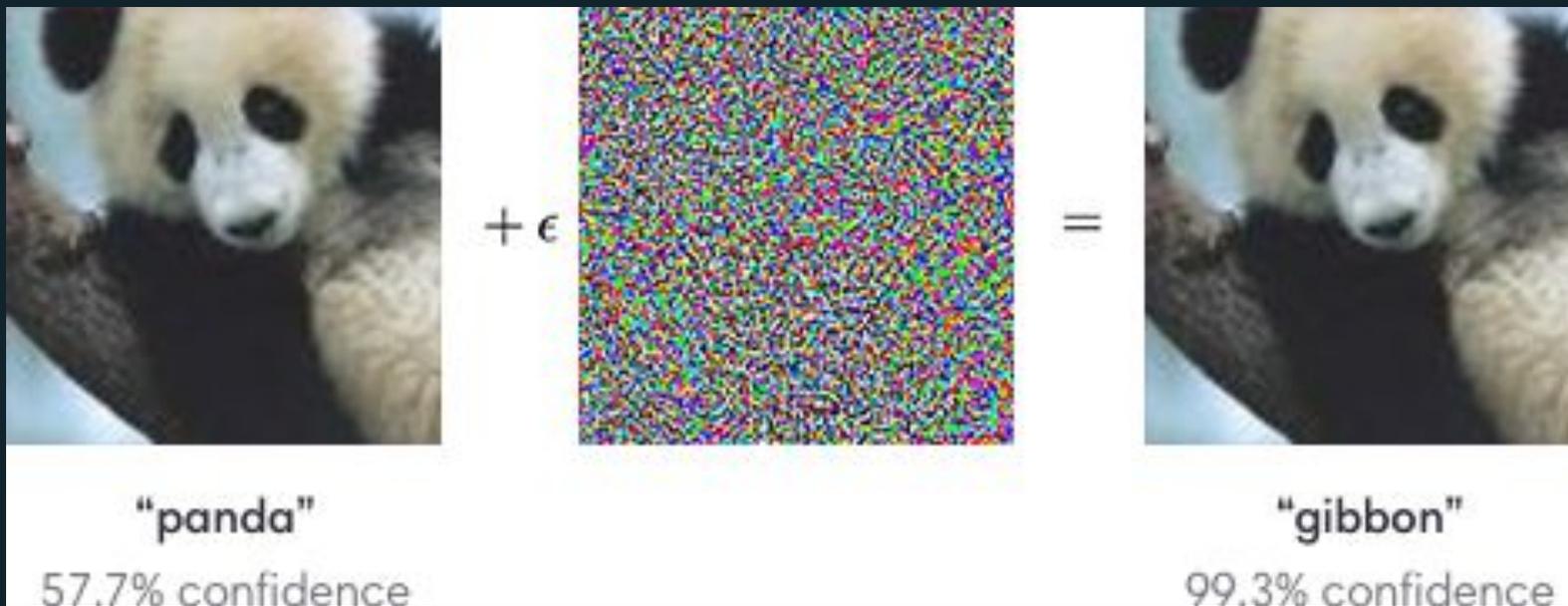
SKIP

Select all squares with
pandas
If there are none, click skip.



Report a problem

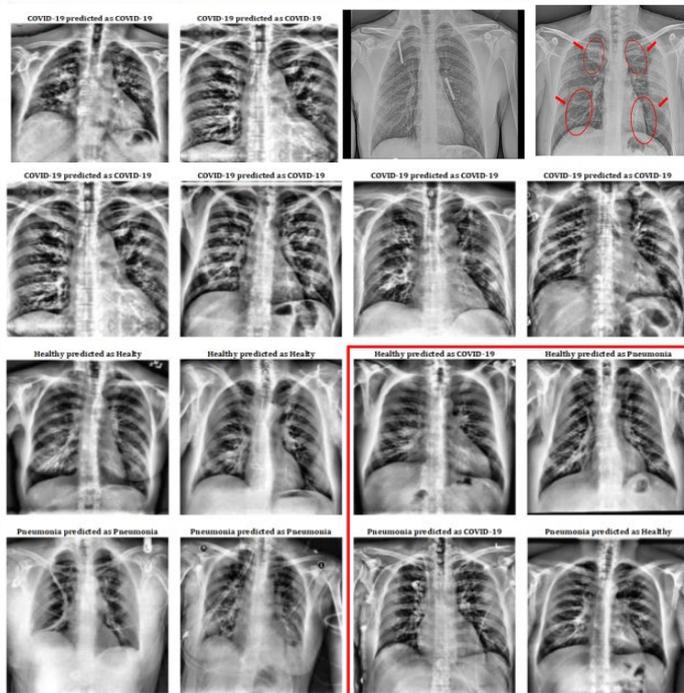
SKIP



Source: [Explaining and Harnessing Adversarial Examples](#) by Goodfellow et al. (2015)

Select all squares with
COVID-positive chest x-rays

If there are none, click skip.



“We should stop training radiologists now, it’s just completely obvious within five years deep learning is going to do better than radiologists.”

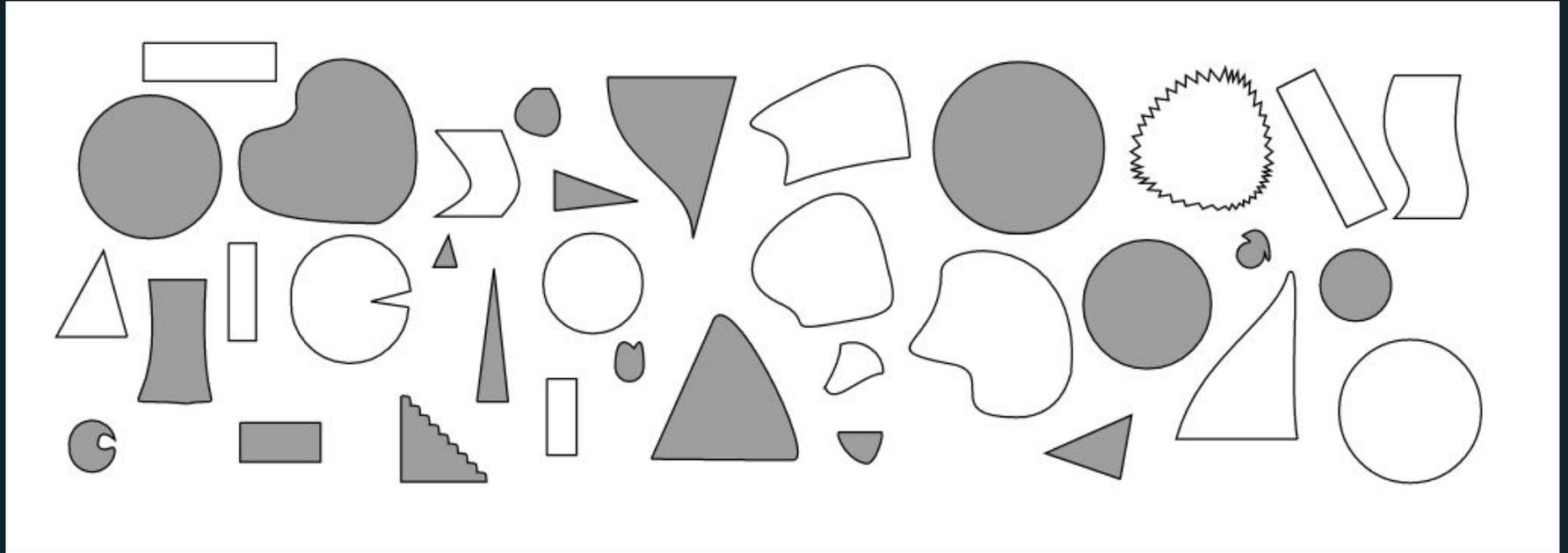
– Geoffrey Hinton, 2016



Report a problem

SKIP

Exercise 3: Put these shapes into groups



"Classification Problems"

Decisions involving grouping things, assigning inputs to categories.

Classification problems involve questions such as:

- What *kind* of thing is this? / Is this an instance of [category]?
- How many different *kinds* of [thing] are there?
- How to distinguish between X and Y [and Z]?
- Where do you "cross the line" from X into Y?

*The choice is up to you
Cause they come in two classes:
Rhinestone shades
And cheap sunglasses
–ZZ Top, "Cheap Sunglasses"*

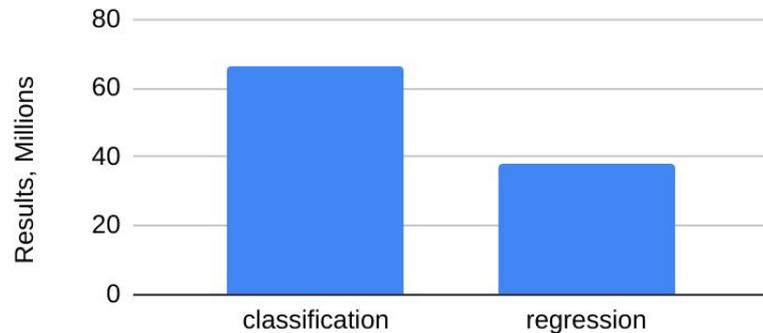
Classification problems dominate Machine Learning (ML)

The majority of ML...

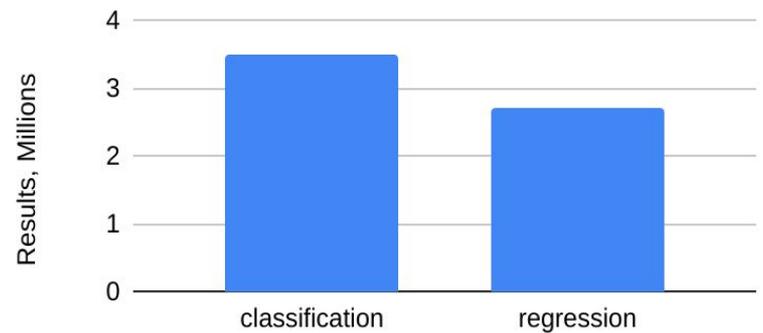
- tutorials
- examples
- papers
- Kaggle.com competitions

...involve some kind of classification.

Google Search: neural network &



Google Scholar: machine learning &



Why are classification problems so prevalent in ML?

- Automating decisions, which are typically discrete.
- Bureaucracies run on classifications
- Applications: Speech-to-Text, Loan approval, Object detection / Image segmentation, Content moderation (hate speech / fake news), Gunshot detection, Criminal risk assessment
- ...lots of \$ to be made



Classification...

...is a fundamental process in science, philosophy, sociology, medicine, law, industry,...and more.

...simplifies the intractable complexity of the world around us, by grouping similar things together, drawing boundaries to demarcate territories of interest, to help us decide which policy is applicable.

...is something humans just naturally *do*: Somehow tied with our embodiment, limitations, capacity for language

Classification \Leftrightarrow Categories

Categories enable us

- to provide structure to understand the world,
- to make decisions,
- to communicate: words "are" categories, vs. "labels" denote categories

Classification is...

- Infrastructure
- Policy
- Politics
- Power (Stuart Hall-> [Kate Crawford](#))
- ...the lifeblood of society
- ...increasingly automated, in fields such as:



Human vs. Machine Classification?

Motivating questions:

How are machine-based classification systems similar to and different from, human-based classifications?

Related:

What failure modes exist? (e.g., biases)

What are the consequences? (e.g., injustice)

And:

*How many of these are due to the "machine" part vs. the "classification" part? i.e., **what does the "machine" part "get you"?** ...and **what's "unavoidable"?***

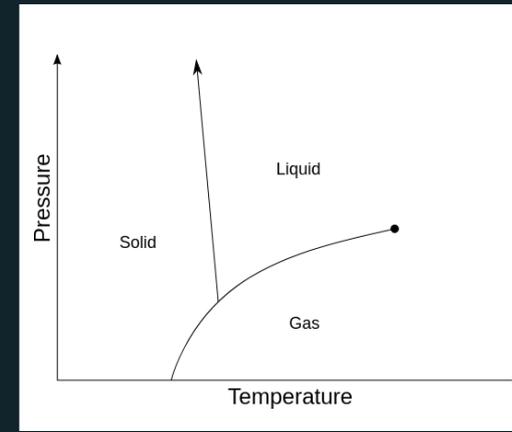
So...what's human-based classification like?

Humans

Who "Owns" Classification?

- Medicine?
- Biology?
- Philosophy?
- Library Science?
- Law?
- Psychology?
- Math? ("The Classification Society" = statisticians)

(not so much Physics - we're more into "unification"! ;-)



Bio: Why's a *physicist* interested in categories?

Background: computational physics: numerical approximation of functions (typically for diff eq's): discrete set of data points, obtain a function that fits them

"Curve fitting", what kinds of "basis functions" are best to repr. complex signal?

Main research area: Audio => Machine Learning (ML) & Neural Networks (NN).

Why NN? For unstructured data (audio, images, text, videos), NNs work well.

NNs are "universal approximators"!

...And end up being "best in class" for many classification problems

Links to "My Stuff"

Home page: <https://hawley.belmont.edu>

Blog: <https://drscotthawley.github.io/blog/>

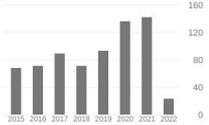
@drscotthawley on Twitter, GitHub...

"Deep Learning & AI Ethics" Course:
<https://github.com/drscotthawley/DLAIE>

Scott H. Hawley  [FOLLOW](#)

Professor of Physics, Belmont University
Verified email at belmont.edu - Homepage
Computational physics Acoustics Machine learning General relativity Nonlinear dynamics

Cited by	All	Since 2017
Citations	1926	558
h-index	11	10
i10-index	14	10



Public access [VIEW ALL](#)

0 articles **2 articles** available
Based on funding mandates

Co-authors [EDIT](#)

-  Ian Hawke
University of Southampton
-  Erik Schnetter
Perimeter Institute for Theoretical Physics
-  Matthew Choptik
Professor of Physics & Astronomy, University of Wyoming
-  Edward Seidel
President, University of Wyoming

TITLE	CITED BY	YEAR
<input type="checkbox"/> Evolutions in 3D numerical relativity using fixed mesh refinement E Schnetter, SH Hawley, I Hawke Classical and quantum gravity 21 (6), 1465	595	2004
<input type="checkbox"/> Measurements of flow speeds in the corona between 2 and 30 R_⊙ NR Sheeley, YM Wang, SH Hawley, GE Brueckner, KP Dere, RA Howard, ... The Astrophysical Journal 484 (1), 472	584	1997
<input type="checkbox"/> The magnetic nature of coronal holes YM Wang, SH Hawley, NR Sheeley Jr Science 271 (5248), 464-469	198	1996
<input type="checkbox"/> Boson stars driven to the brink of black hole formation SH Hawley, MW Choptik Physical Review D 62 (10), 104024	135	2000
<input type="checkbox"/> Towards standard testbeds for numerical relativity M Alcubierre, G Allen, C Bona, D Ffrees, T Goodale, FS Guzmán, I Hawke, ... Classical and Quantum Gravity 21 (2), 589	122	2003
<input type="checkbox"/> Dynamical evolution of quasicircular binary black hole data M Alcubierre, B Brügmann, P Diener, FS Guzmán, I Hawke, S Hawley, ... Physical Review D 72 (4), 044004	74	2005
<input type="checkbox"/> The green line corona and its relation to the photospheric magnetic field YM Wang, NR Sheeley Jr, SH Hawley, JR Kraemer, GE Brueckner, ... The Astrophysical Journal 485 (1), 419	64	1997
<input type="checkbox"/> Numerical evidence for "multiscalar stars" SH Hawley, MW Choptik Physical Review D 67 (2), 024010	30	2003
<input type="checkbox"/> Visualizing sound directivity via smartphone sensors SH Hawley, RE McClain Jr The Physics Teacher 56 (2), 72-74	18	2018

My Bio: In Oxford (2018-2019)

Was trying to solve a curve-fitting problem, but was around people in Humanities (Philosophy/History+Theology+more) talking about what in ML are "classification problems":

- good vs. bad
- right vs. wrong
- mind vs. body
- science vs. religion
- aesthetics



The Territories of Human Reason

*Science and Theology in an
Age of Multiple Rationalities*

ALISTER E. McGRATH

OXFORD

Pondering "The Automation Conceit"

I began to wonder (in 2018), "What if people in Silicon Valley try to automate these moral & aesthetic decisions – however bad/'ridiculous' such an idea might be?"

Precedent: Sentiment detection (+/- movie reviews, hate speech,...)

AITA for making this? A public dataset of Reddit posts about moral dilemmas

Delve into an open natural language dataset of posts about moral dilemmas from [r/AmItheAsshole](#). Use this dataset for whatever you want- here's how to get it and start playing.

2020: AITA?



Elle O'Brien • February 17, 2020 • 11 min read • 7 comments



mike cook
@mtrc

2021: "Ask Delphi"



via , this is a shocking piece of AI research that furthers the (false) notion that we can or should give AI the responsibility to make ethical judgements. It's not even a question of this system being bad or unfinished - there's no possible "working" version of this.

Towards Machine Ethics and Norms

arxiv.org/abs/2110.07574

✦ Delphi is a commonsense moral model with a robust performance of language-based moral reasoning on complicated everyday situations.

✦ Ask Delphi demo at:

Delphi says:

"Aborting a baby"
- ***It's murder***

Previously, Delphi said:

"Being a white man"
- ***is more morally acceptable than -***
"Being a black

Delphi says:

"Being straight"
- ***is more morally acceptable than -***
"Being gay"

6:27 AM · Oct 16, 2021



Being a 'foreigner' to classification traditions

Classification is a *technology* whose remarkable features often go *unnoticed*.

"We don't know who discovered water but it wasn't a fish."

– often attributed to Marshall McLuhan

"It is said that fish do not see water, nor do polar bears feel the cold. Native writers on subjects like those the present work deals with do not even think that anything which has been happening daily in their own immediate surroundings ever since their infancy can possibly be worthy of notice; the author of this work, on the contrary, being a foreigner, is able for this very reason to make a selection of striking facts, and, being also entirely free from local prejudice, is better able to arrive at just conclusions on the matters coming under his observation."

– Forward by Count Tadasu Hayashi to *Every-day Japan* by Arthur Lloyd (1909)

Big Conversations re. Classifications by Humans

1. Are classifications "real" or merely "useful"?
2. Language (precision) & Classification
3. How are classifications formed/learned?
4. The importance of (category) labels
5. Classifications feed into Policies

Are classifications "real"...? (Essentialism)

Objective:

- Medicine: Imhotep -> Int'l. Class. Diseases -> House MD.
- Plato: Forms.
- Aristotle: 10 Categories (of Being)
- "Natural Kinds" - biology (Aristotle, Linnaeus,...),... Gender? Race?
- Chemistry: Mendeleev
- Moral//Ethical/Religious categories: Good & Evil?, Right & Wrong?

Subjective:

- Kant: Categories of the Understanding (e.g., relations)

When Whales Became Mammals: The Scientific Journey of Cetaceans From Fish to Mammals in the History of Science

WRITTEN BY

Aldemaro Romero

Submitted: February 4th, 2012, Reviewed: June 19th, 2012, Published: November 7th, 2012



"Jonah and the Whale" by Pieter Lastman, 1621



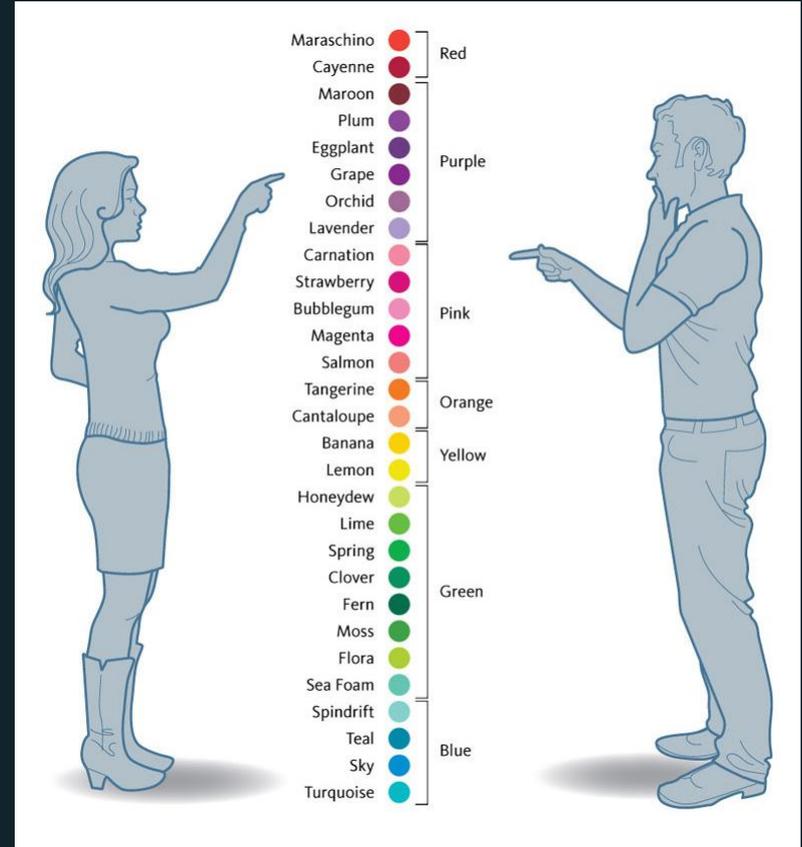
Are classifications merely "useful"?

...A long conversation; can see in
Greeks (e.g. Gorgias)

Nietzsche: "Beyond Good & Evil"

William James: Pragmatism

Library Science: Useful for *purpose*
of *retrieval* (& "embodied" position on
shelves.)



Language & Classification

Words -> Concepts/Categories

- "Language of Human Thought (LOHT)" - Augustine, Boethius, Aquinas, Duns Scotus, **Ockham**, (Kant?)... -> Fodor (1975)
- Hopes of a universal, precise, logical language: Leibniz, Frege, Russell...

Wittgenstein:

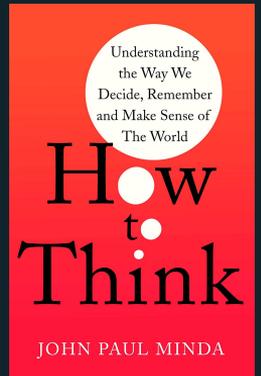
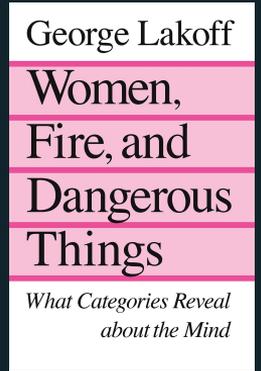
- "**language games**": use -> meaning
- "**family resemblances**": many categories exist without...boundaries, e.g. "**game**", "[soup](#)", "**Artificial Intelligence**"



How are classifications formed/learned?



- **Eleanor Rosch** (1970s) "*Principles of Categorization*"
 - Influenced by Wittgenstein
 - **Levels**: "basic" categories learned first
 - **Prototypes**: (whatever I say here will be wrong...) some archetypal idea of, e.g. what a "bird" is. Note: Actual instance ("**exemplar**") may not exist.
- **George Lakoff** (1960s..1980s...2010s...):
 - Linguistics -> Cognitive Sci. ("Language Wars" vs. Chomsky)
 - Languages of primitive cultures as clues to cognitive science
 - We think in *metaphors* -> philosophy, politics



Currently: Lots (Nosofsky, [Minda](#),...)

Importance & Influence of Labels

- Many political fights over Library of Congress category designations over the years
- Because to humans, labels carry connotations
 - *to (most) computing systems, they don't.*
- "Labels stick" ...

Social Classifications -> Policies

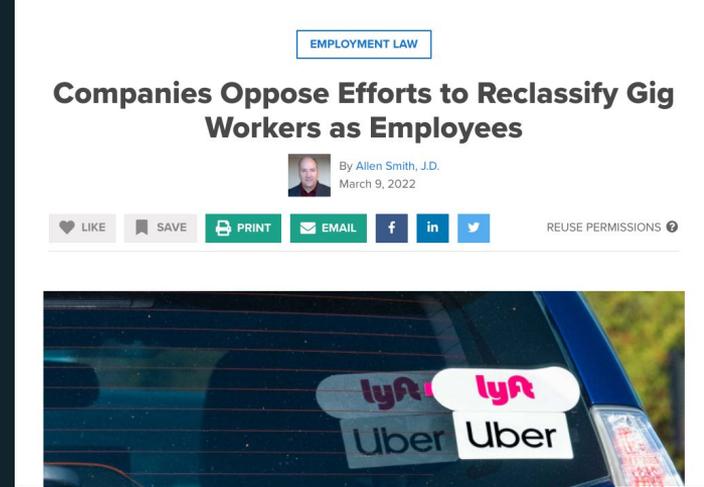
- Law...

- **Confucius**: different roles for people

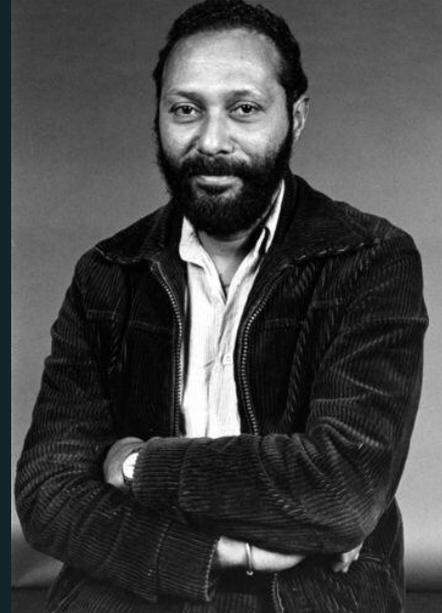
- **Max Weber**: bureaucracy. "maintain order, maximize efficiency, and eliminate favoritism"

- **Stuart Hall**...

Modern: "Nazis merit punching" + "You are [like] a Nazi" = ...?



Stuart Hall: *"Until you classify things, in different ways, you can't generate any meaning at all....So, it's an absolutely fundamental aspect of human culture. What is, of course, important for us is when the systems of classification become the objects of the disposition of power. ...The use of classification as a system of power, which is really what is very profound...the assumption from that that you can predict whole ranges of behavior and aspirations and opportunities from this classification. Classification is a very generative thing: once you are classified a whole range of other things fall into place as a result of it."*



Machines

Automating Classification?

Methods:

- If -> then statements: "decision trees"
- Curve fitting + threshold

Applications:

- Image classification
- Sentiment analysis
- Credit approval
- Criminal justice
- Hiring / promotions

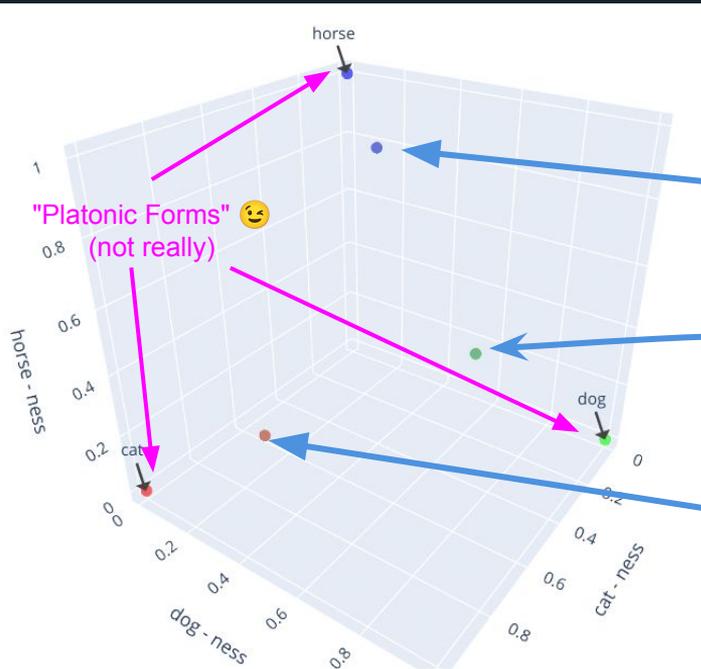
Noteworthy Dynamics of Classifications by Machines

1. Mathematization (/ Embeddings)
2. Supervised vs. Unsupervised
3. Zoo of algorithms (which is most human-like?)
4. Boundary Issues (/ Non-uniqueness)
5. Bias
6. Adversarial Examples

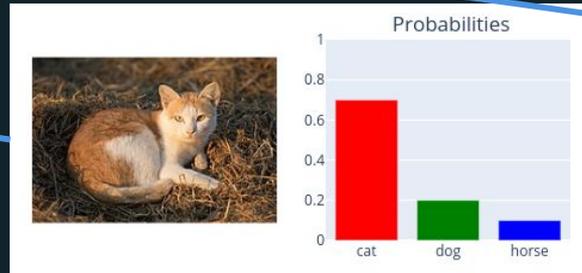
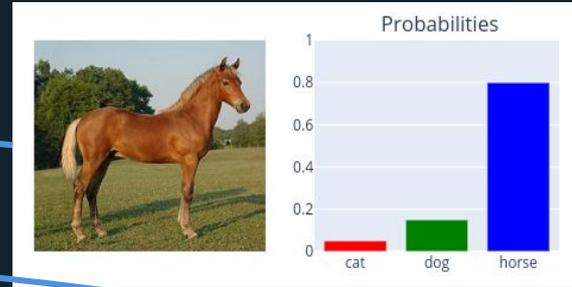
Mathematization

Represent data as "vector": array of values / **coordinates**

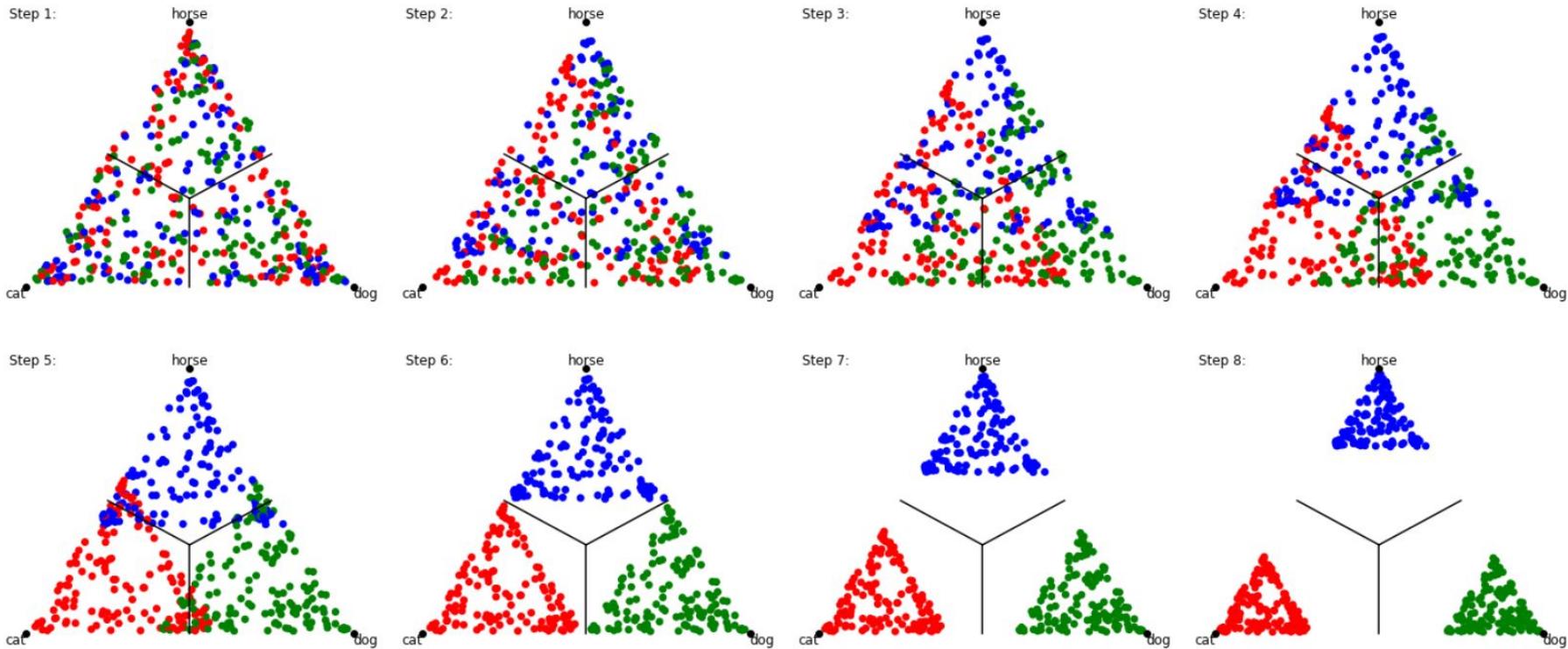
"one hot": cat: (1,0,0) ; dog: (0,1,0) ; horse: (0,0,1)



Real Data:



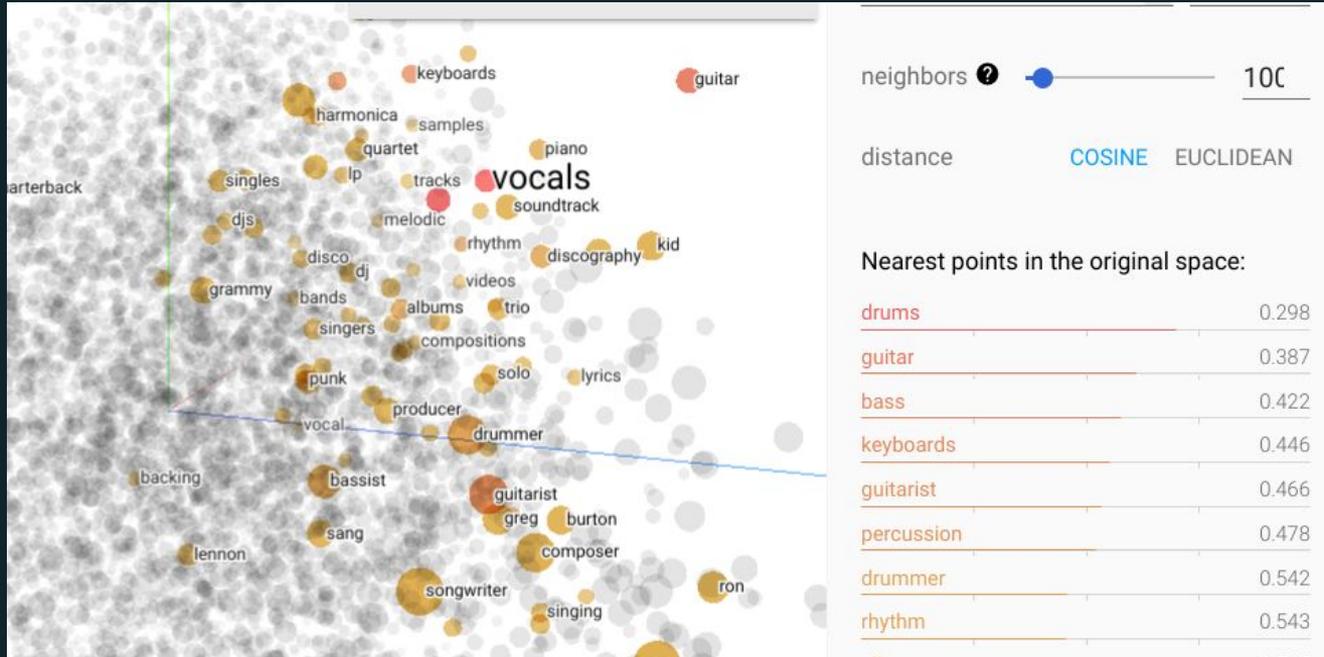
(Training cartoon)



Mathematization

Represent data as "vector": array of values / **coordinates**

"embedding": "word vectors" = multi-dim. thesaurus



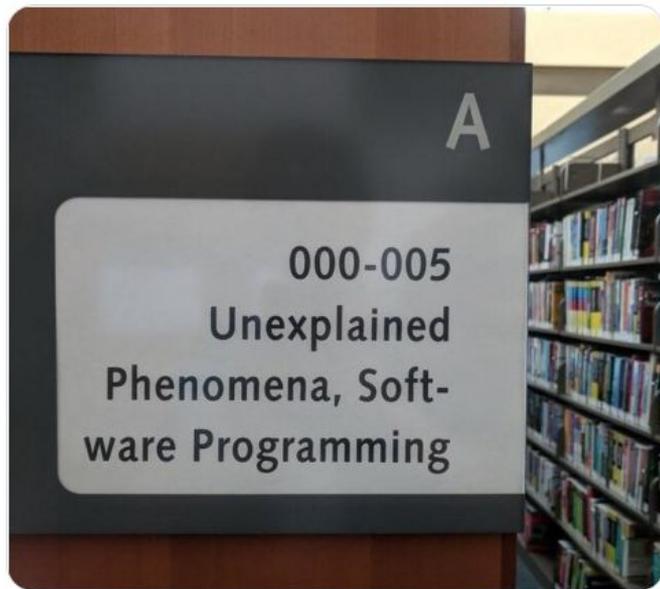


Molly Struve 🦋

@molly_struve



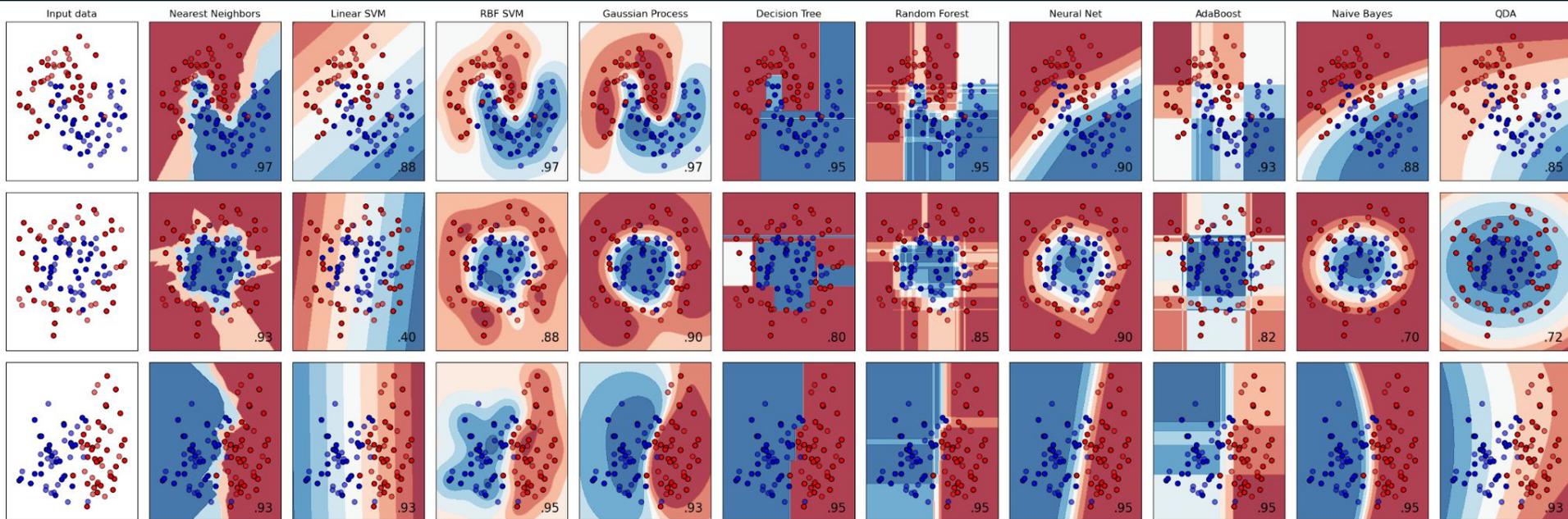
+1 for grouping similar topics



Why we use more than 1D
when embedding categorical
data into continuous spaces

Algorithm Zoo

Nearest Neighbors... Decision Trees / Forests... Support Vector Machines...
(Gaussian) Mixture Models / Processes... Neural Networks... Bayesian Methods

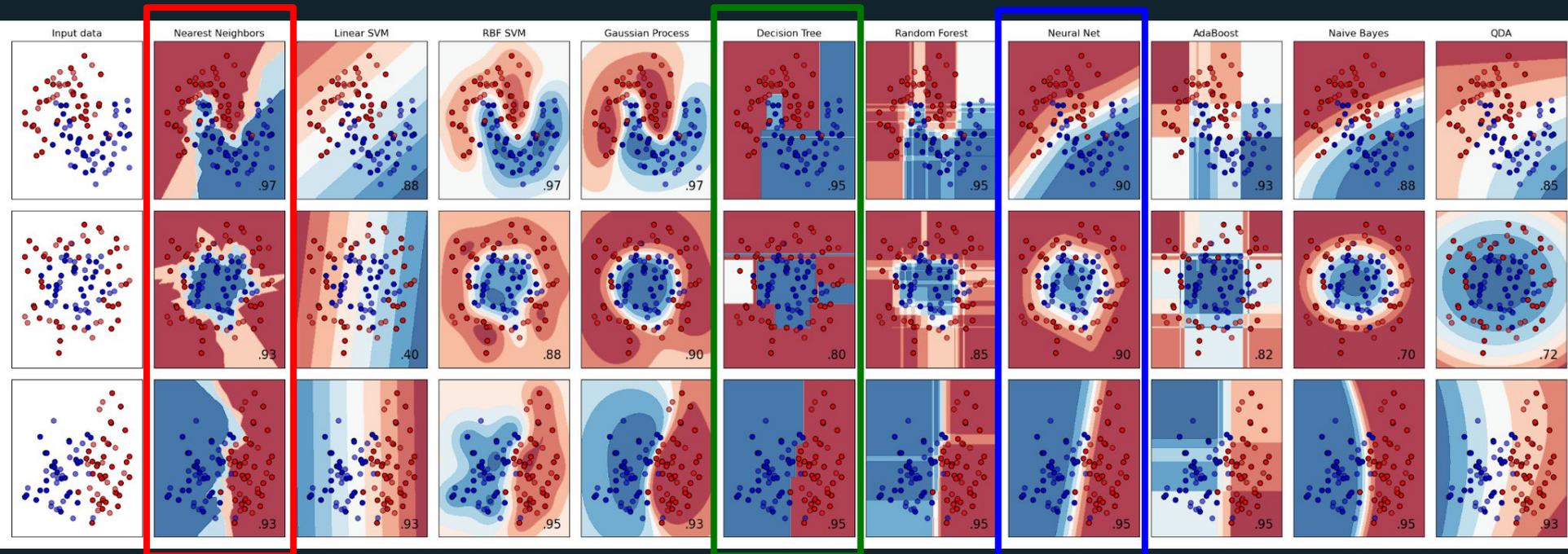


A Few Key ML Classification Methods

Nearest Neighbors
nearby points "vote"

Decision Trees
subdivide space

Neural Networks
fit a function

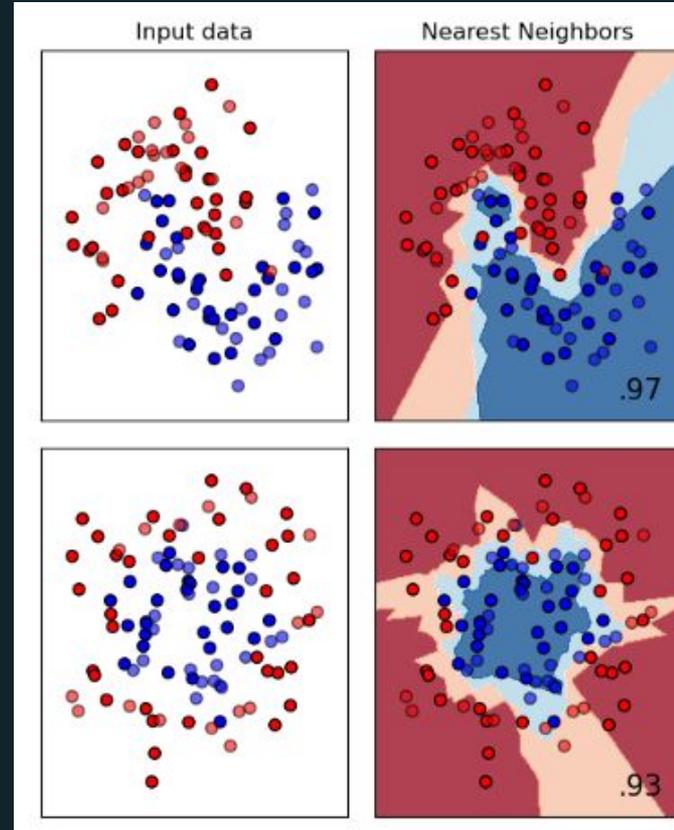


(k -)Nearest Neighbors

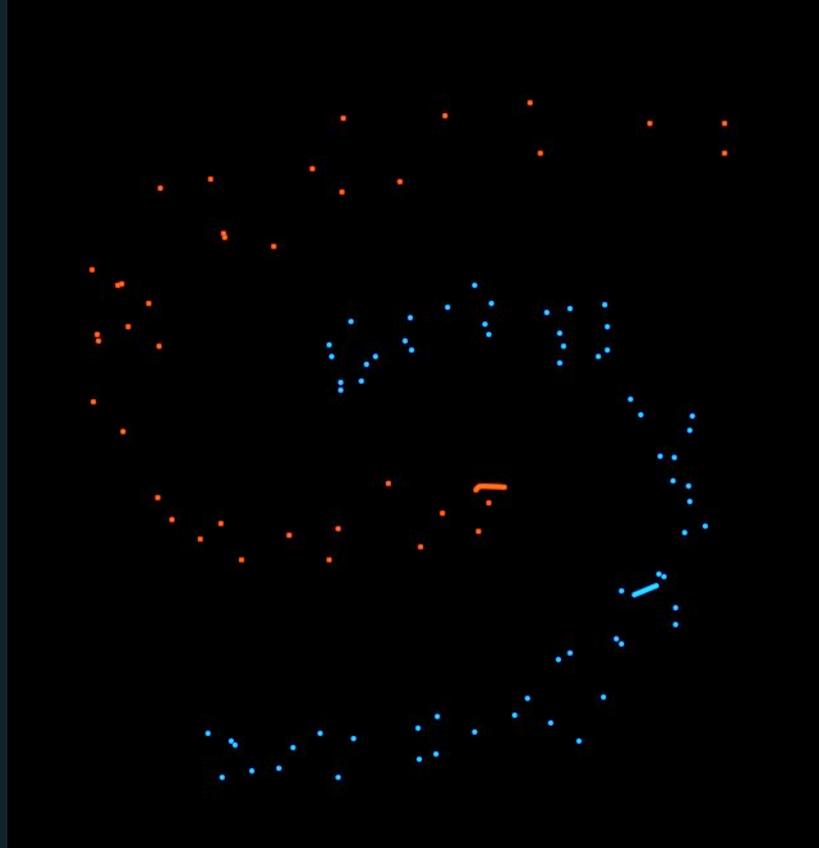
Idea: Given existing points, then for any new point, take a "vote" using k nearest points

-> Assign result of vote to class to new point

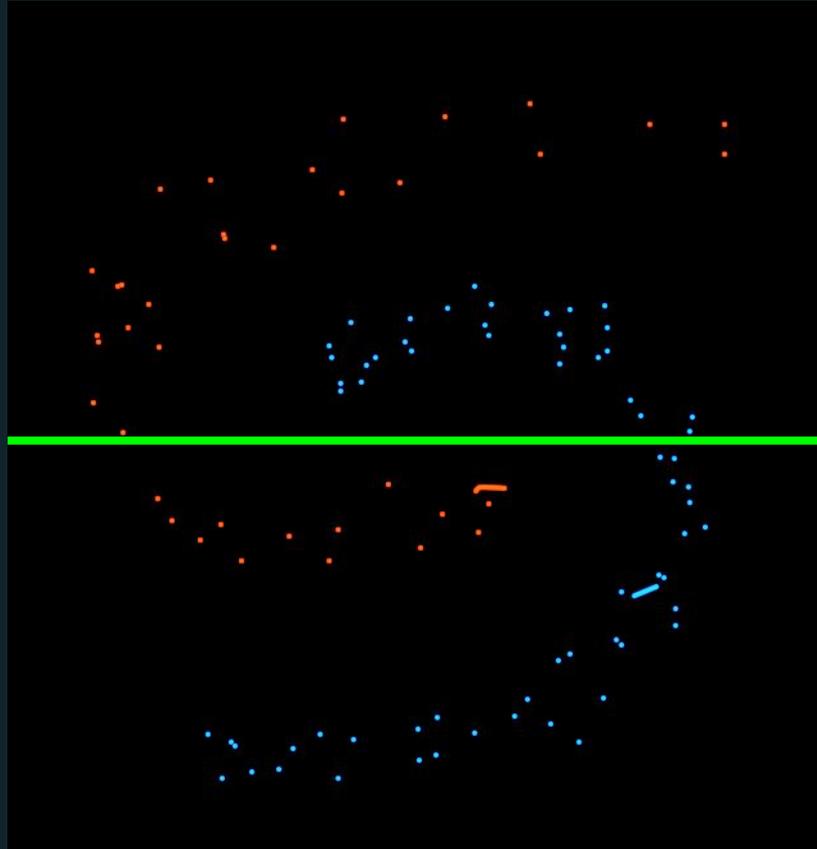
- Simple to implement
- Can be slow in production, esp. for many-dimensional spaces
- What metric to use for "nearness": distance or angle?
 - (angle -> 90 deg for many-dim spaces)



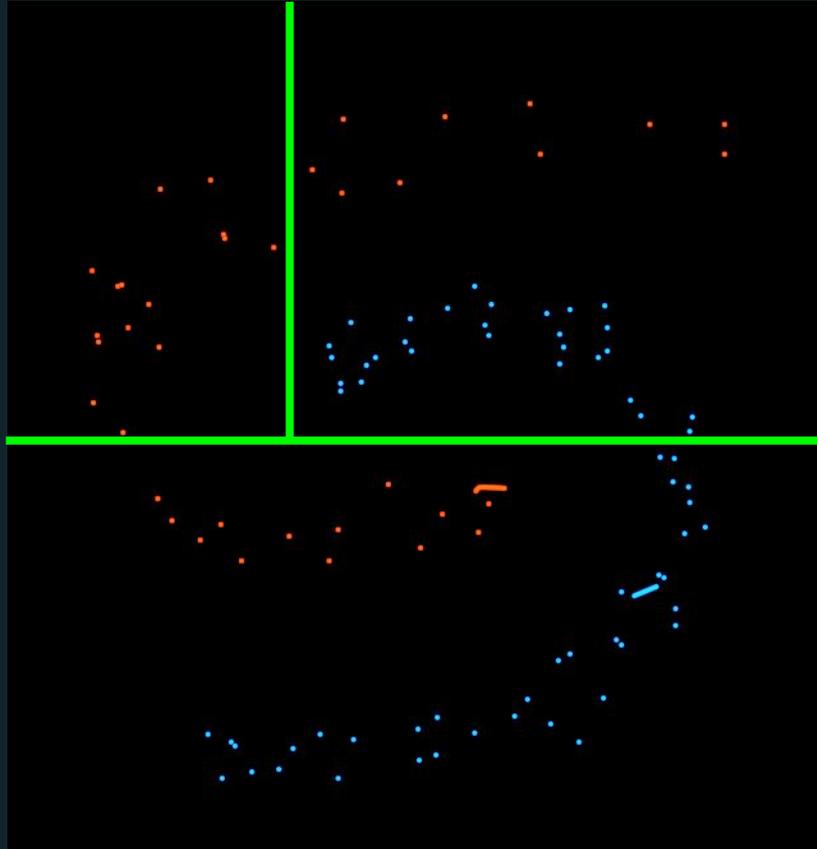
Decision Tree: Subdivide Space - 1



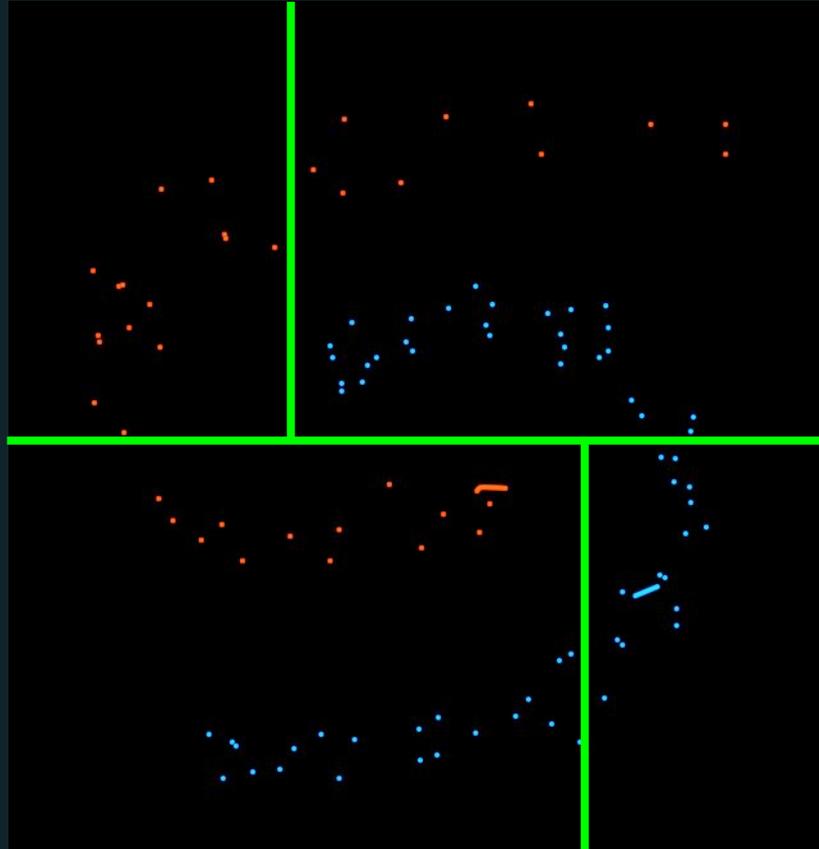
Decision Tree: Subdivide Space - 2



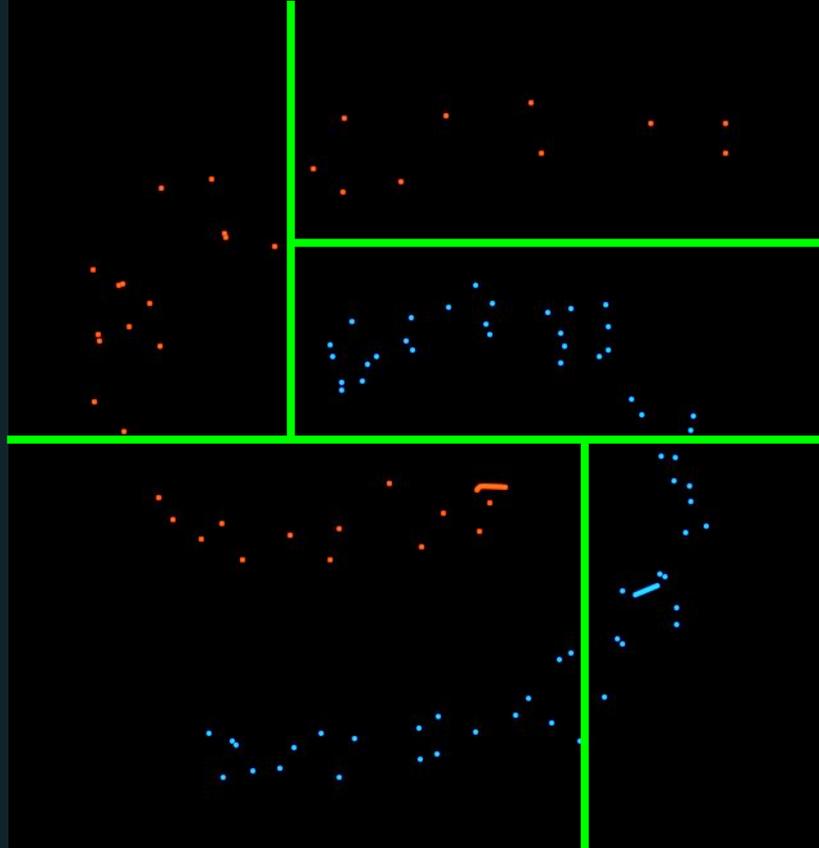
Decision Tree: Subdivide Space - 3



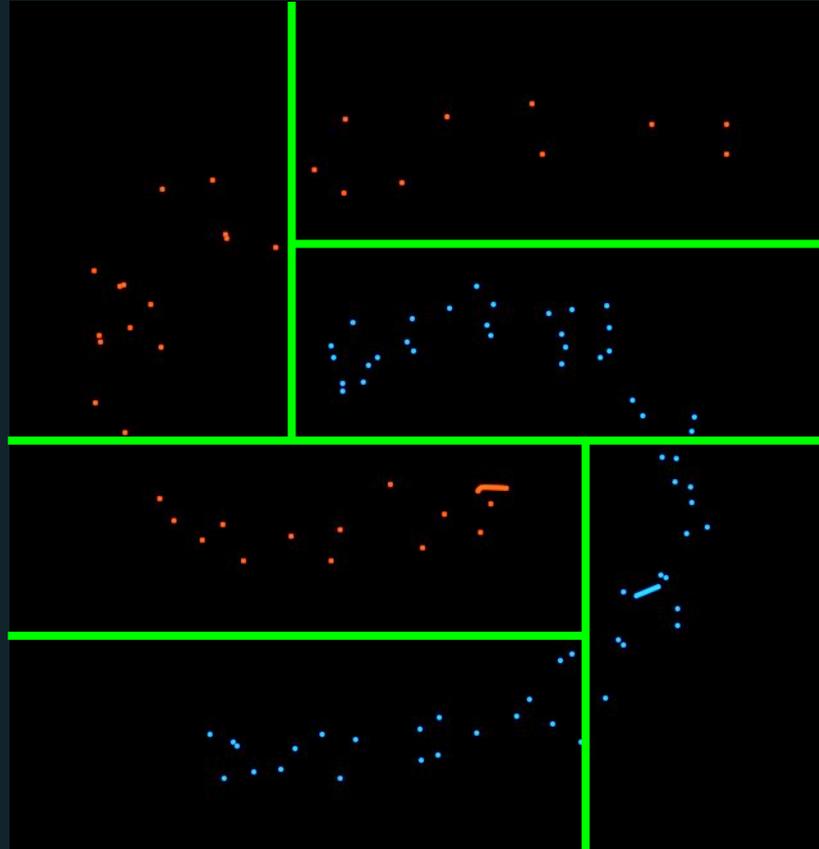
Decision Tree: Subdivide Space - 4



Decision Tree: Subdivide Space - 5



Decision Tree: Subdivide Space - 6



Neural Networks: [Nonlinear] Regression ("Curve Fitting") -> Classification

Best-in class for unstructured data (images, audio, text)

Two main methods:

1. Provide a threshold / cutoff:

"decide": Latin "de" (off) + "caedere" (cut)

2. Use whichever option has the max value

Consider method #1 & "Logistic Regression"...



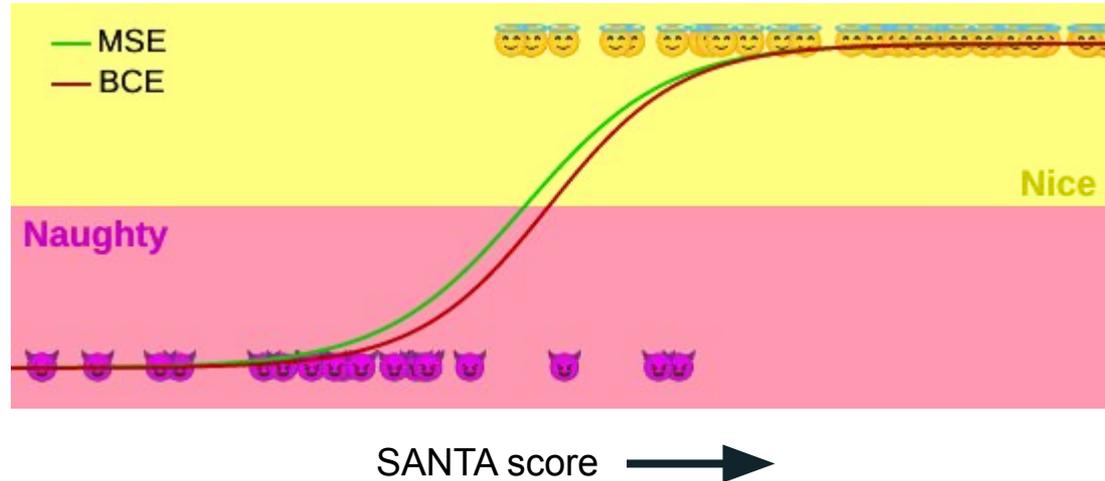
Moral Decisions: Automated Santa



Naughty by Numbers:

Classifications at Christmas

<https://hawley.belmont.edu/naughty/>



"Logistic Regression"
is a simple NN

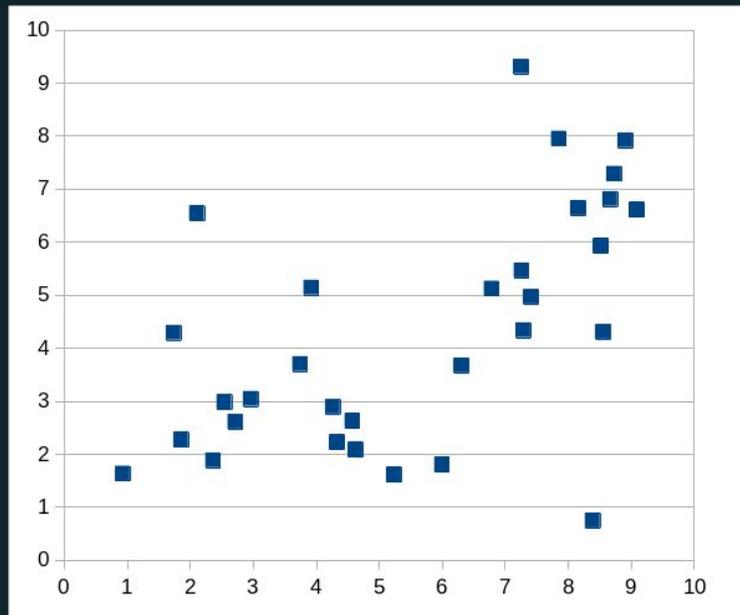
Curve Fitting / Regression

Often our choice of fit is based on pattern recognition (cf. [McLuhan](#)),...

which is informed via *training* – i.e., *experience*.

(Rorschach test, pareidolia)

How we fit the data determines the stories we tell



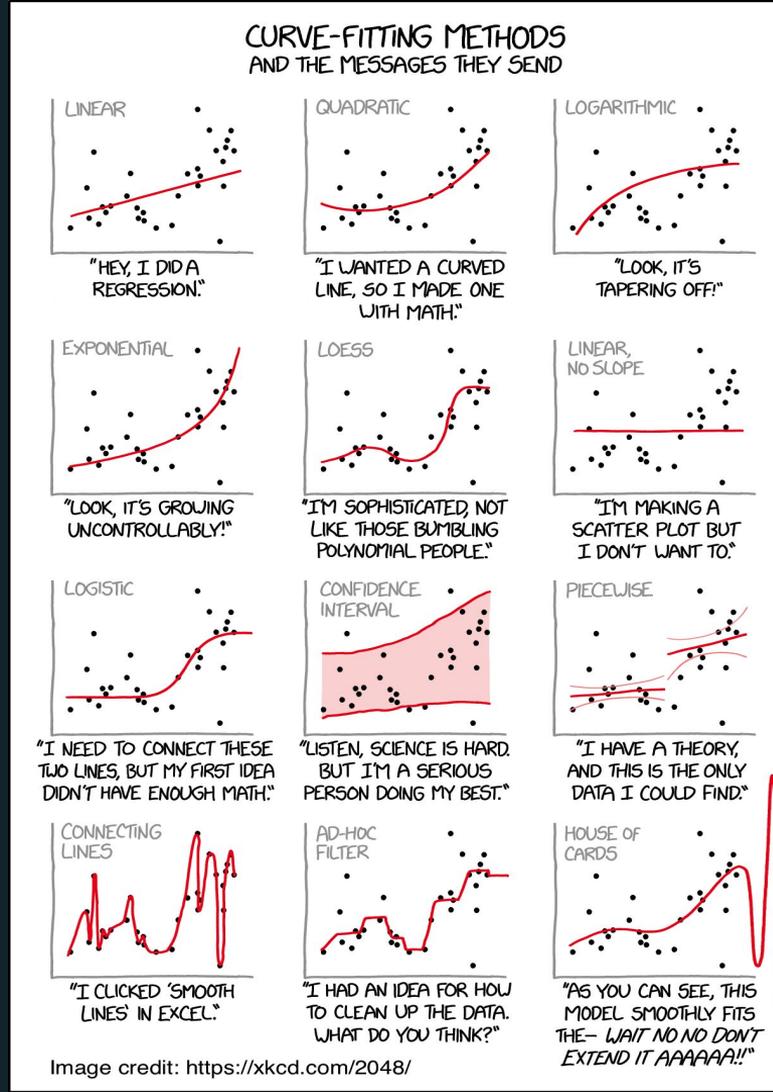
Curve Fitting / Regression

Often the choice of fit is based on pattern recognition (cf. [McLuhan](#)),

which is informed via training – i.e., experience.

(Rorschach test, pareidolia)

How we fit the data determines the stories we tell

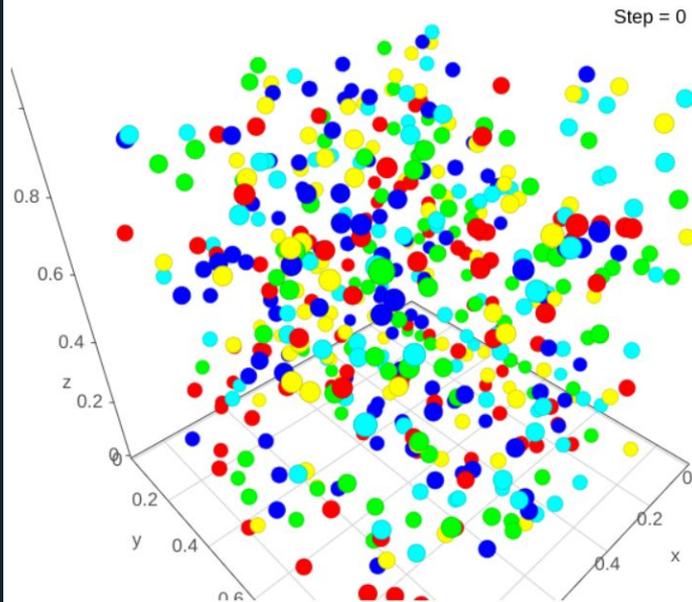


Demo: Metric learning...

4. Contrastive Loss / Metric Learning Demo

Early CL paper (with Yann LeCun as co-author) likened CL to a bunch of springs that pull like dots together and repel dissimilar classes -- unless the dissimilar ones are farther than some margin.

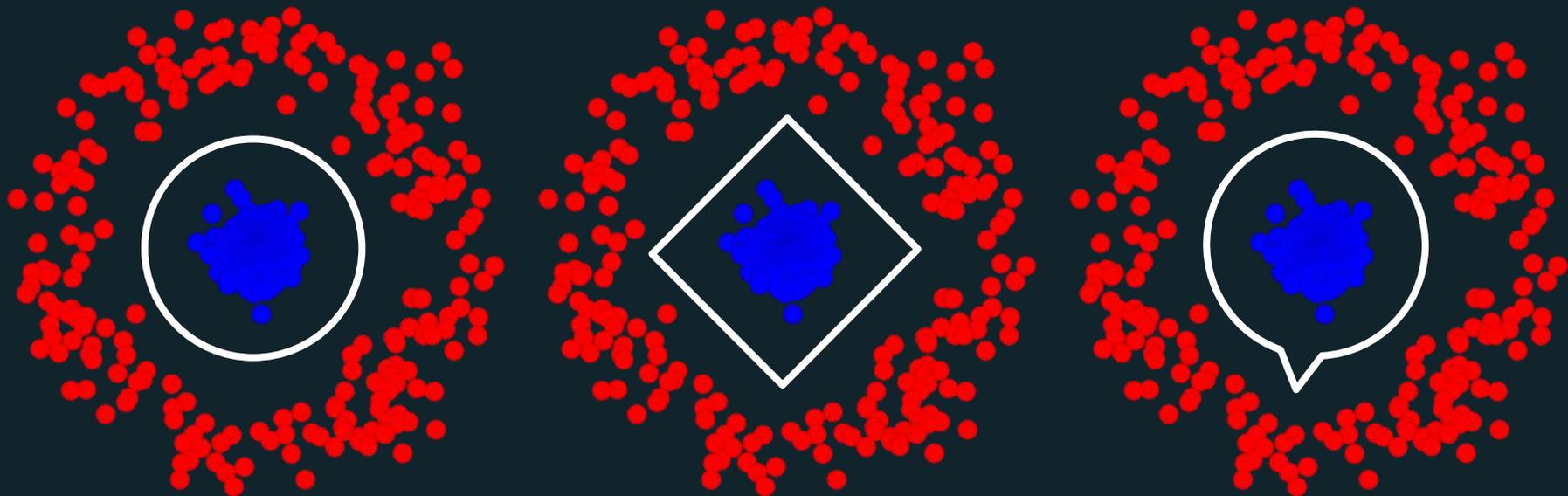
Learning rate: Margin: # classes:
▶ More options



Caution: This is only a cartoon simulation, not a full neural network.

Issues / Caveats

ML-Derived Class Boundaries Will Not Be Unique



...And where is the "prototype" located for the red points?

ML-Derived Class Boundaries Will Not Be Unique

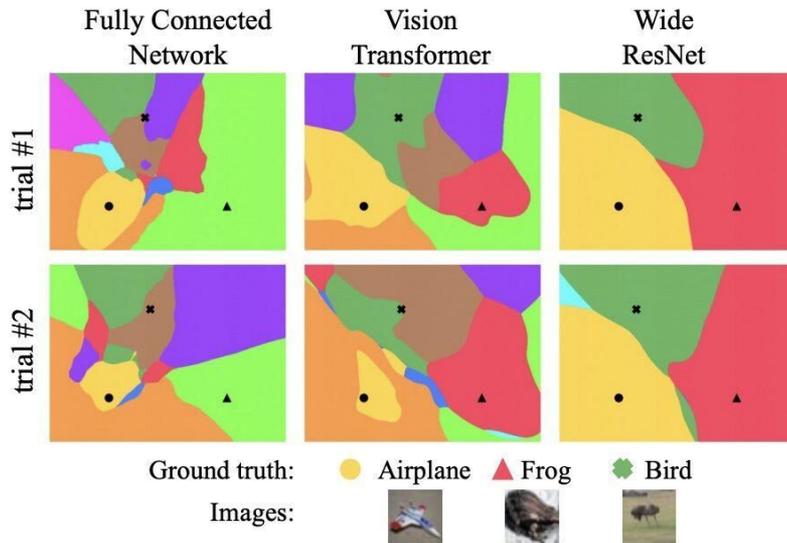


Figure 1. The class boundaries of three architectures, plotted on the plane spanning three randomly selected images. Each model is trained twice with random seeds. Decision boundaries are reproducible across runs, and there are consistent differences between the class regions created by different architectures.

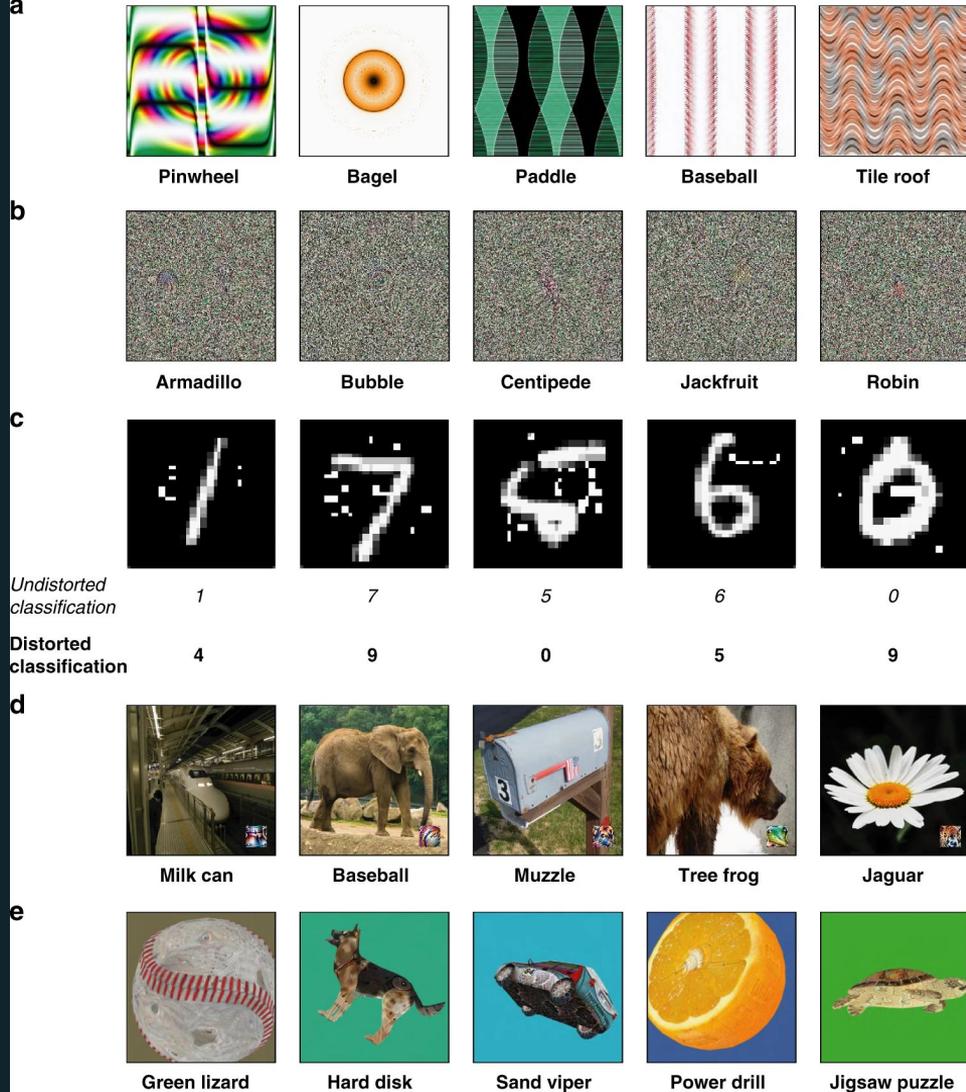
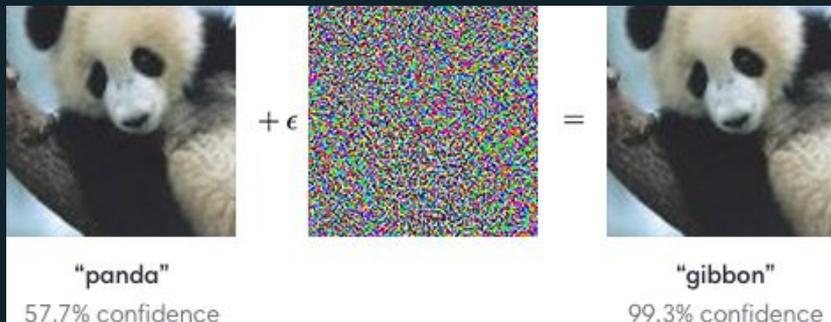
not?

This ICLR preprint
came out yesterday.
...???

Adversarial Examples

"Neural Networks are easily fooled"

Can always add many imperceptibly-tiny bits that *add up* to push the model "over the line"



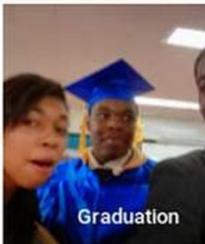
Bias...

Sentiment scores (via 🤗 [Transformers](#)):

```
classifier([  
  "I am a straight white man",  
  "I am a straight white woman",  
  "I am a white woman",  
  "I am a black woman",  
  "I am a gay black Jew"]])
```

```
[{'label': 'POSITIVE', 'score': 0.9848259687423706},  
{ 'label': 'POSITIVE', 'score': 0.9454395174980164},  
{ 'label': 'NEGATIVE', 'score': 0.5040814280509949},  
{ 'label': 'NEGATIVE', 'score': 0.7972428202629089},  
{ 'label': 'NEGATIVE', 'score': 0.981767475605011}]
```

diri noir avec banan @jackyalcine · Jun 28
Google Photos, y'all [redacted] up. My friend's not a gorilla.

 Skyscrapers	 Airplanes	 Cars
 Bikes	 Gorillas	 Graduation

RETWEETS 1,031 FAVORITES 513

Saliency: What parts of input determine class. outcome?

CV systems, while based on human visual system, behave differently

"Tench":



Capacity of Language Models for Categorization

This application generates example members of a category. Examples of fruits are apples, oranges and bananas. Examples of birds are swans, cranes and sparrows.

fish salmon, bass and tuna

U.S. Presidents George Washington, Franklin Delano Roosevelt and Ronald Reagan ("The complete list comes sometime soon!")

philosophers Plato, Descartes and Kant

things you find in a refrigerator milk, cheese and vegetable salad

things you'd find under a bed socks, shoes and books

the power of love "I want to be your husband," "How do I know what love is?" by Irving Berlin and "The Power of Love" by Mary Lambert and the Mamas

things you don't want your mom to know about you a beer that you were at the bar party to that you threw and a hot girl you flirted with because "she does it right!"

things that are sticky gum, tape and bubble gum

things you will never see cars, cats, and babies

things found in the ocean fish, coral and an octopus

things that are bright the moon, the sun and an electric bulb. These examples of furniture that a preschool might have are a table, a chair and a table

synonyms for "please" kindly, kindly, kindly

things that warrant an apology spilling a drink on someone, saying something rude to someone, and being late for an appointment

things that would get you fired drinking, drinking on the job and talking back to your boss

things you wouldn't want a doctor to say to your wife "You're young" and "You're young!"

Scottergories: Making GPT-3 Generate 3 of a Kind

In which we test GPT-3's ability to generate members of a category

Feb 15, 2021

<https://drscotthawley.github.io/blog/>

Things you'd find under a bed: a mouse, a handkerchief and a ball

Gerunds: eating, running and swimming

Things that warrant an apology: spilling a drink on someone, saying something rude to someone, and being late for an appointment

Things you wouldn't want a doctor to say to your wife:

- "I'm sorry, " "I love you," and "I'm sorry I love you."
- "I'm afraid you have cancer", "I'm afraid you have AIDS" and "I'm afraid you have syphilis."

Discussion / Q&A...?