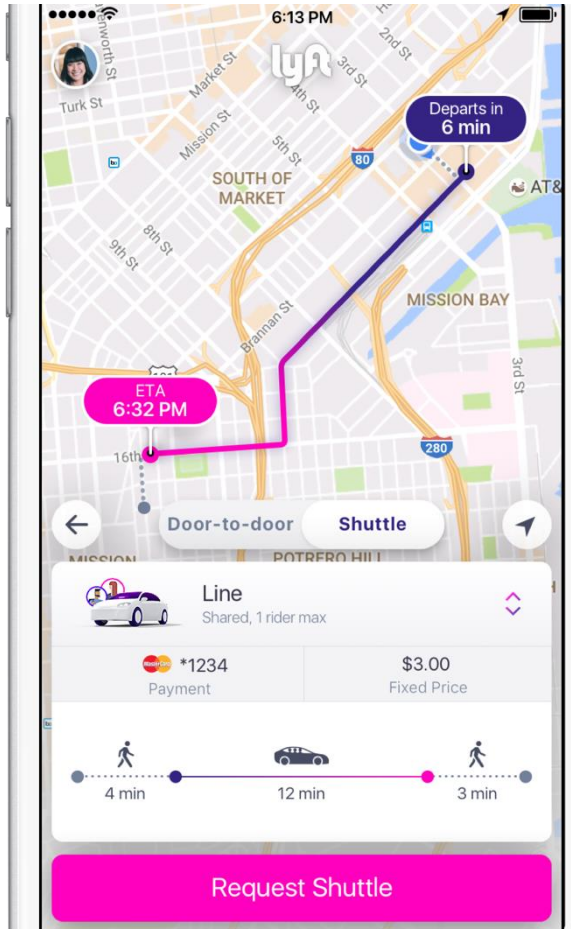
CS181: Principles of Machine Learning and AI

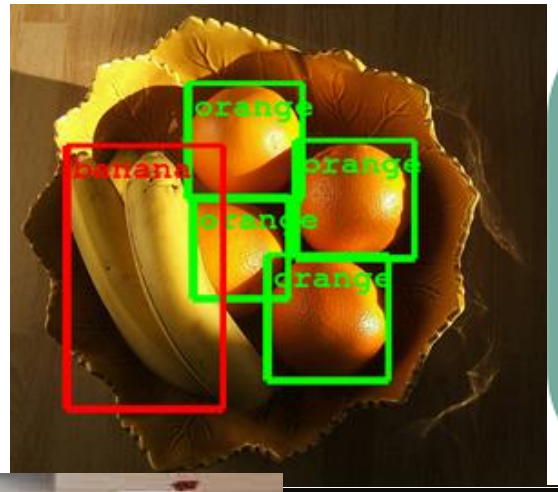
Spring 2024

David Alvarez-Melis and Finale Doshi-Velez
Harvard Computer Science

Staff: Alex Cai, Charumathi Badrinath (head TFs), Angela Li, Brian Ham, Audrey Chang, Ahmad Abdel-Azim, Gabriel Sun, Shirley Zhu, Jeffrey Xu, Alyssa Huang, Derek Zheng, Sam Jones, Neeyanth Kopparapu, Andrew Zhao, Joshua Park

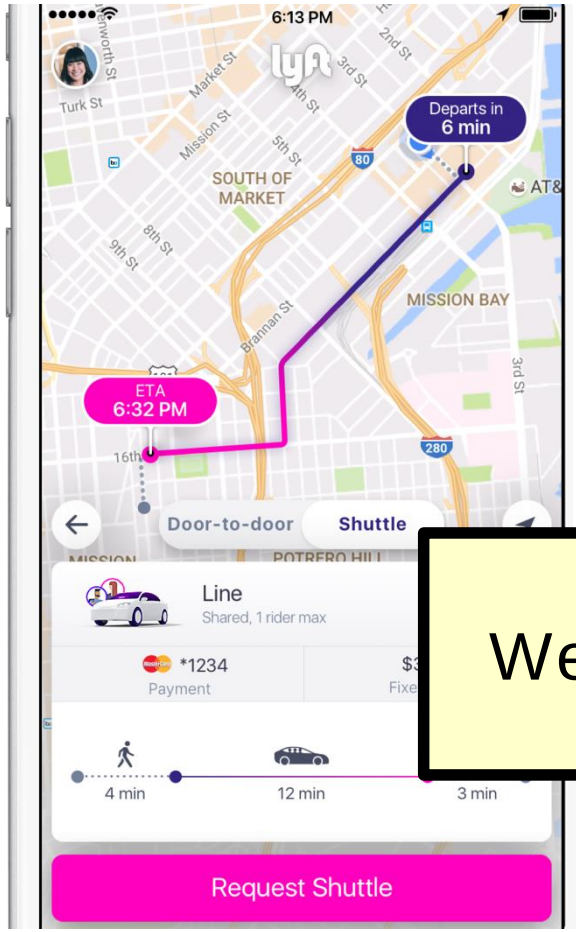
What this course is about

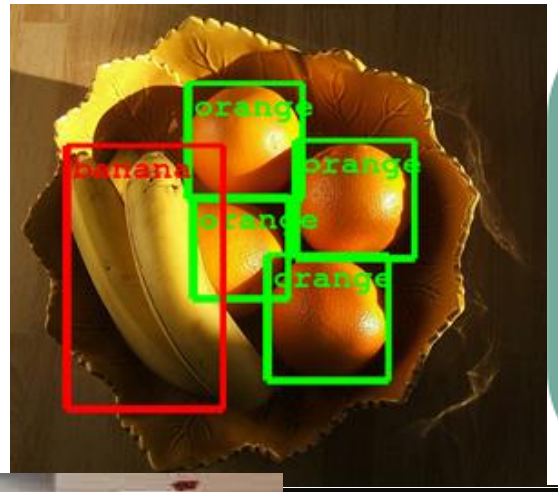


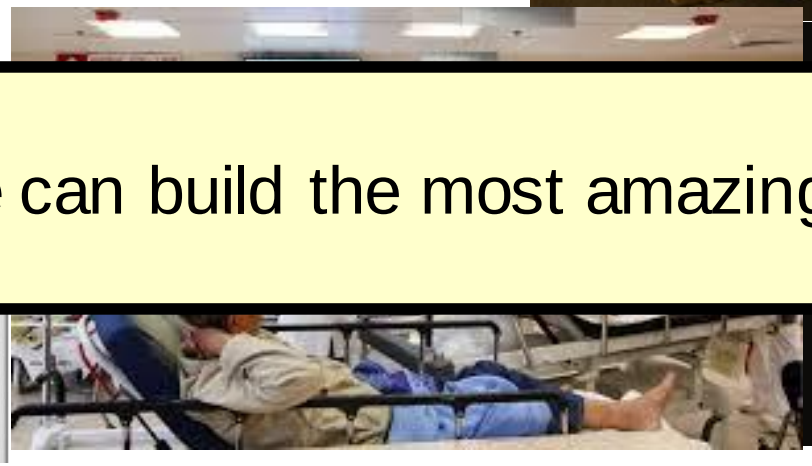


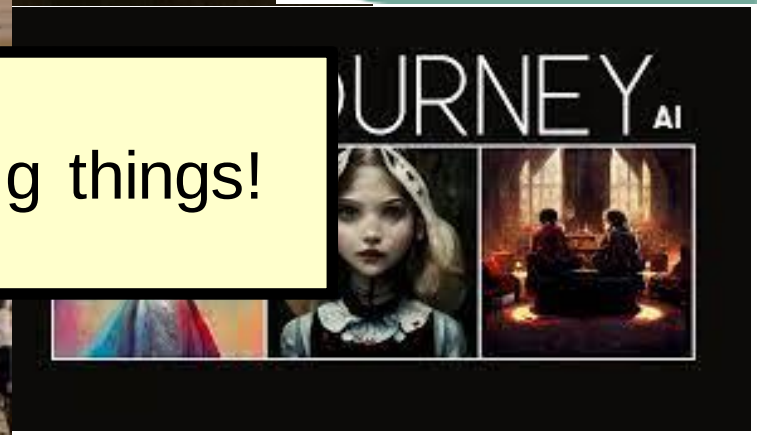
https://s.aolcdn.com/hss/storage/midas/cb433fb6c234b014394ad56714ceee8b/200707597/google-image-recognition-tech.jpg
https://cdn.shopify.com/s/files/1/0684/3445/products/Partner_AmazonEcho_small_031932b3-c31d-4977-a922-5ebbe052bad6.jpg?v=1497504913
https://cdn.lyft.com/brochure/shuttle-screenshot.1e79cd66.png
https://cdn3.i-scmp.com/sites/default/files/styles/980x551/public/images/methode/2016/05/29/b0dc6826-2577-11e6-80b1-a87df553e801_1280x720.JPG?itok=QVZxRMvy

What this course is about









We can build the most amazing things!

https://s.aolcdn.com/hss/storage/midas/cb433fb6c234b014394ad56714ceee8b/200707597/google-image-recognition-tech.jpg
https://cdn.shopify.com/s/files/1/0684/3445/products/Partner_AmazonEcho_small_031932b3-c31d-4977-a922-5ebbe052bad6.jpg?v=1497504913
https://cdn.lyft.com/brochure/shuttle-screenshot.1e79cd66.png
https://cdn3.i-scmp.com/sites/default/files/styles/980x551/public/images/methode/2016/05/29/b0dc6826-2577-11e6-80b1-a87df553e801_1280x720.JPG?itok=QVZxRMvy

What this course is about

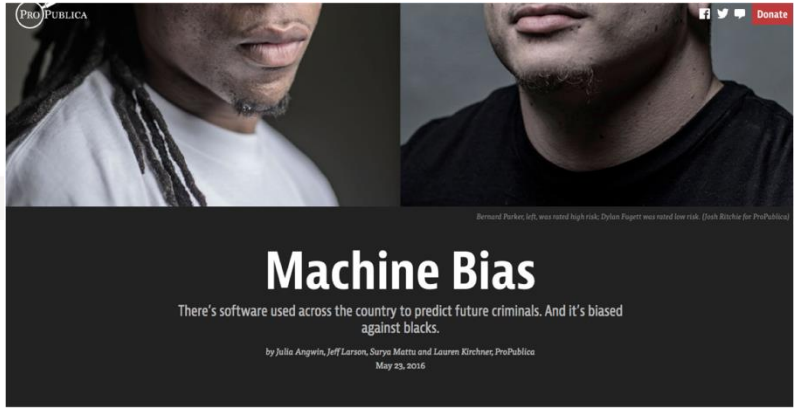
BUSINESS NEWS OCTOBER 9, 2018 / 11:12 PM / 2 MONTHS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

HEALTH TECH

3 lessons from Stanford's Covid-19 vaccine algorithm debacle

By CASEY ROSS @caseymross *and* ERIN BRODWIN @erbrod / DECEMBER 21, 2020



https://twitter.com/geraldmellor/status/712880710328139776
https://www.statnews.com/2020/12/21/stanford-covid19-vaccine-algorithm/
https://www.theguardian.com/education/2020/aug/20/england-exams-row-timeline-was-ofqual-warned-of-algorithm-bias
https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

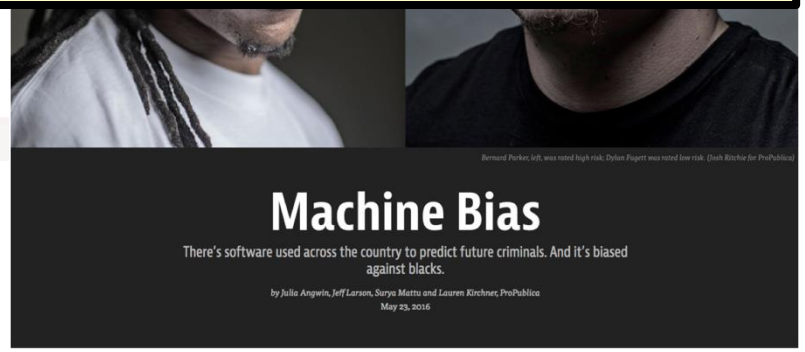
What this course is about

BUSINESS NEWS OCTOBER 9, 2018 / 11:12 PM / 2 MONTHS AGO

Amazon scraps secret AI recruiting tool that showed bias against women

Rigor, professionalism,
ethics, and inclusion are all essential!

By CASEY ROSS @caseymross and ERIN BRODWIN @erbrod / DECEMBER 21, 2020

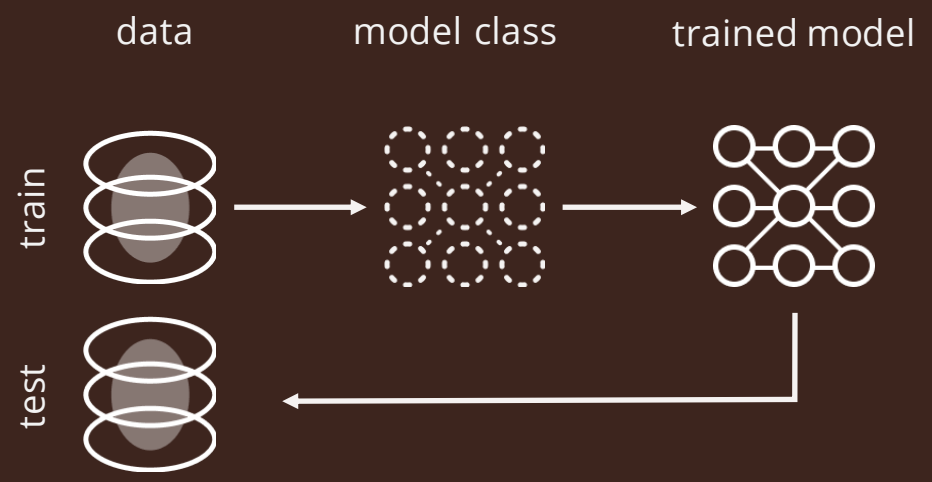


https://twitter.com/geraldmellor/status/712880710328139776
https://www.statnews.com/2020/12/21/stanford-covid19-vaccine-algorithm/
https://www.theguardian.com/education/2020/aug/20/england-exams-row-timeline-was-ofqual-warned-of-algorithm-bias
https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

What happens after CS 181?

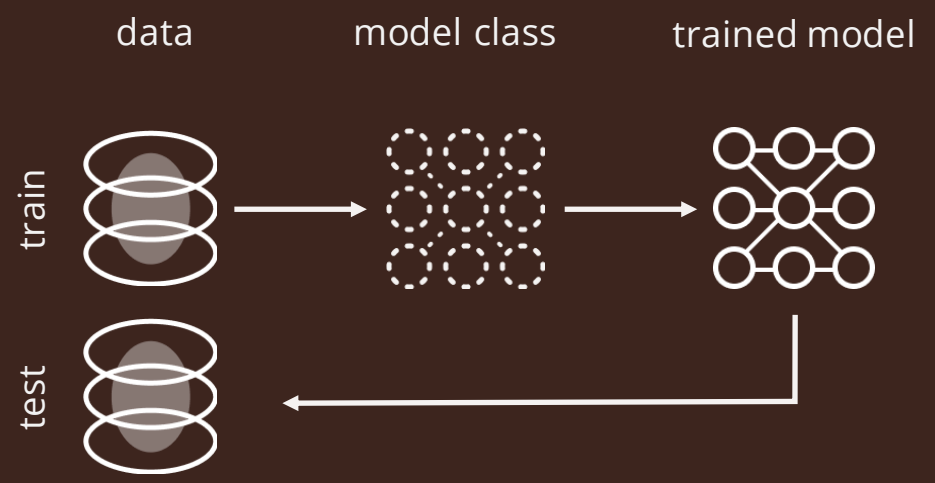
ML 101 in a nutshell

Given: task description, data, model class, loss objective

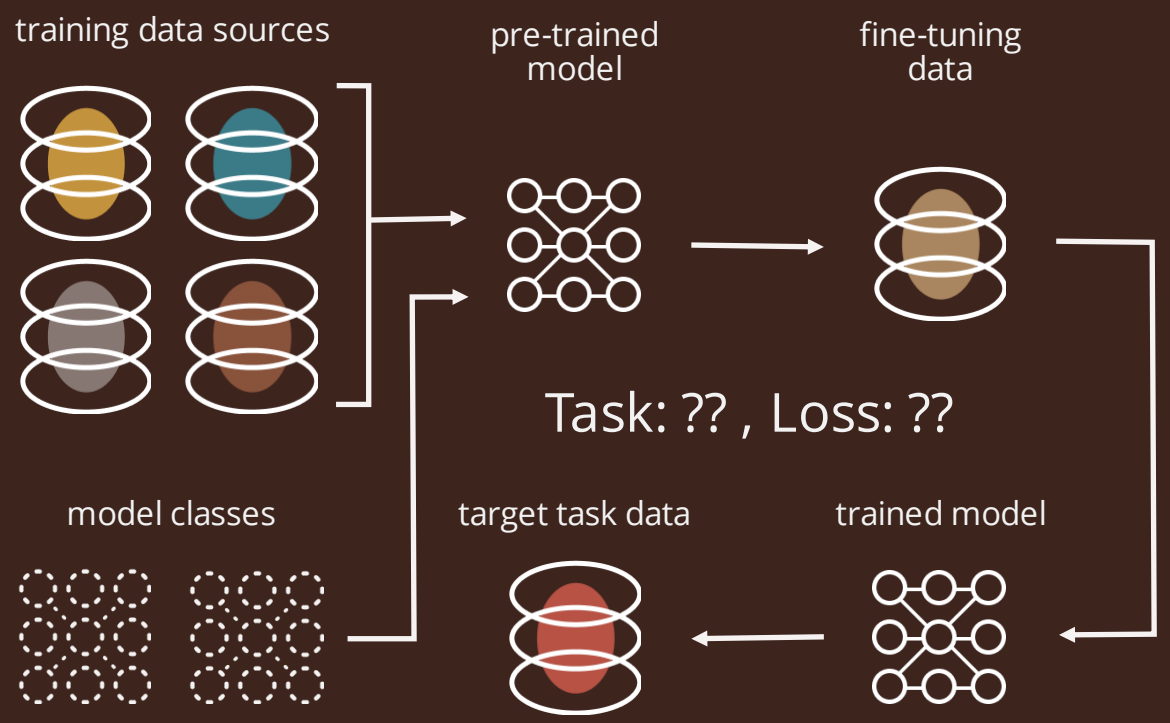


The Bad News

Given: task description, data, model class, loss objective



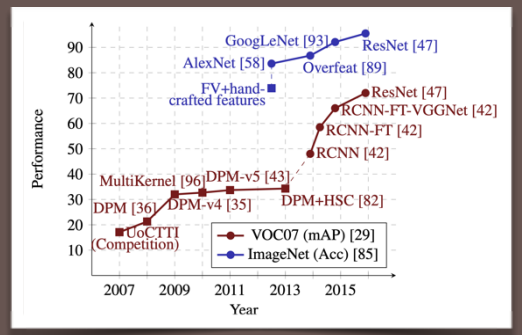
Machine Learning Courses



Machine Learning in the 'wild'

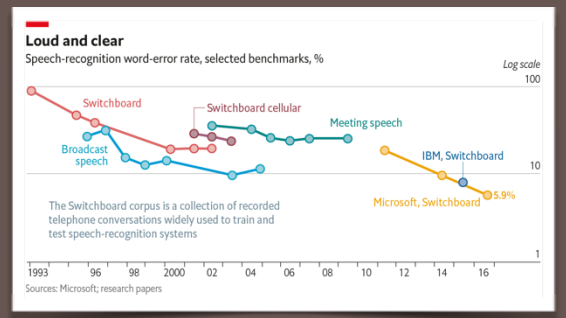
A Tale of Two Fields

Image Classification



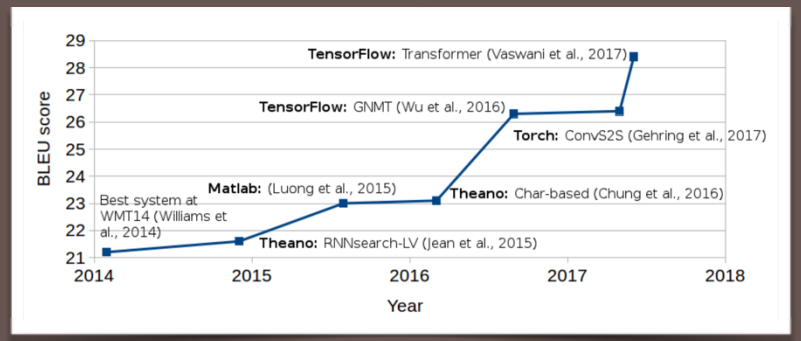
Credit: [Sharif Razavian, 2017]

Speech Recognition



Credit: The Economist

Machine Translation



Credit: [Stahlberg et al., 2018]

SONIA PAUL BACKCHANNEL 03.20.2017 12:00 AM

Voice Is the Next Big Platform, Unless You Have an Accent

...tand my mom — until we ...and do pretty much

Harvard Business Review Technology | The Dumb Reason Your AI Project Wi...

The Dumb Reason Your AI Project Will Fail

by Terence Ts...

June 08, 2020

nature

Explore our content · Journal information · Publish with us · Subscribe

NEWS FEATURE · 09 OCTOBER 2019

Why deep-learning AIs are so easy to fool

Artificial-intelligence researchers are trying to fix the flaws of neural networks.

Douglas Heaven

Future Finance

Who to Sue When a Robot Loses Your Fortune

The first known case of humans going to court over investment losses triggered by autonomous machines will test the limits of liability.

By Thoma...

'Benchmark' Machine Learning

Machine Learning in the 'wild'

The Good News

The same fundamental principles we will learn in this class can be re-used and extended to tackle "ML in the Wild" too:

e.g. probabilistic modeling, generalization, bias-variance tradeoffs, representation, capacity, overfitting, etc

'Benchmark' Machine Learning

Machine Learning in the 'wild'

My group's research

Bridging the gap between 'textbook ML' and 'real ML' requires* new methods & theory for:

manipulating multiple heterogeneous noisy data sources

understanding model (and data) failures

**among many other things*

'Benchmark' Machine Learning

Machine Learning in the 'wild'

Data-Centric ML:

Better datasets for better (reliable & trustworthy) models

Data "Learning"

How can we (mathematically) optimize the training data we feed intro our models?

Model ↔ Data Interpretability

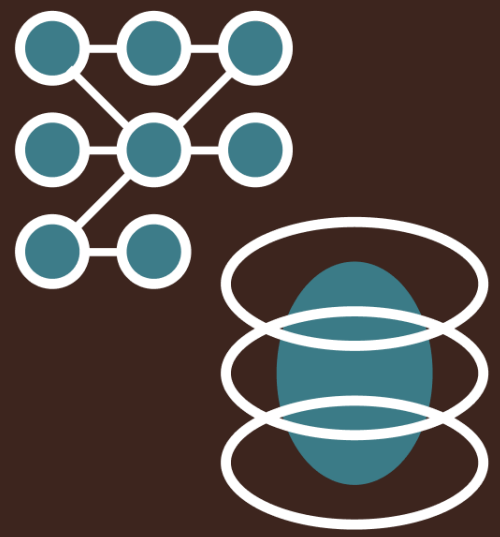
What can we learn about the inner workings of a model through the lens of how it processes data?

ML with Structured Data

How should we deal with data with rich geometric structure (e.g., Molecules, DNA sequences, graphs) ?

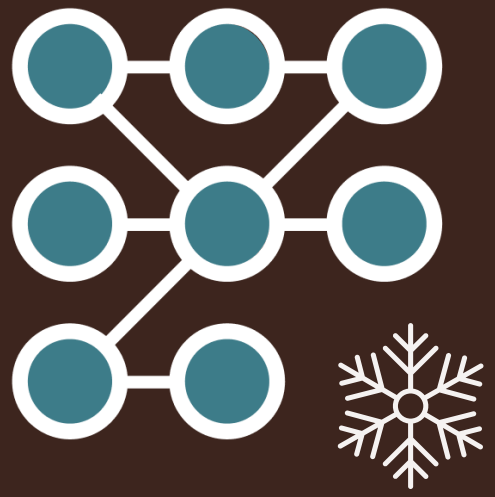
Data learning:

Adapting models by adapting datasets



Work partly done while at Microsoft Research in collaboration with:

Jiaojiao Fan, Nicolò Fusi, Jimmy Hall, Lester Mackey, Neil Tenenholtz

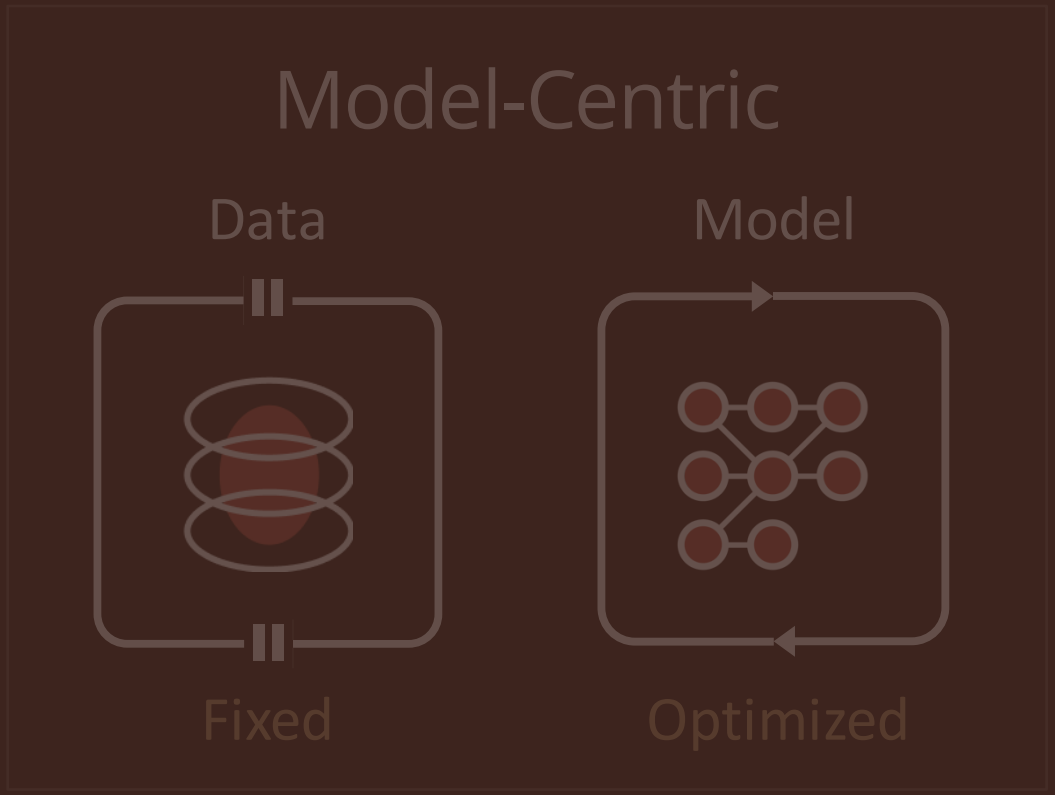


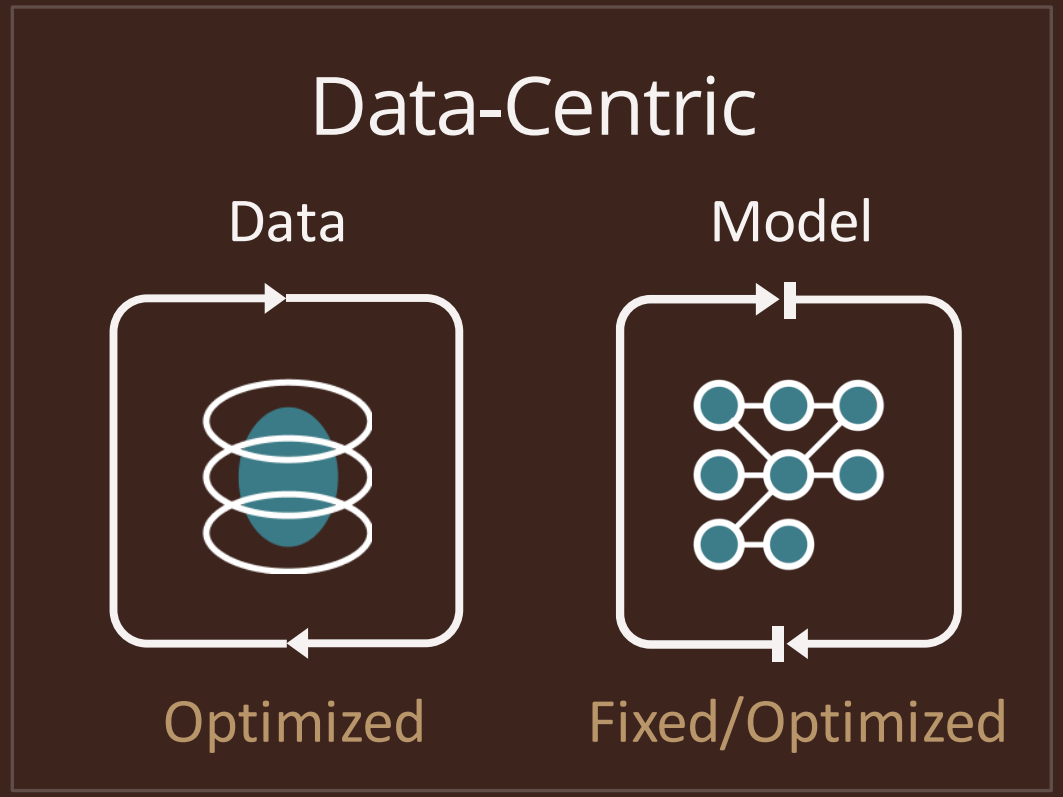
What does adaptation mean

in the era of

large pretrained models?

A new data-centric learning paradigm



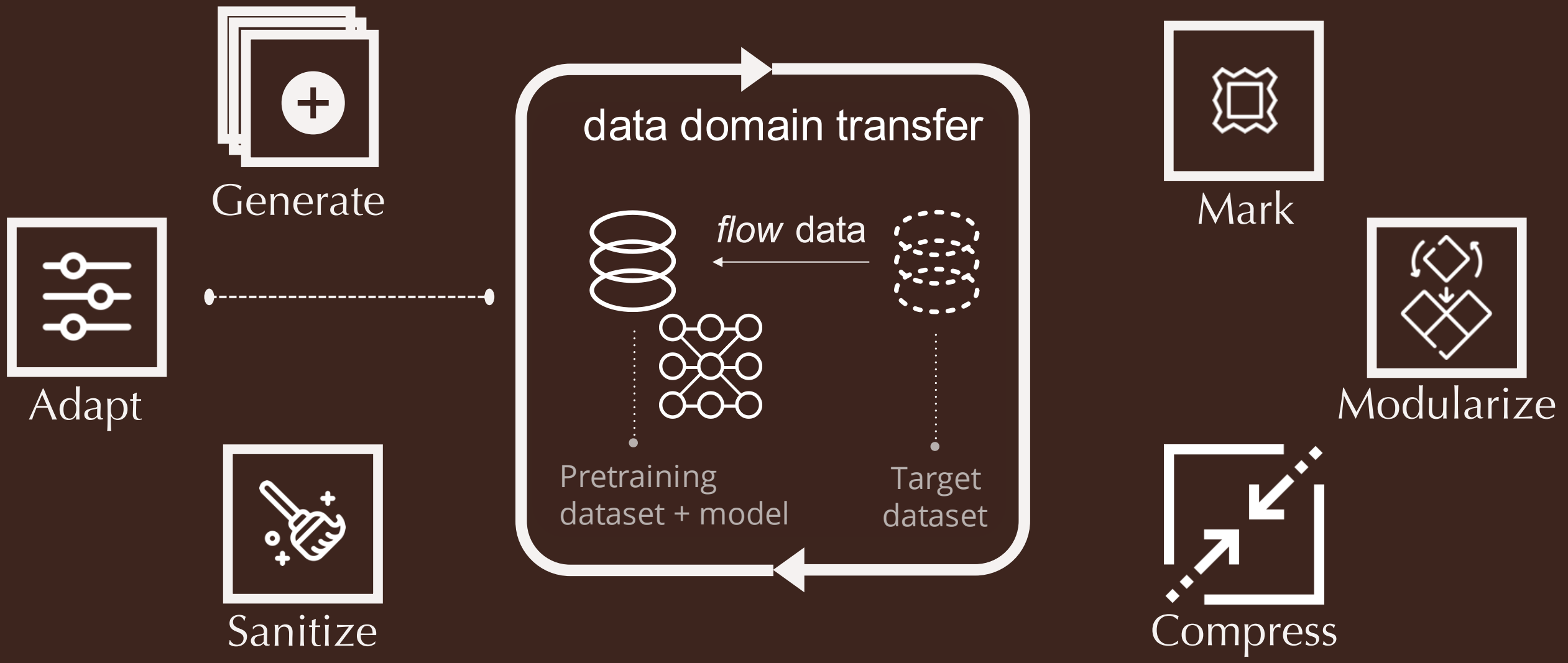


A new class of optimization problems:

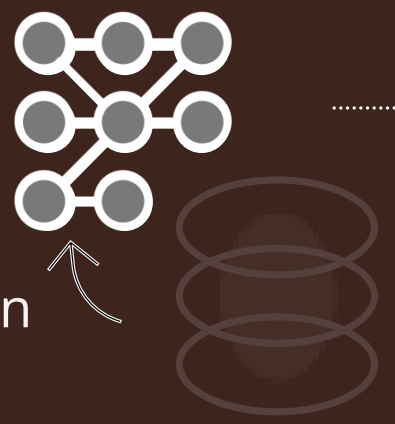
- a **dataset** is the optimization variable
- objective defined on entire datasets

Theoretical backbone: Optimal Transport

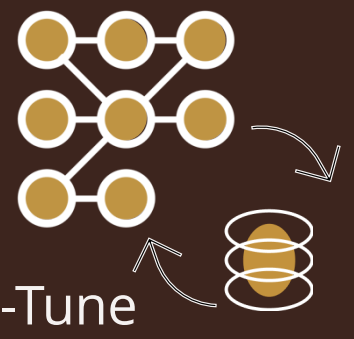
Data Learning Tasks



How are large pre-trained models used in practice?



Status-quo: adapt model, data is fixed



This is often inefficient or altogether impossible (e.g., LLM via API access)

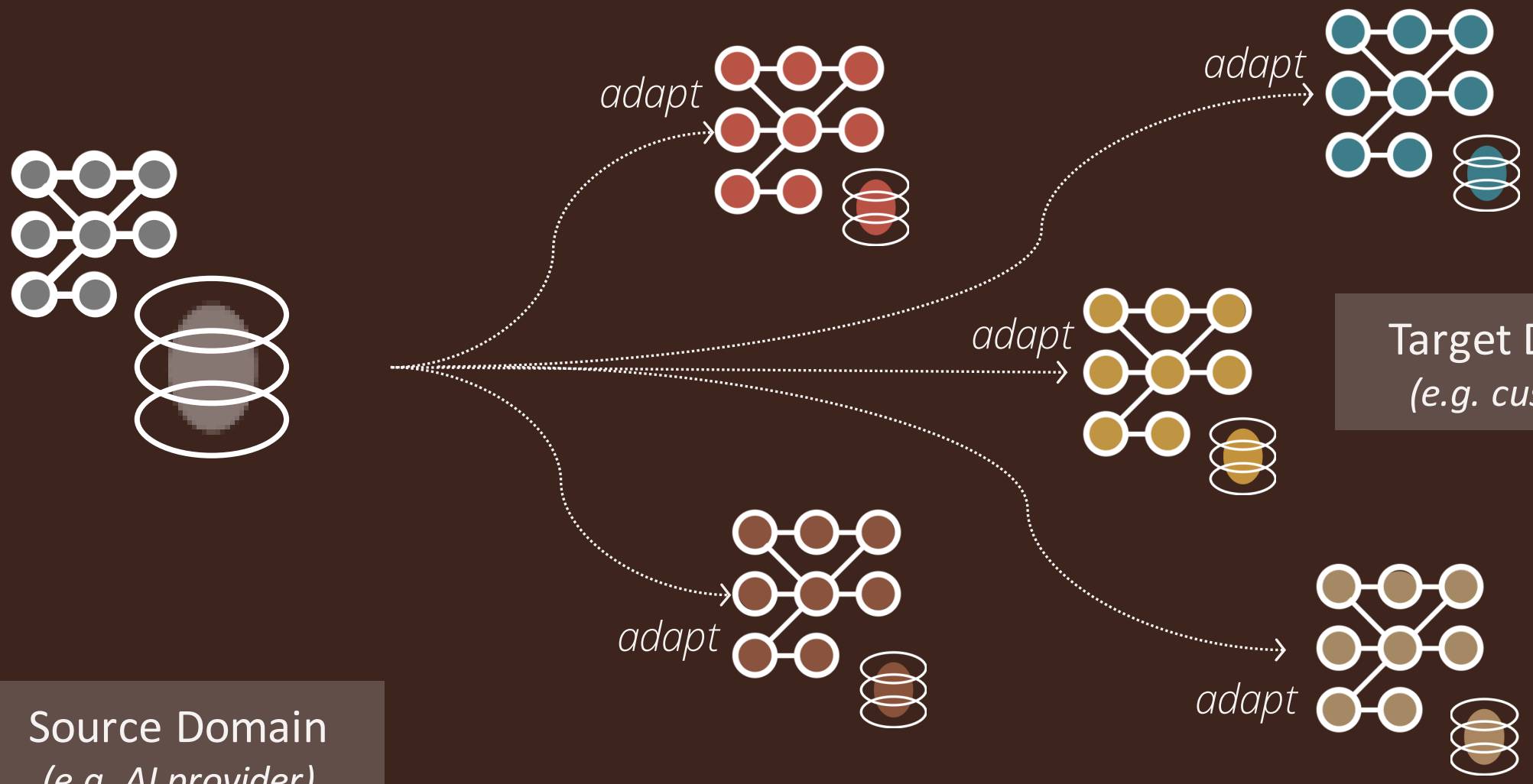
pre-trained model
(on general-purpose data)

fine-tuned model
(on limited specialized data)

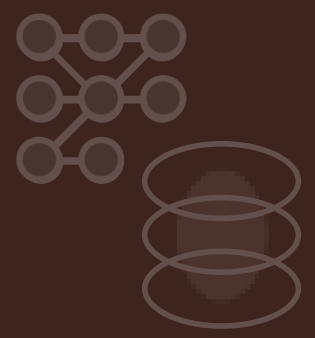
Source Domain

Target Domain

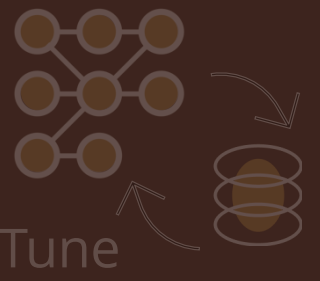
Large model adaptation doesn't scale well



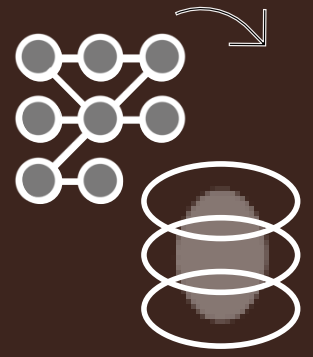
Efficient adaptation by modifying data



Model-centric: adapt model, data is fixed



Predict



Data-Centric: **adapt data**, model is fixed

Pre-Training Domain
(general-purpose data)

Adapting the data allows us to impose additional constraints (privacy / size / etc.)!

Target Domain
(limited specialized data)

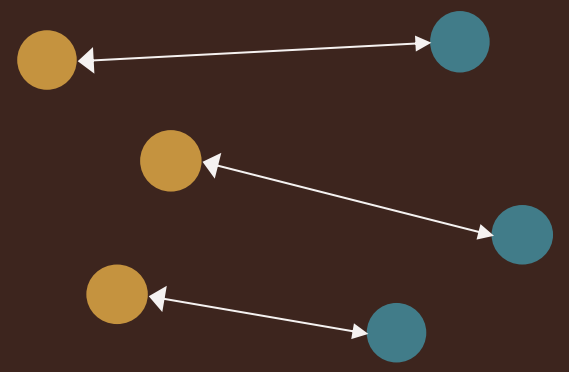
Our Main Tool: Optimal Transport

- A sub-field at the intersection of optimization, statistics and geometry.

Compare

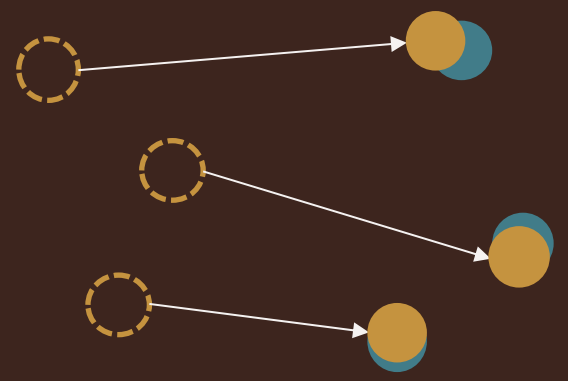
$d(P_A, P_B)$

Match



$\pi(x_i, y_j)$

Transform



$\mathrm{T}(x_i) \approx y_j$

- Centuries worth of theory: Monge (1780s), Kantorovich (1940s), Villani, Figalli, etc.
- Fast algorithms (~2014 – now).

Transforming Datasets via Gradient Flows

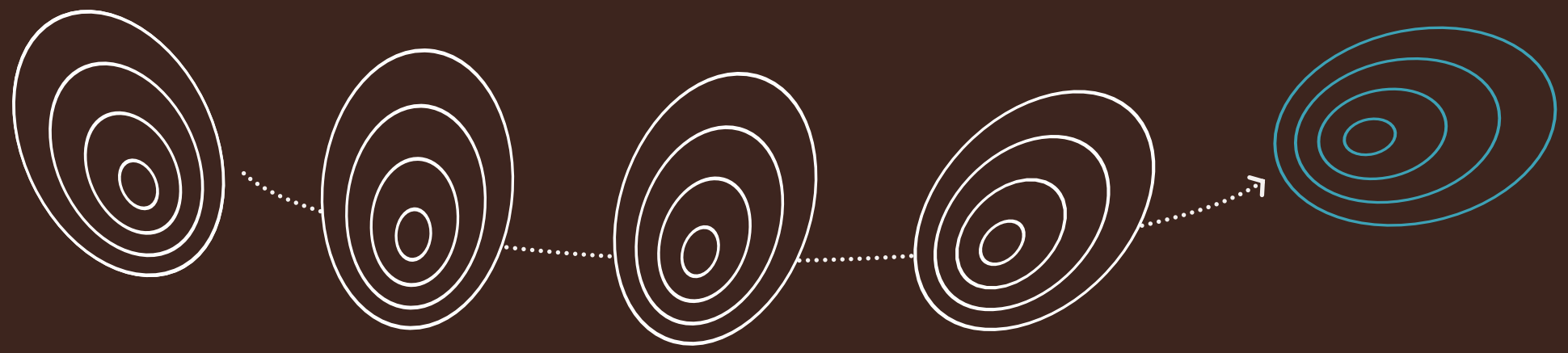
Distance to target dataset* ↑

gradient flow using optimal transport ↑

Initial dataset + objective + constraints ⟶ Learned dataset

↓ *e.g. pretraining data*

↓ *Size? Privacy? Invariances?*

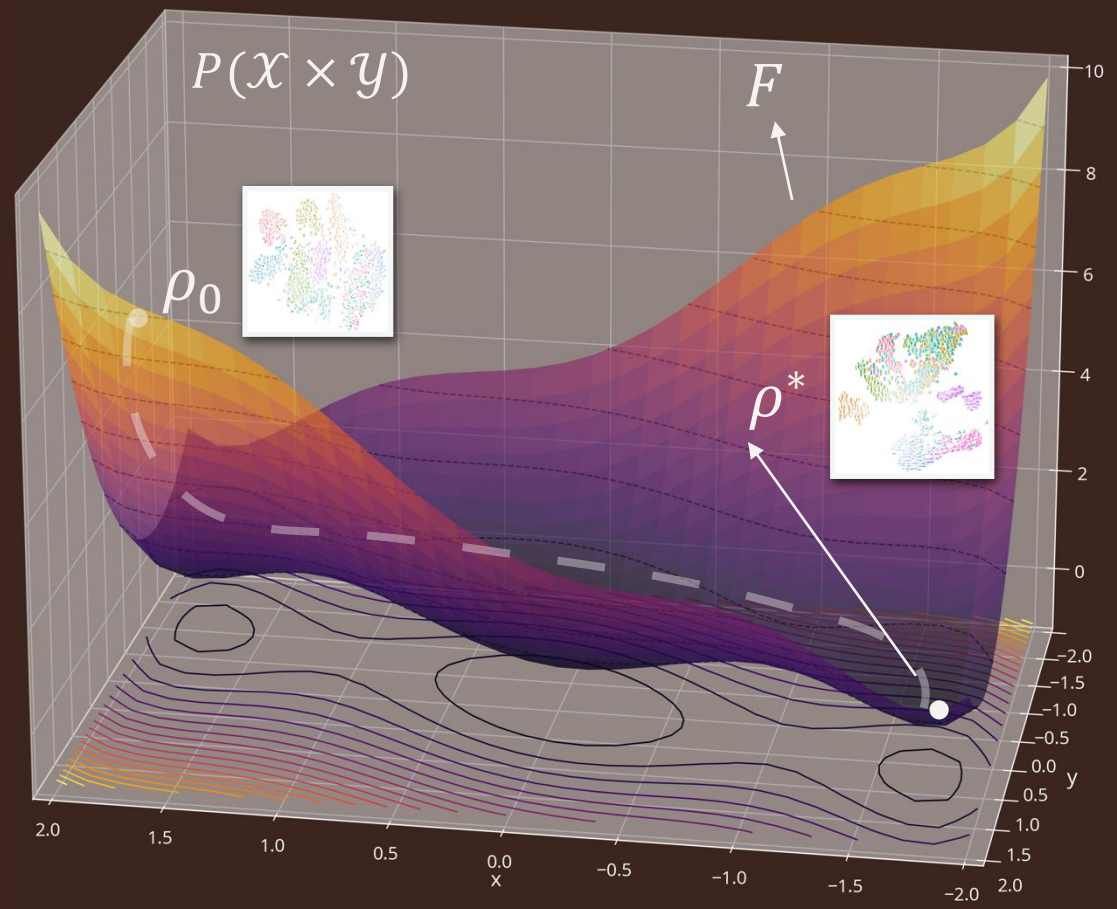


*Optimal Transport Dataset Distance (OTDD) [NeurIPS 2020]

[A-M & Fusi, ICML 2021]

Gradient Flows in Probability Space

Need to solve optimization problem over (infinite dimensional, non-euclidean!) probability space



Gradient flow in probability space:
[Ambrosio; Santambrogio; Figalli; Villani; etc]

$$\partial_t \rho(t) = -\nabla \cdot \left(\rho(t) \nabla \frac{\delta F}{\delta \rho}(\rho(t)) \right), \rho(0) = \rho_0$$

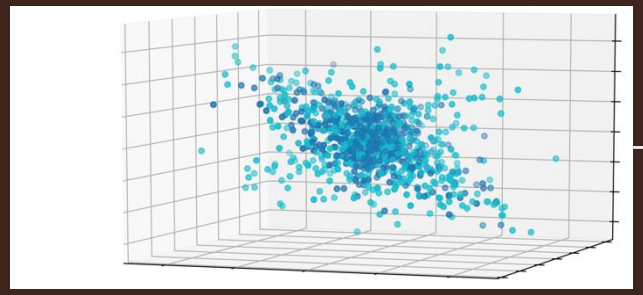
finite particle approximation

time discretization

$$z_{t+1}^{(i)} \leftarrow z_t^{(i)} - \lambda \nabla_{z^{(i)}} F(\widehat{\rho_t}) \quad i = 1, \dots, n$$

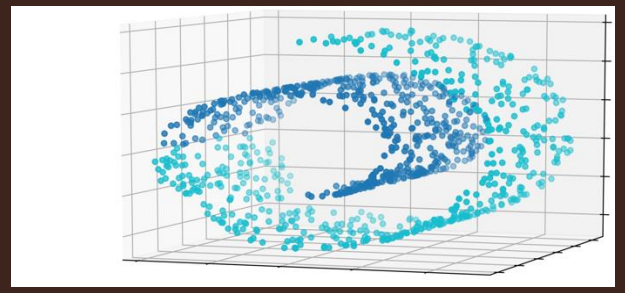
Application: Dataset Shaping

Initial Dataset ρ_0



Want: similar to ρ^* but linearly separable

Reference Dataset ρ^*



Functional Objective $F(\rho)$

$\mathrm{OTDD}(\rho, \rho^*)$

(distance only)

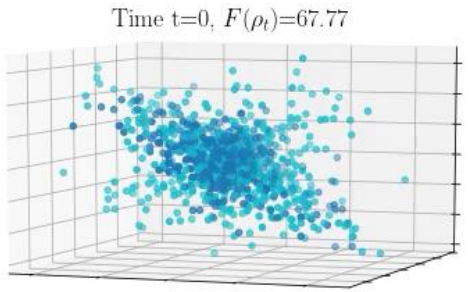
$\mathcal{V}(\rho)$

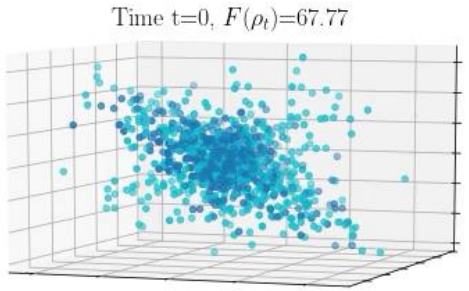
(constraint only)

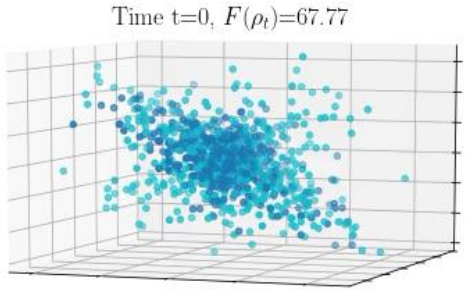
$\mathrm{OTDD}(\rho, \rho^*) + \mathcal{V}(\rho)$

(distance + constraint)

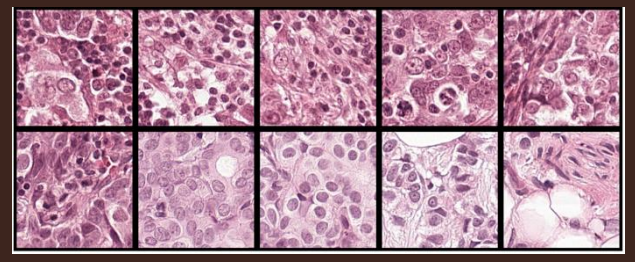
Optimization Iterates ρ_t







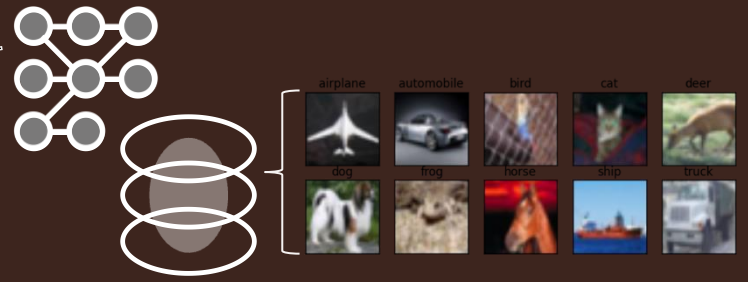
Flowing data to repurpose models



Target Dataset: histopathology

'Flow' towards CIFAR10

Predict

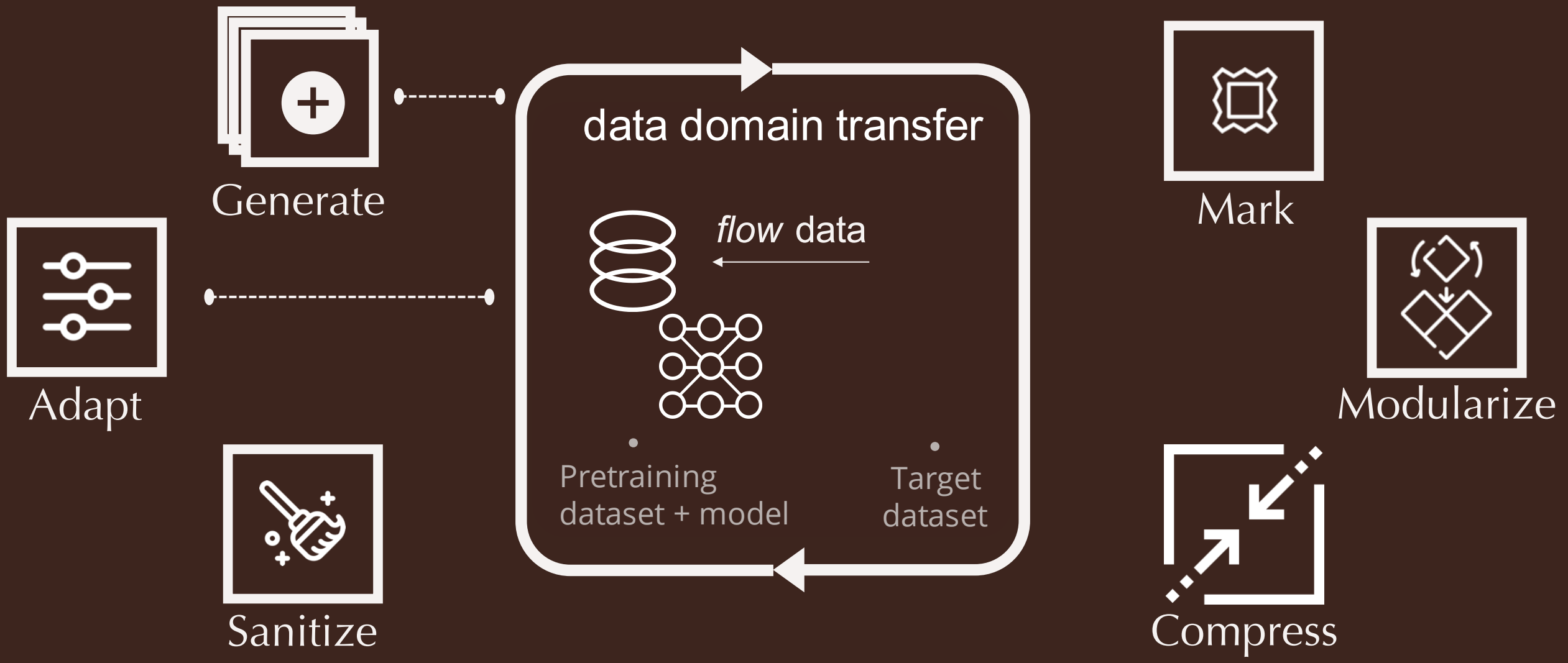


Model: NN trained on CIFAR 10
(frozen after pre-training)

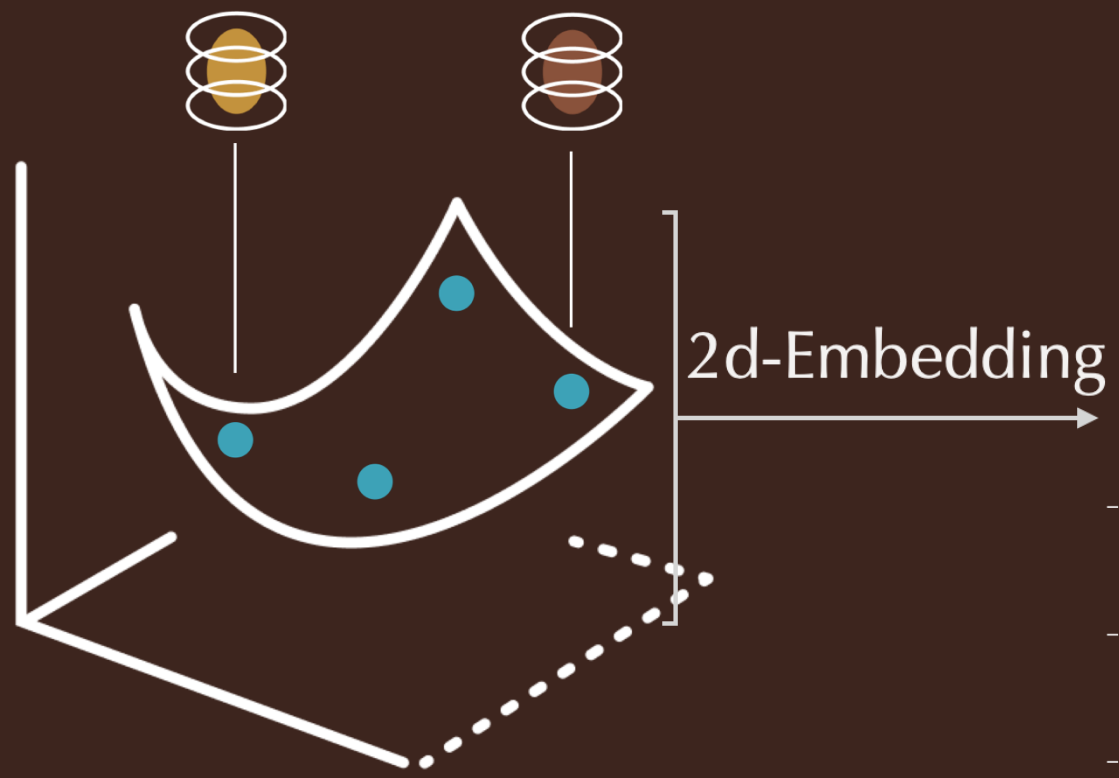
Key Takeaway:
models trained only on histopathology images can be repurposed without modifying them!

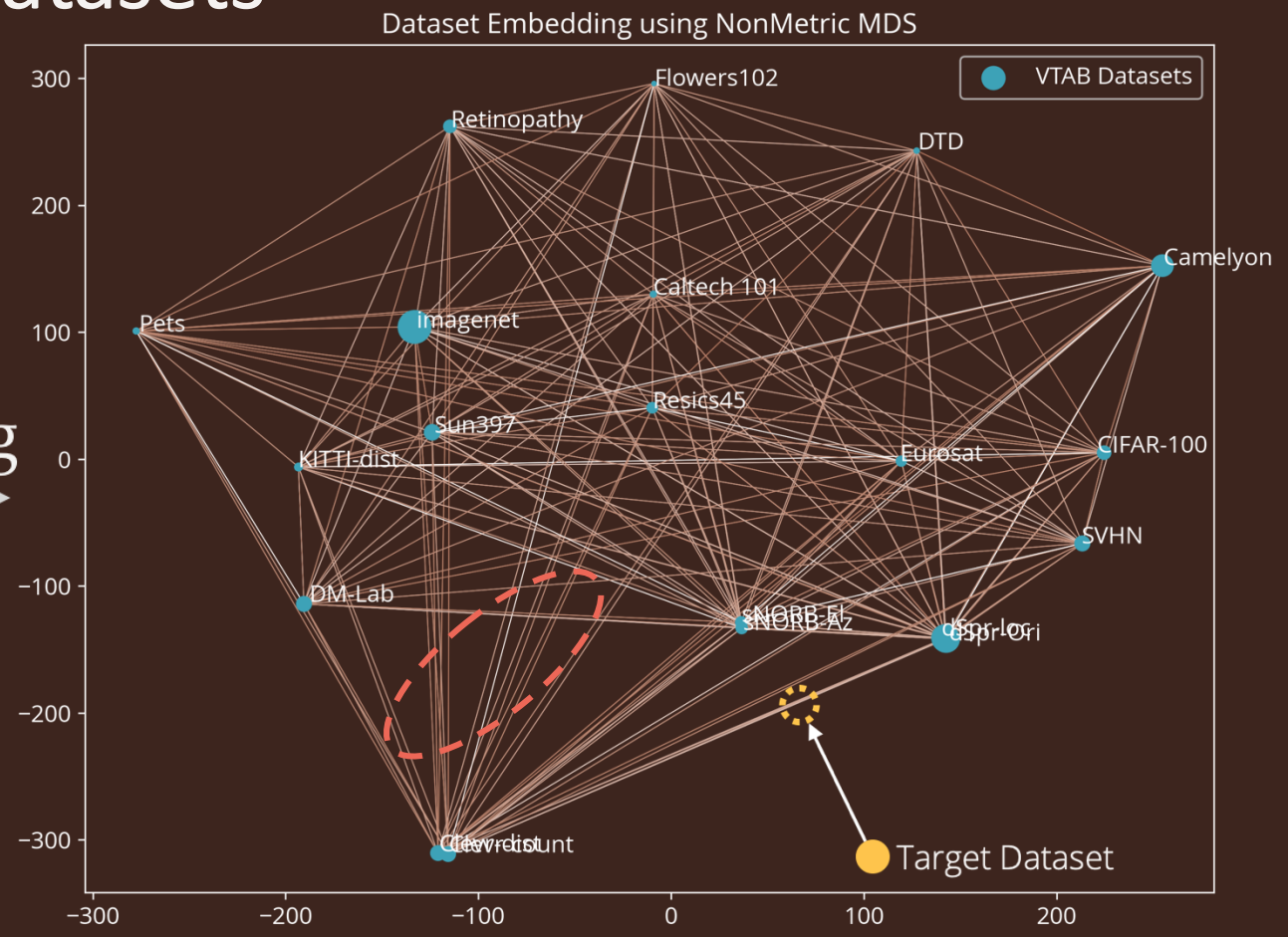
→ random guessing baseline

Data Learning Tasks



Generation in the Space of Datasets





Space of Datasets
(general-purpose data)

Can we 'fill gaps' in this manifold to generalize better?

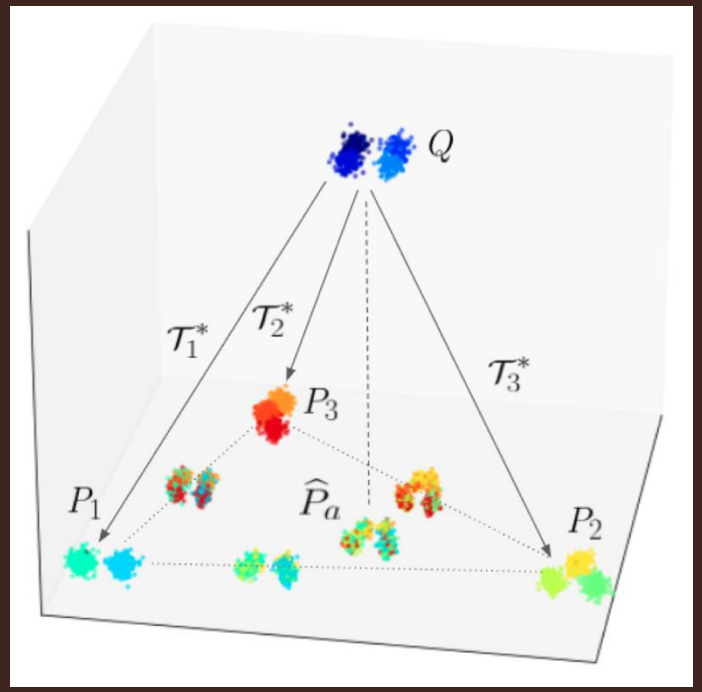
or 'project' into manifold to create closest training dataset?

Adaptation by Synthesis

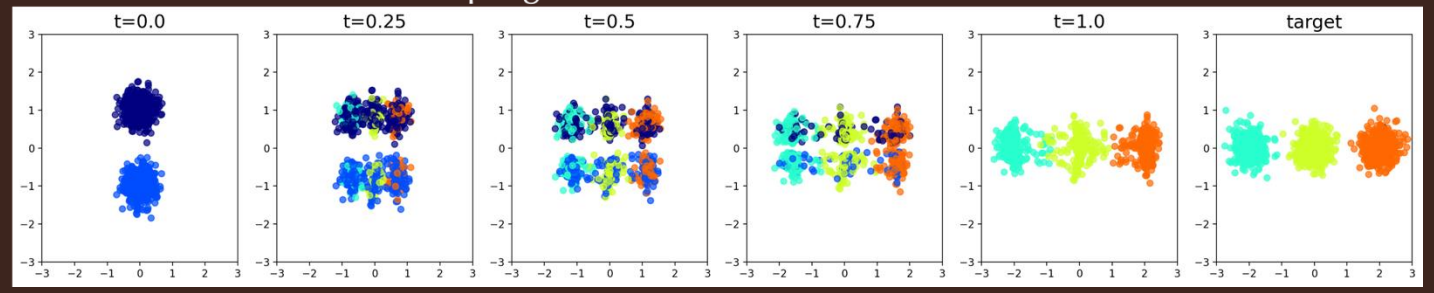
Pre-training domains: $D_i \sim P_i$. Target domain $D_0 \sim Q$.

Goal: generate the dataset in 'convex hull' of P_i closest to Q.

Formalized as multi-dataset interpolations, using generalized geodesics from OT



Example geodesic between two datasets



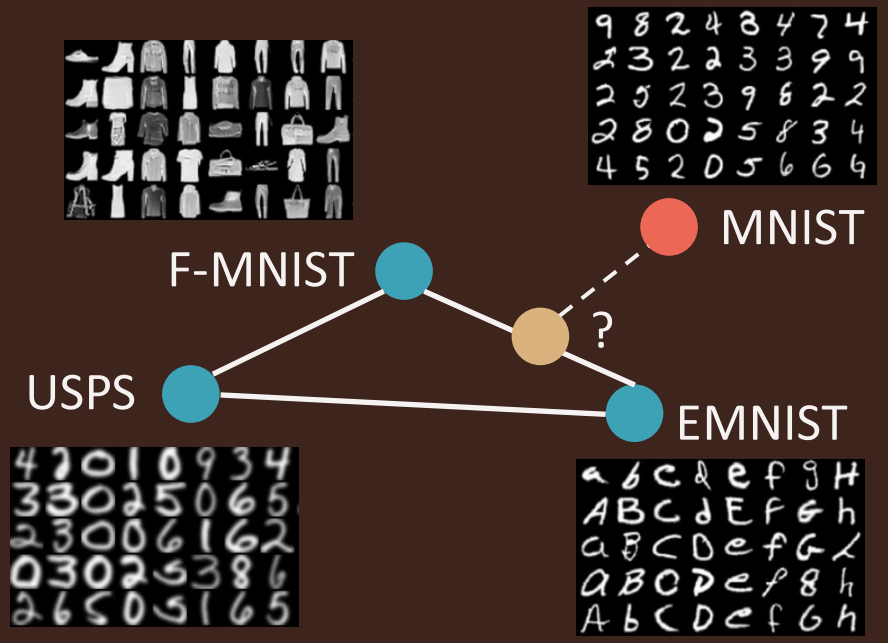
Want mixture weights a minimizing $\mathrm{Dist}(P_a, Q)$.

Can't be found directly, instead we solve:

$$a = \underset{u \in \mathrm{cvx}(X)}{\mathrm{argmin}}\, \mathrm{W}_{2,Q}(P_a, Q) = \sum_{i=1}^{m} a_i \mathrm{W}_{2,Q}^2(P_i, Q) - \frac{1}{2} \sum_{i \neq j}^{m} a_i \mathrm{W}_{2,Q}^2(P_i, P_j)$$

[A-M & Fan, UAI 2023]

Adaptation by Synthesis



Pre-training on synth dataset closest to target + adapting:

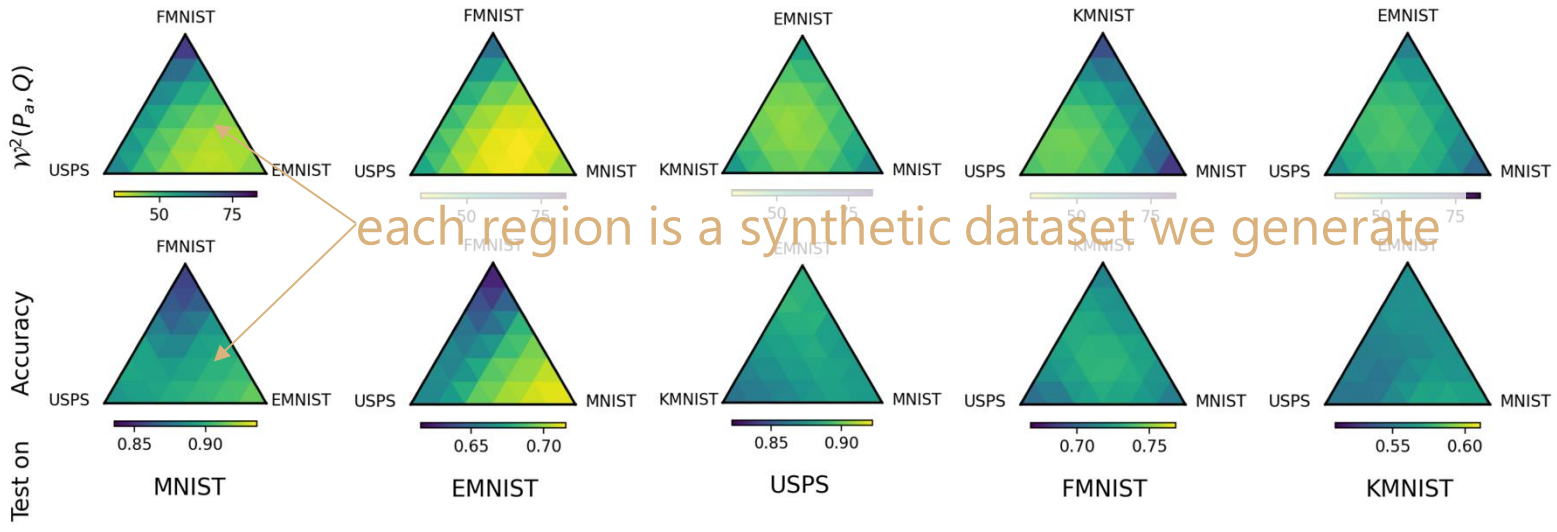
Table 1: **Pretraining on synthetic data**. Shown is 5-shot transfer accuracy (mean ± s.d. over 5 runs).

Methods	MNIST-M	MNIST	USPS	FMNIST	KMNIST	EMNIST
OTDD barycentric projection	**42.10±4.37**	**93.74±1.46**	86.01±1.50	**70.12±3.02**	**52.55±2.73**	**67.06±2.55**
OTDD neural map	40.06±4.75	88.78±3.85	83.80±1.60	70.02±2.59	50.32±3.10	65.32±1.80
Mixup	33.85±2.22	88.68±1.57	**88.61±2.00**	66.74±3.79	48.16±3.38	60.95±1.38
Train on few-shot dataset	19.10±3.57	72.80±3.10	80.73±2.07	60.50±3.07	41.67±2.11	53.60±1.18
1-NN on few-shot dataset	20.95±1.39	64.50±3.32	73.64±2.35	60.92±2.42	40.18±3.09	39.70±0.57

larger scale results (e.g. ImageNet⟺VTAB) on paper

Dataset Distance
(between synth and target datasets)

Transfer Accuracy
(pretrain on synth, adapt to target)



Take-Home Messages

'Usual' Adaptation: Focuses exclusively on model

Data-Centric Way: Data informs (or replaces!) model adaptation

- *Predict when and where to adapt using data*
- *Adapt the model implicitly by 'flowing' data*
- *Synthesize new datasets optimized for adaptation*

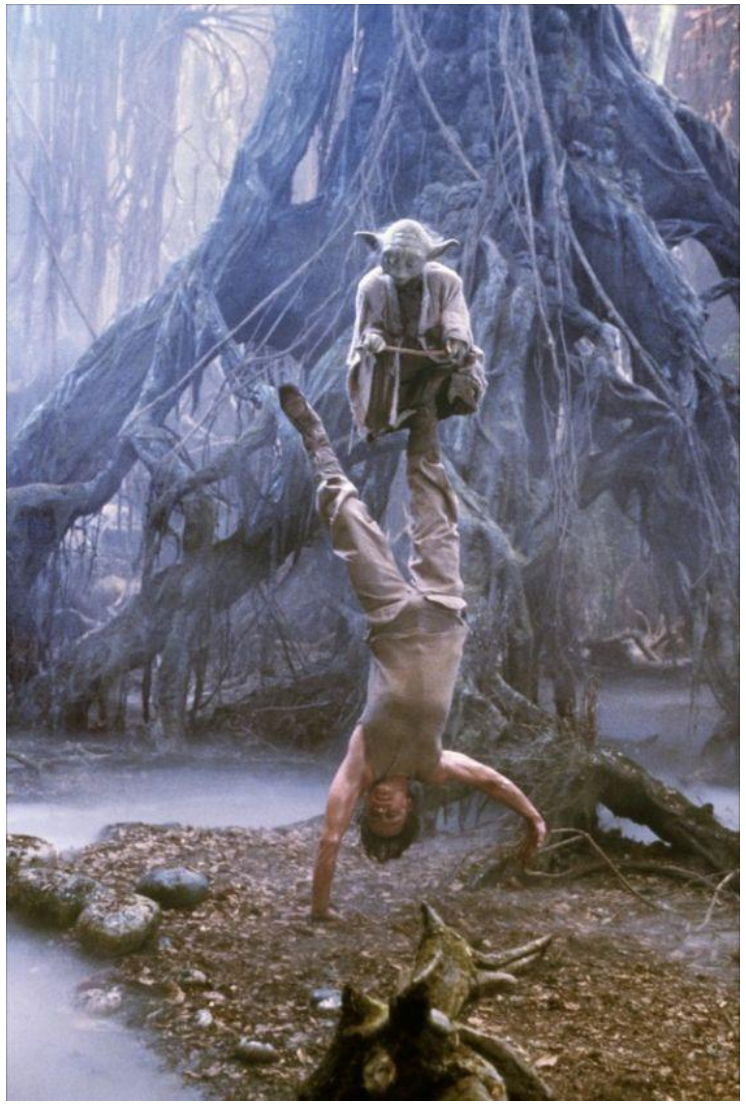
How we will get there: Zen and Hard Work

Zen

- Make appropriate model choices
- Understand the principles such that you can use new techniques
- Identify sources of error, in the maths and in the sociotechnical context
- Evaluate carefully

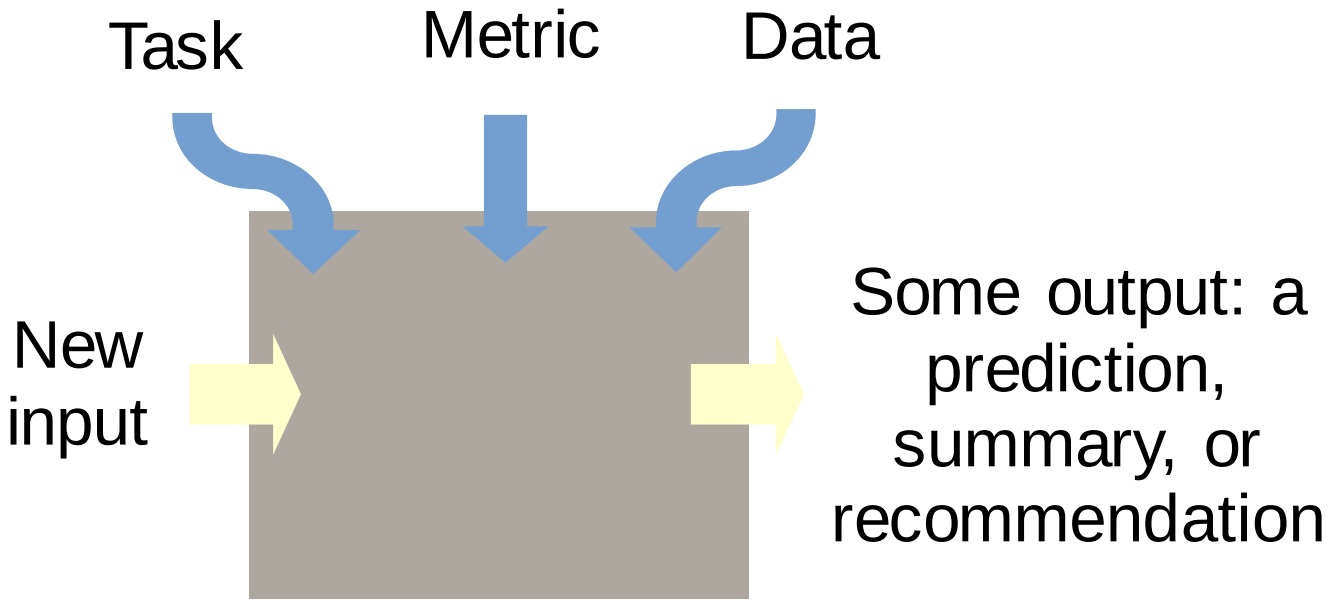
Zen

- Make appropriate model choices
- Understand the principles such that you can use new techniques
- Identify sources of error, in the maths and in the sociotechnical context
- Evaluate carefully



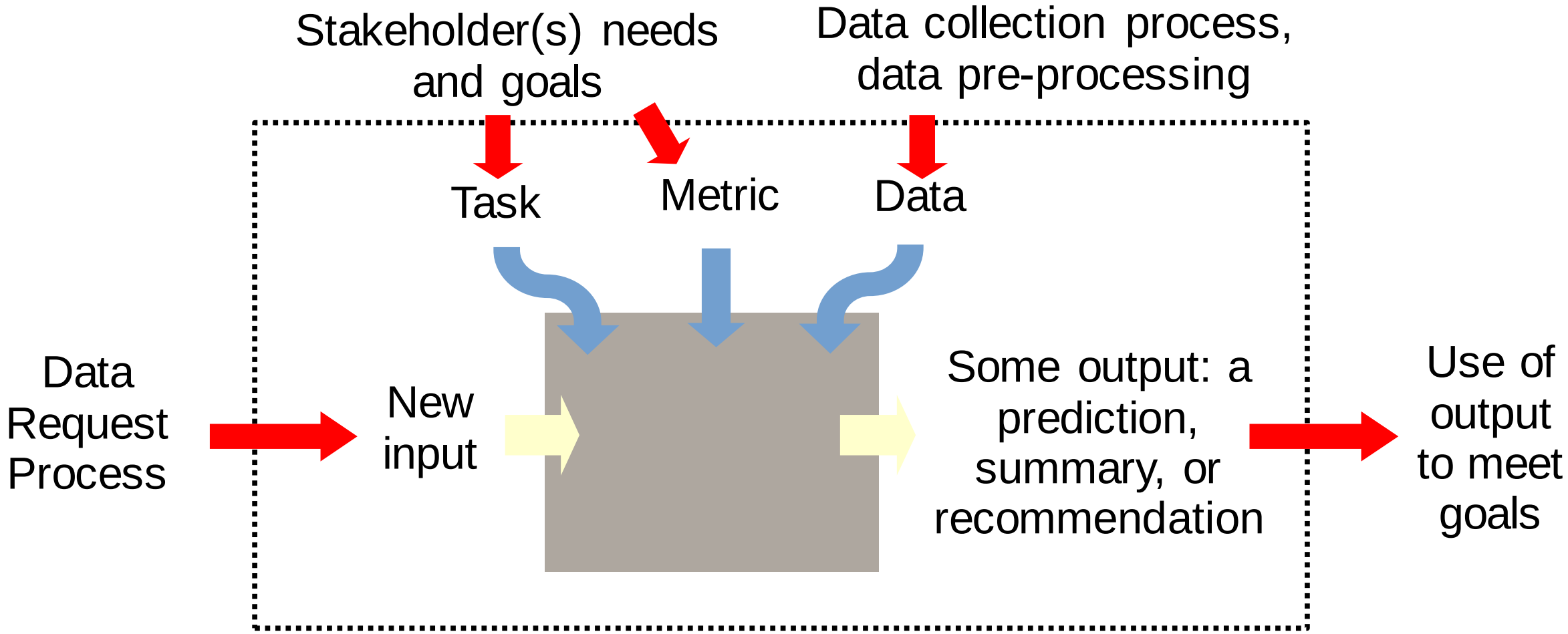
How we will get there: ML Taxonomy, Course Structure

High-Level Setup



Professional ethics requires that you get that box right.

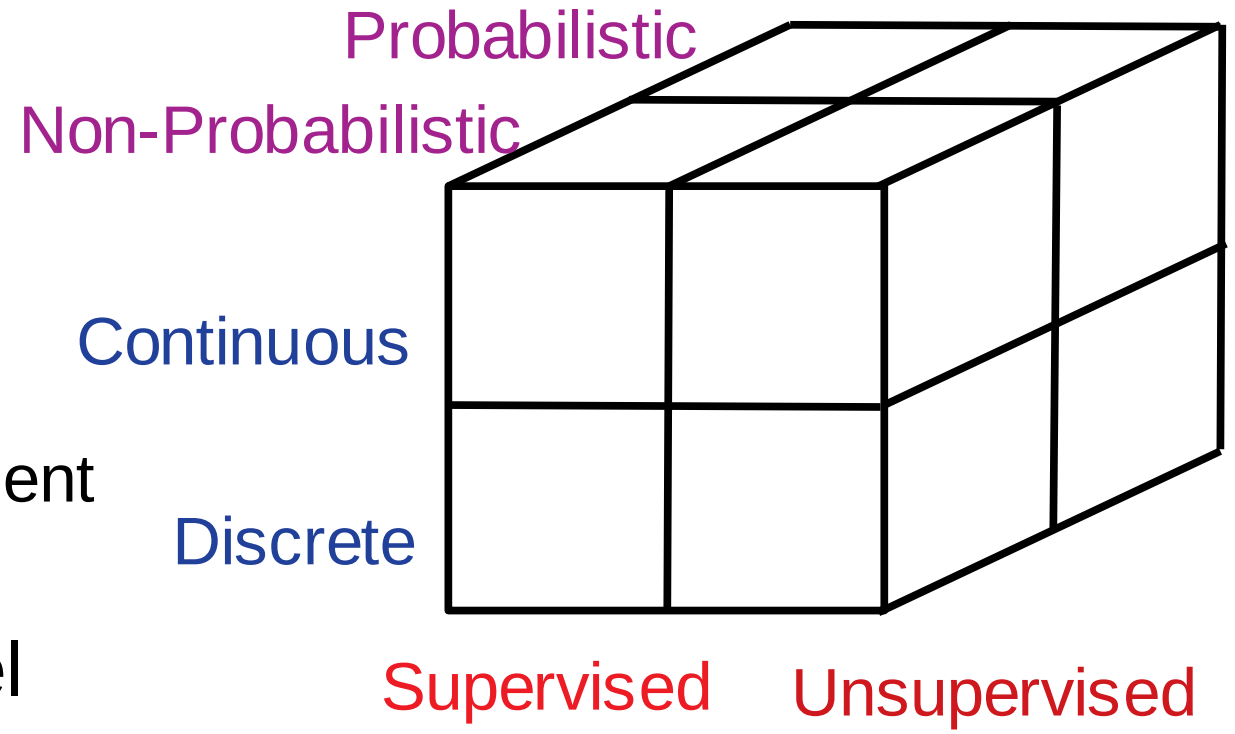
High-Level Setup



Societal ethics demands that you do not ignore the larger picture.

Structure of the Course

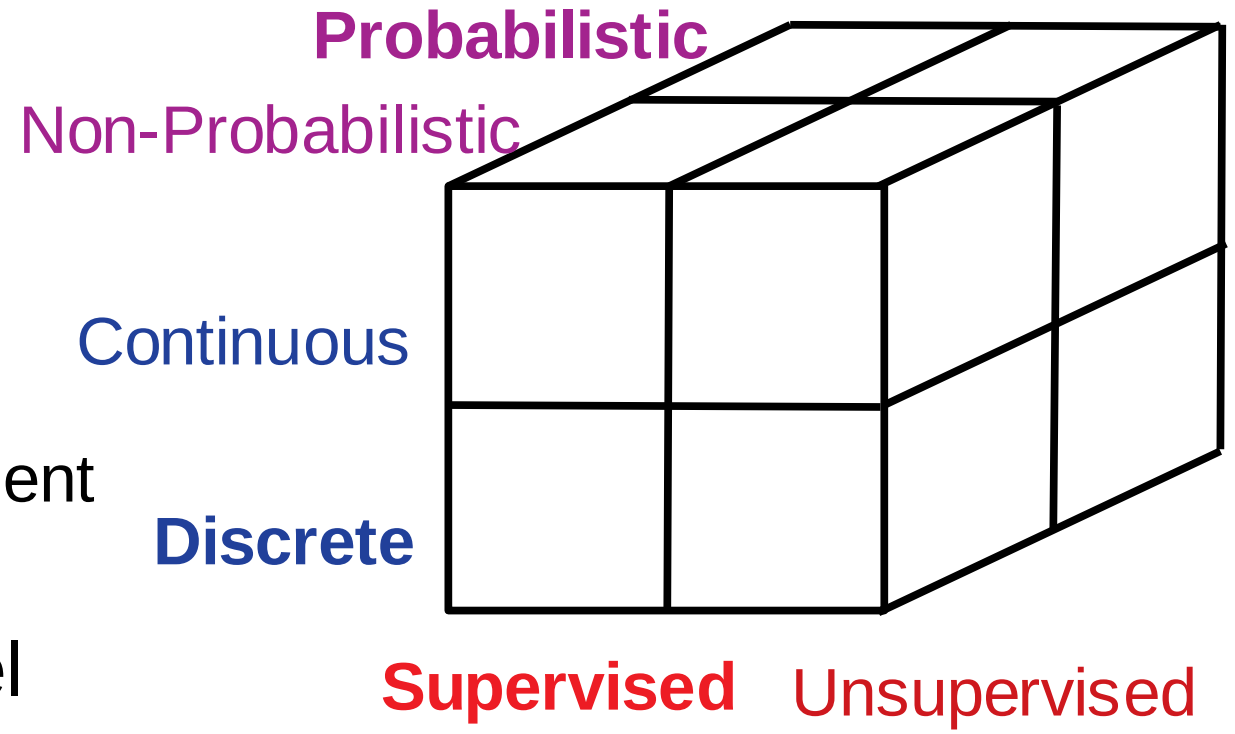
- Follow the “cube”++
  - Supervised: regression, classification
  - Unsupervised: clustering, embeddings
  - Graphical Models, Reinforcement Learning
- Excursions to discuss model classes, model selection, evaluation...



+ graphical models, reinforcement learning

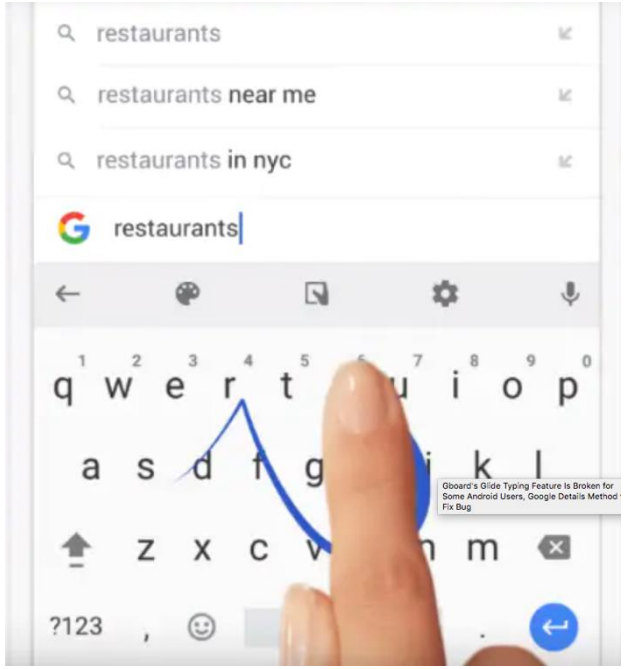
Structure of the Course

- Follow the “cube”++
  - Supervised: regression, classification
  - Unsupervised: clustering, embeddings
  - Graphical Models, Reinforcement Learning
- Excursions to discuss model classes, model selection, evaluation...

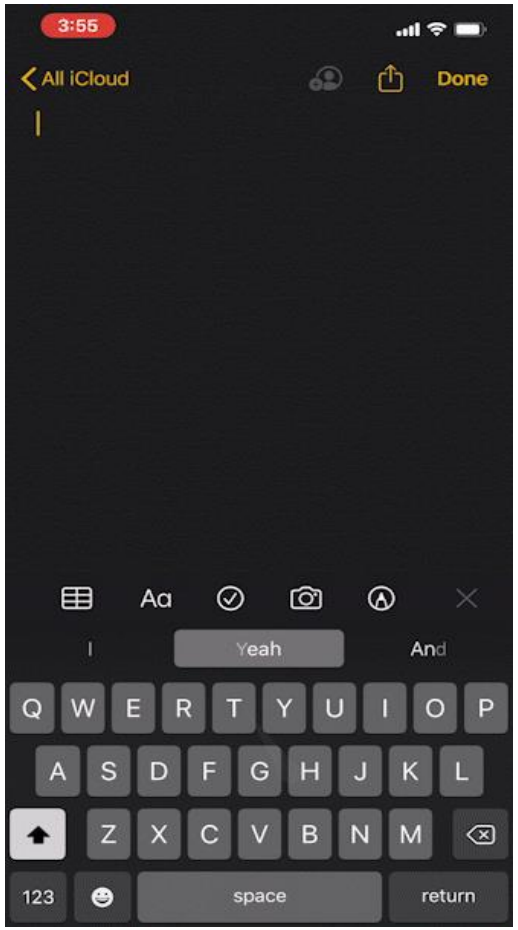


+ graphical models, reinforcement learning

Example: “Swipe typing”



Android Glide Typing



iOS QuickPath

Supervised:
Output is a
Prediction

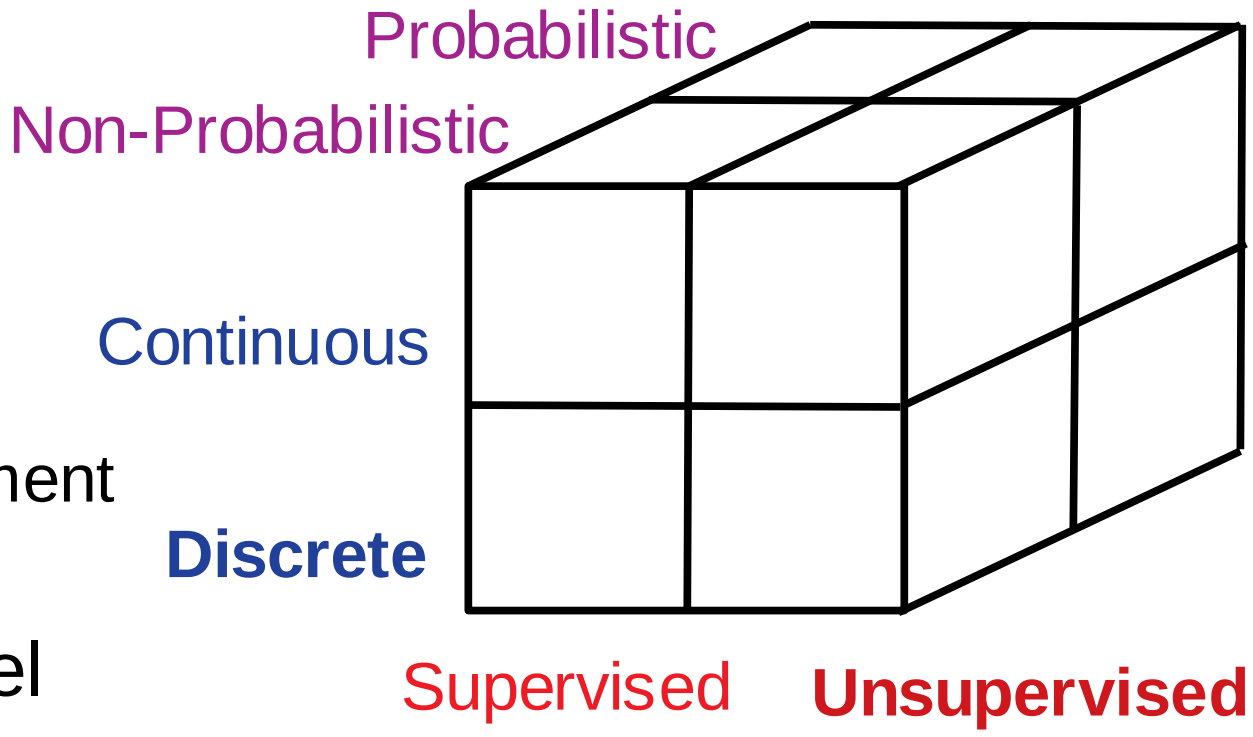
Discrete:
Output is a
Word

Probabilistic:
Works by
building a
language model

https://www.macworld.com/article/3432168/how-to-use-the-quickpath-swipe-keyboard.html
https://en.wikipedia.org/wiki/Swype
https://gadgets.ndtv.com/apps/news/gboard-for-android-glide-typing-bug-fix-google-1917639

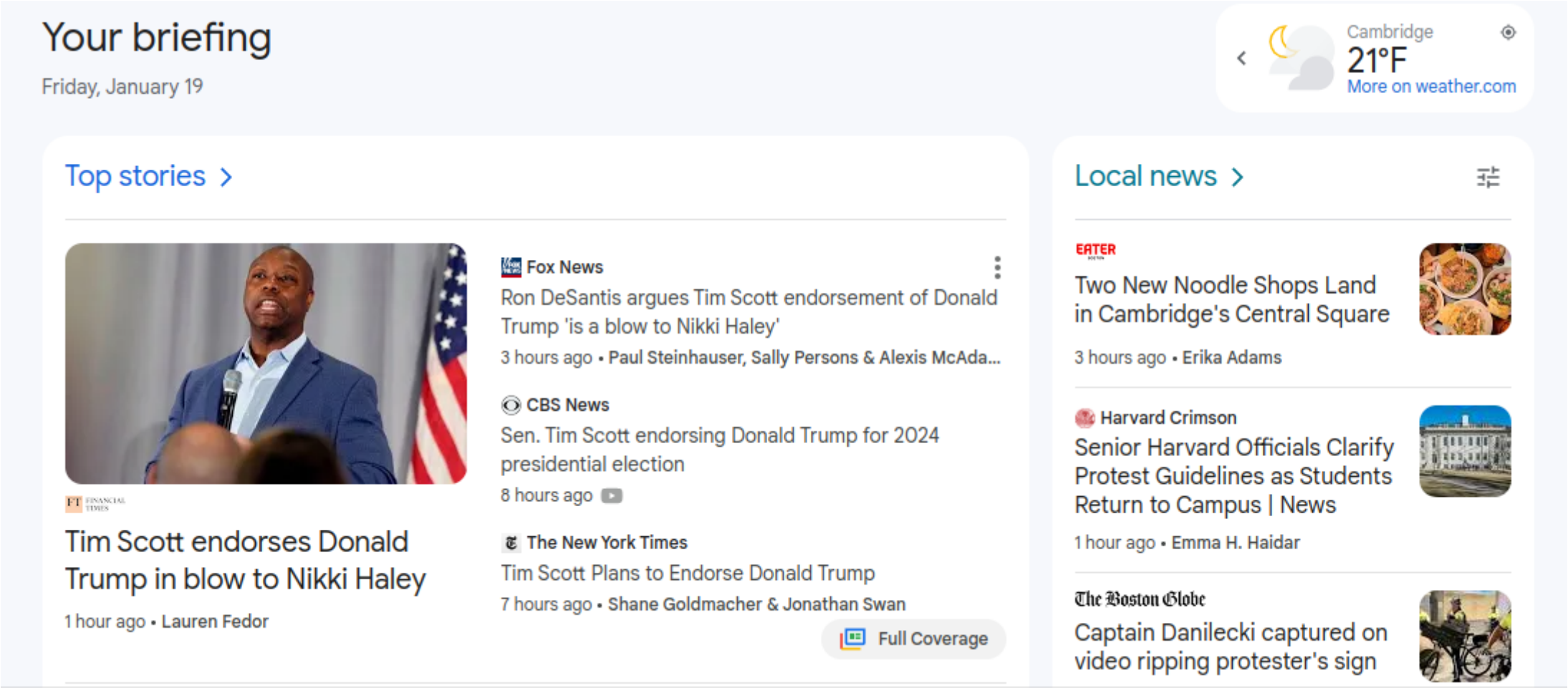
Structure of the Course

- Follow the “cube”++
  - Supervised: regression, classification
  - Unsupervised: clustering, embeddings
  - Graphical Models, Reinforcement Learning
- Excursions to discuss model classes, model selection, evaluation...



+ graphical models, reinforcement learning

Example: Google News

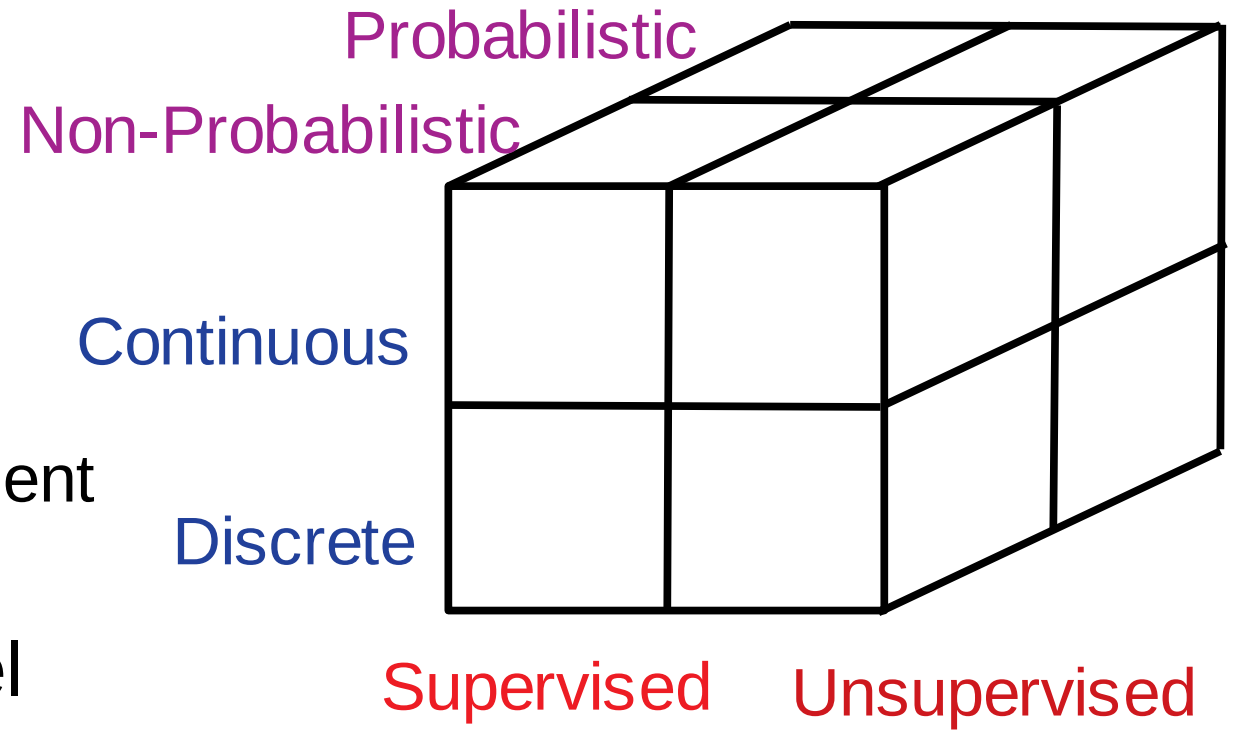


Unsupervised: Output is a **Summary**

Discrete: Outputs are **Clusttered**

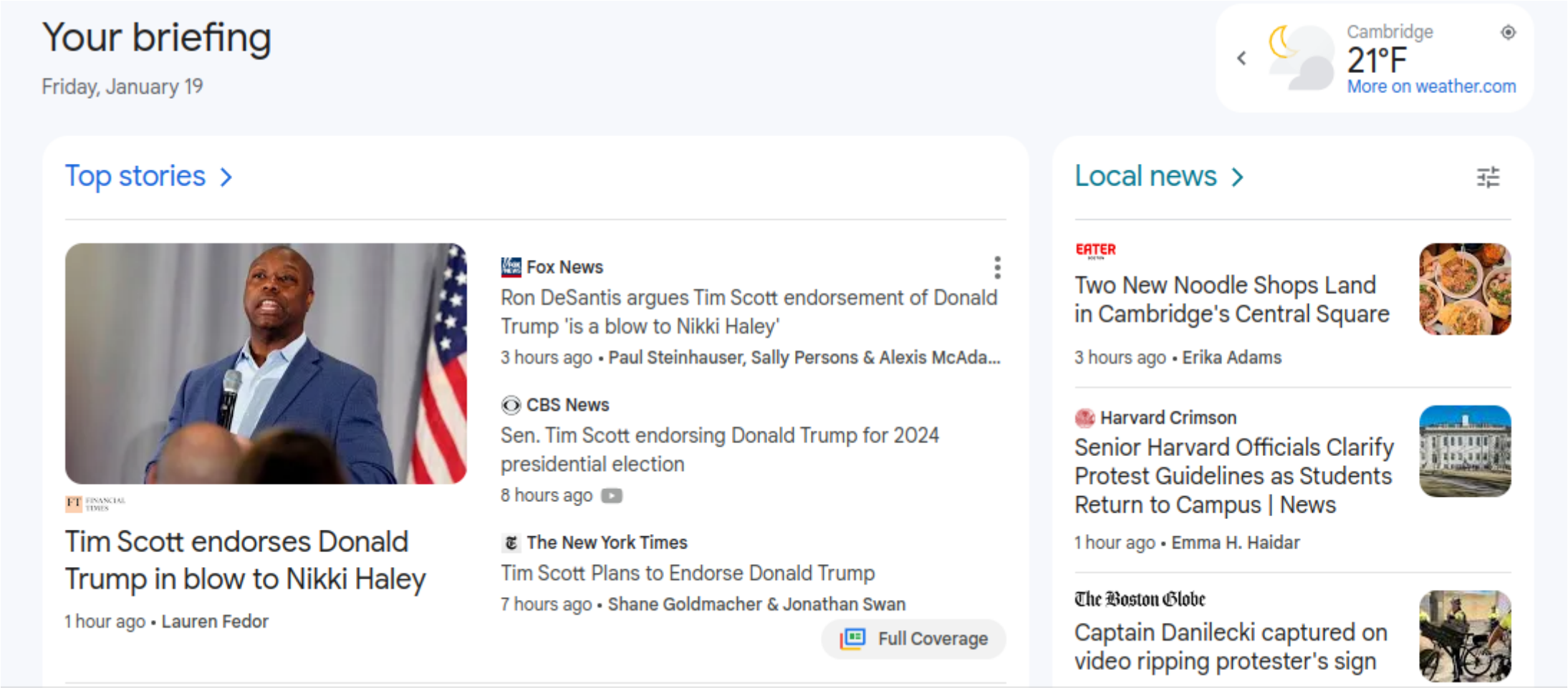
Structure of the Course

- Follow the “cube”++
  - Supervised: regression, classification
  - Unsupervised: clustering, embeddings
  - Graphical Models, Reinforcement Learning
- Excursions to discuss model classes, model selection, evaluation...



+ graphical models, **reinforcement learning**

Example: Google News



RL: Output is a **Recommendation**

Lecture Structure

- Story/real-life example
- Lecture content, quiet questions option
- Concept checks

Lectures will include real world examples and concept exercises; we'll also discuss ethics

Logistics

- HW0 due Friday: this is a check on your prereqs.
- In person. Let's help keep each other safe by masking if ill/exposed.
- Lecture recaps, CS181 "textbook" as resources (including if ill). You are responsible for finding ways to keep up with your learning!
- Regular homeworks: biweekly (8p Friday), done individually, combining theoretical, programming, and broader thinking elements (LaTex, Python)
- Six late days, at most two per problem set
- Two timed, closed-book midterms
- Sections: section zero: math+code this week, then weekly sections on content
- Lots of office hours (in person and on zoom)
- Ed: for all discussion and questions on course content (content, clarification, private message)
- Yes, you may sit in, audit, etc.!

Logistics

- HW0 due Friday: this is a check on your prereqs.
- In person. Let's help keep each other safe by masking if ill/exposed.
- Lecture recaps, CS181 "textbook" as resources (including if ill). You are responsible for finding ways to keep up with your learning!
- Regular homeworks: biweekly (8p Friday), done individually, combining theoretical, programming, and broader thinking elements (LaTex, Python)
- Six late days, at most two per problem set
- Two timed, closed-book mid
- Sections: section zero: math
- Lots of office hours (in perso
- Ed: for all discussion and qu
  message)
- Yes, you may sit in, audit, et

We're in difficult times.
Let's make a collective commitment to be human to each other.

Collaboration Policy

- You may work with others, but your write-up must be entirely written by yourself in your own words. You may help each other debug code, but again, the code must be written by you.
  - Think of GenAI as a collaborator: cite if/how you use tools.
- *Include the names of anyone you worked with in your write-up.*
- Do not copy parts of another person's assignment or jointly type up an assignment.
- You can make use of textbooks and online sources to help in answering questions, but you must cite your sources (and you should be ready to explain your answer to a member of the teaching staff)
- Do not look up solutions to specific questions from the internet or other sources (e.g. friends from previous years)

A Short Break (and then, … nonparametric regression)