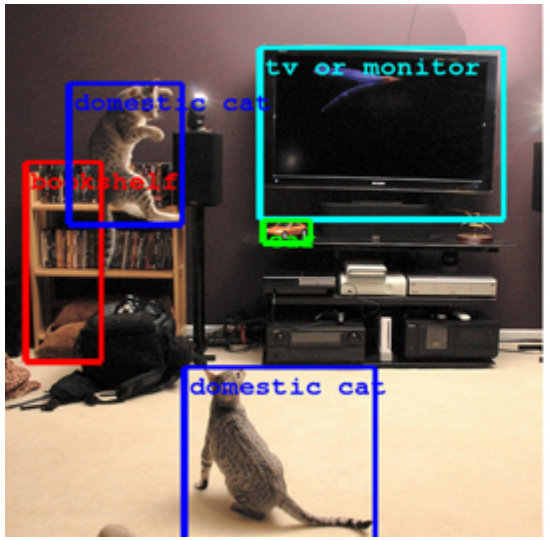
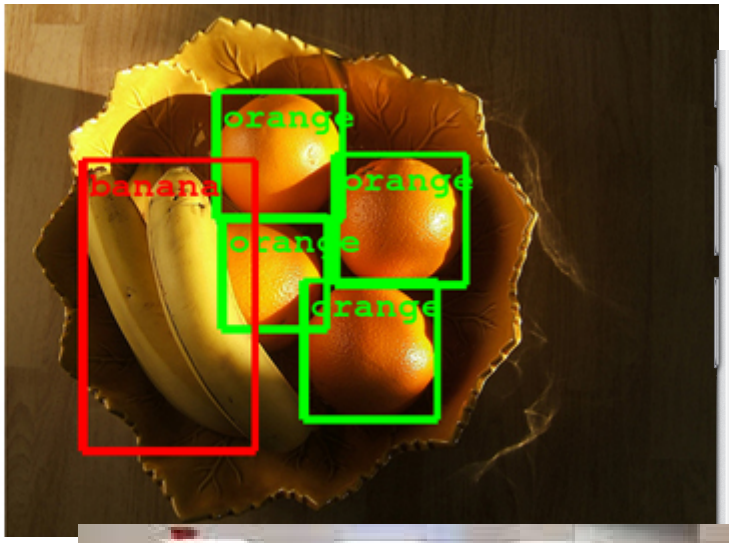
CS181: Introduction to Machine Learning

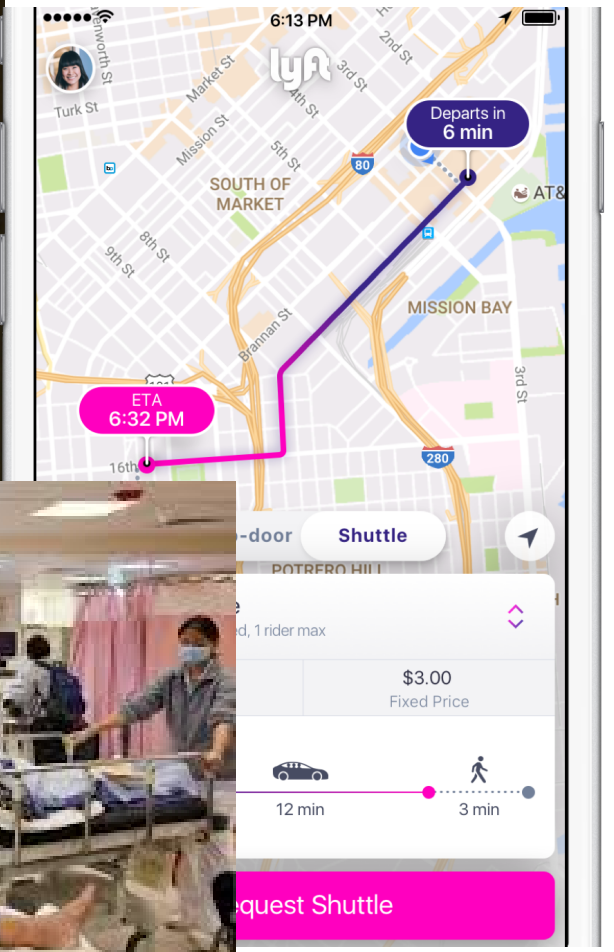
Finale Doshi-Velez

HARVARD
School of Engineering
and Applied Sciences

What this course is about

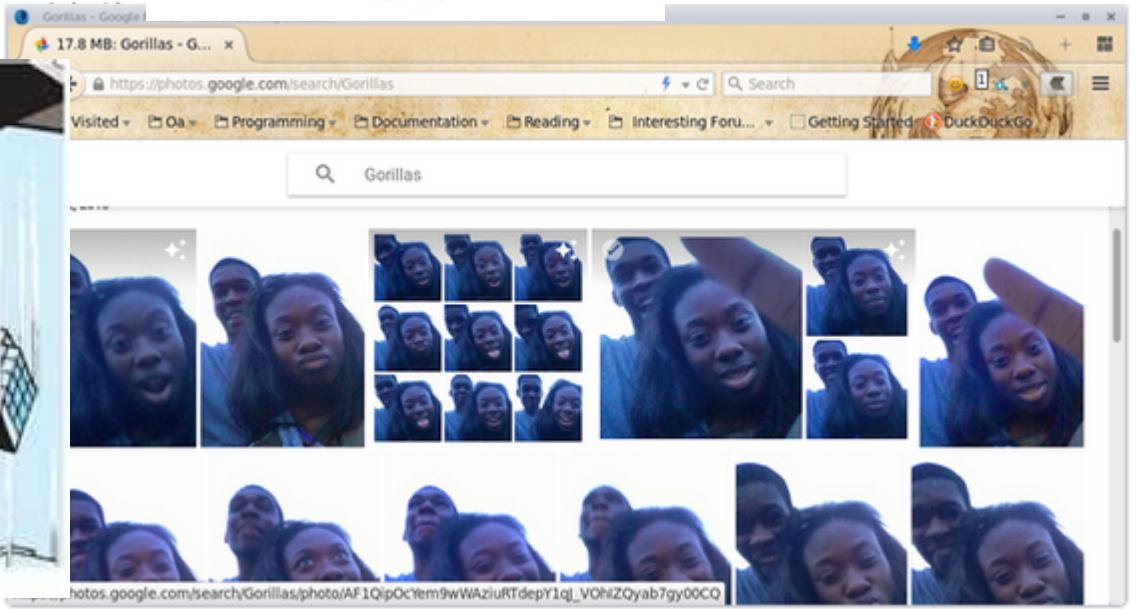






https://s.aolcdn.com/hss/storage/midas/cb433fb6c234b014394ad56714ceee8b/200707597/google-image-recognition-tech.jpg
https://cdn.shopify.com/s/files/1/0684/3445/products/Partner_AmazonEcho_small_031932b3-c31d-4977-a922-5ebbe052bad6.jpg?v=1497504913
https://cdn.lyft.com/brochure/shuttle-screenshot.1e79cd66.png
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



https://twitter.com/geraldmellor/status/712880710328139776/photo/1?ref_src=twsrc%5Etfw&ref_url=https%3A%2F%2Fwww.theverge.com%2F2016%2F3%2F24%2F11297050%2Ftay-microsoft-chatbot-racist
https://twitter.com/jackyalcine/status/615331869266157568?ref_src=twsrc^tfw&ref_url=https%3A%2F%2Fwww.usatoday.com%2Fstory%2Ftech%2F2015%2F07%2F01%2Fgoogle-apologizes-after-photos-identify-black-people-as-gorillas%2F29567465%2F
https://needamortgageloan.files.wordpress.com/2012/05/subprime.jpg
http://static3.businessinsider.com/image/519285ffecad046054000014-506-253/clippy-microsofts-talking-paperclip-is-back.jpg

What this course is about

You will develop a core machine learning understanding which will serve as a foundation for becoming a

- Informed product lead
- ML developer
- ML researcher

What this course is about

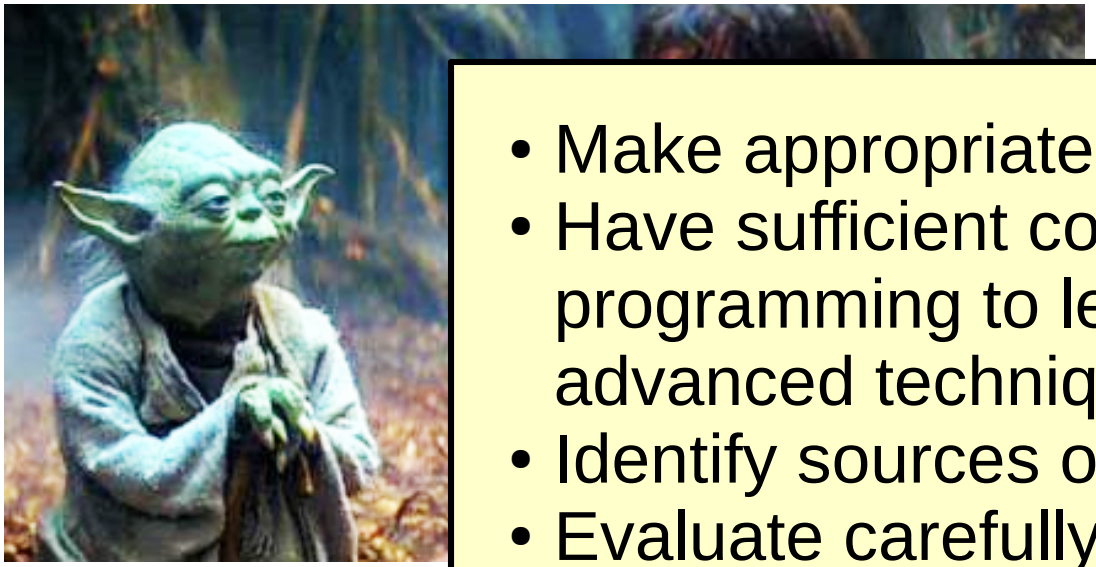
You will develop a core machine learning understanding which will serve as a foundation for becoming a

- Informed product lead
- ML developer
- ML researcher

What this course is about

You will develop a core machine learning understanding which will serve as a foundation for becoming a

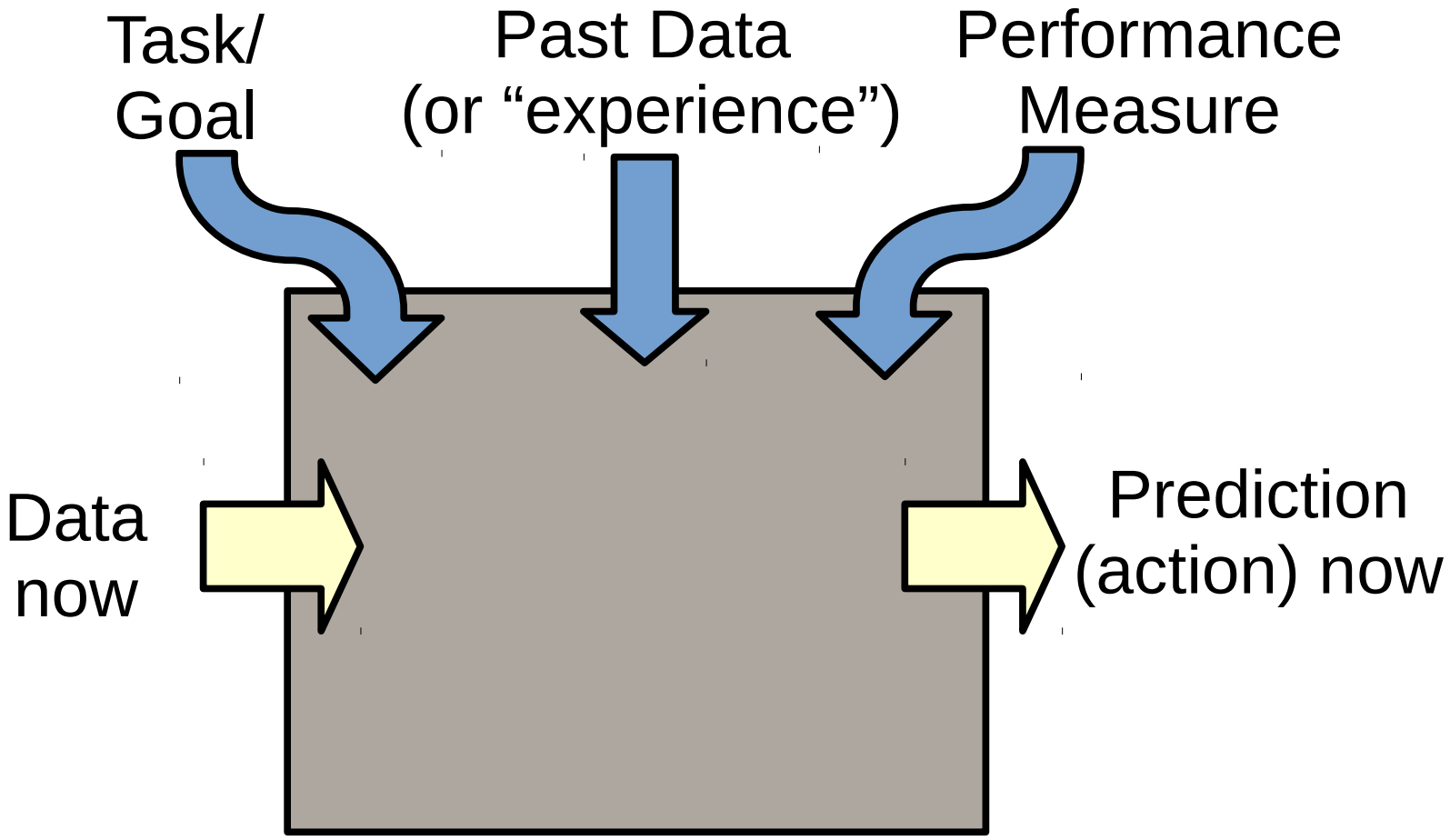
- Informed product lead
- ML developer
- ML researcher



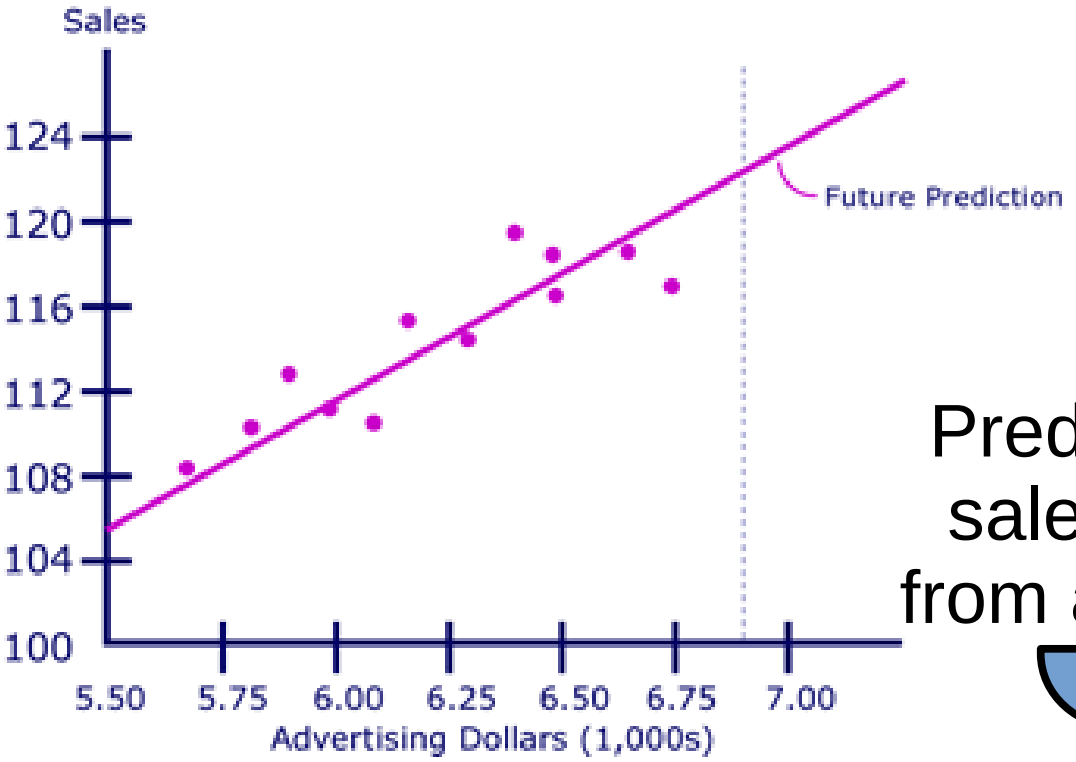
- Make appropriate model choices
- Have sufficient concepts, math, and programming to learn newer, more advanced techniques
- Identify sources of error
- Evaluate carefully

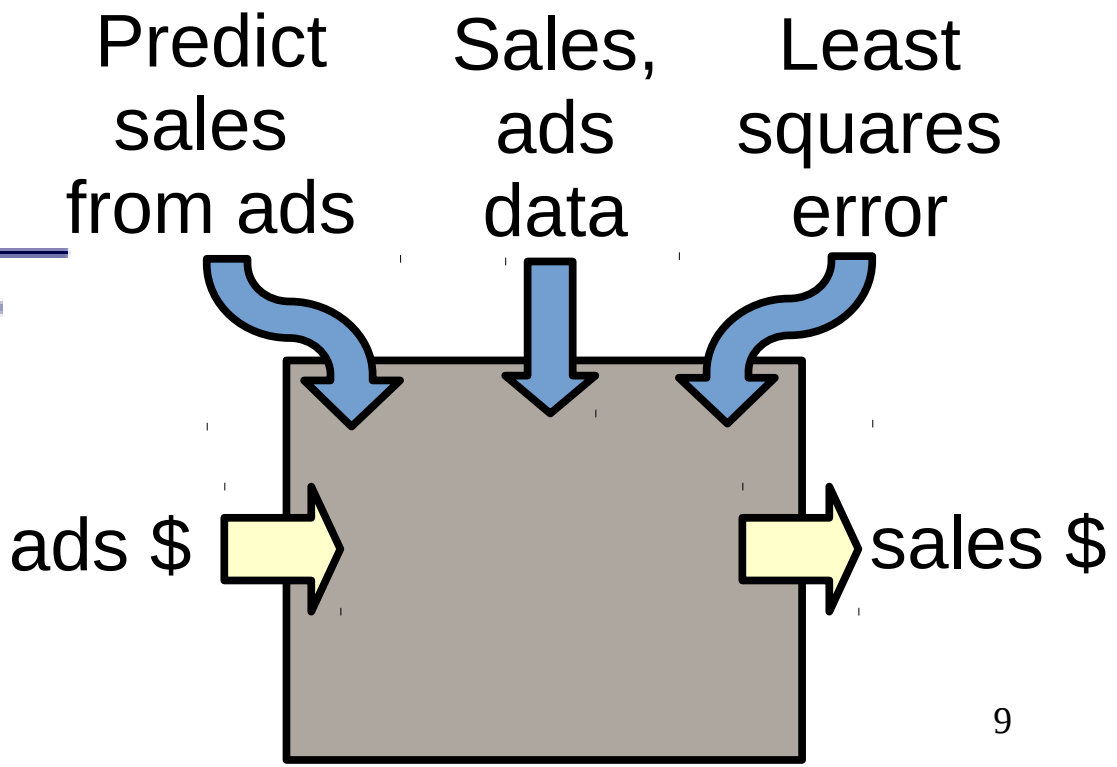
Let's get to it!

What is machine learning?



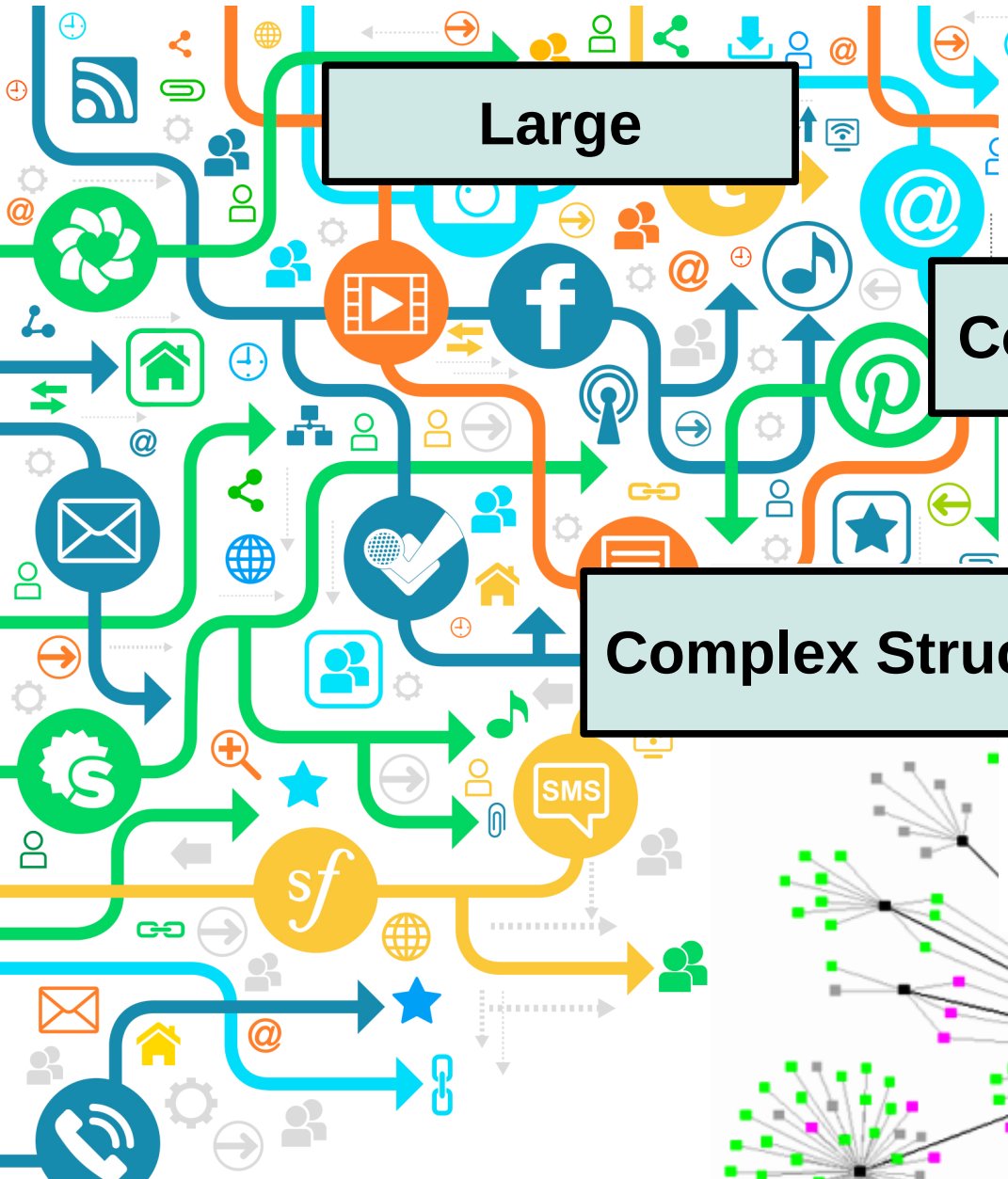
The starting point...

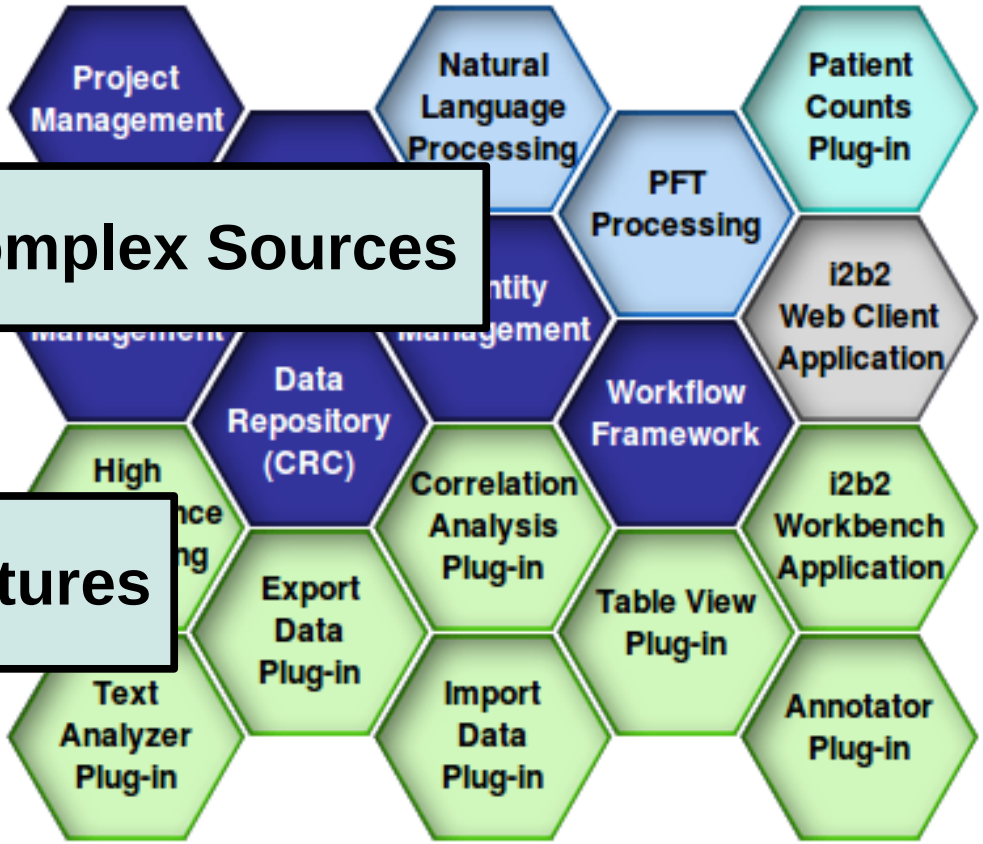


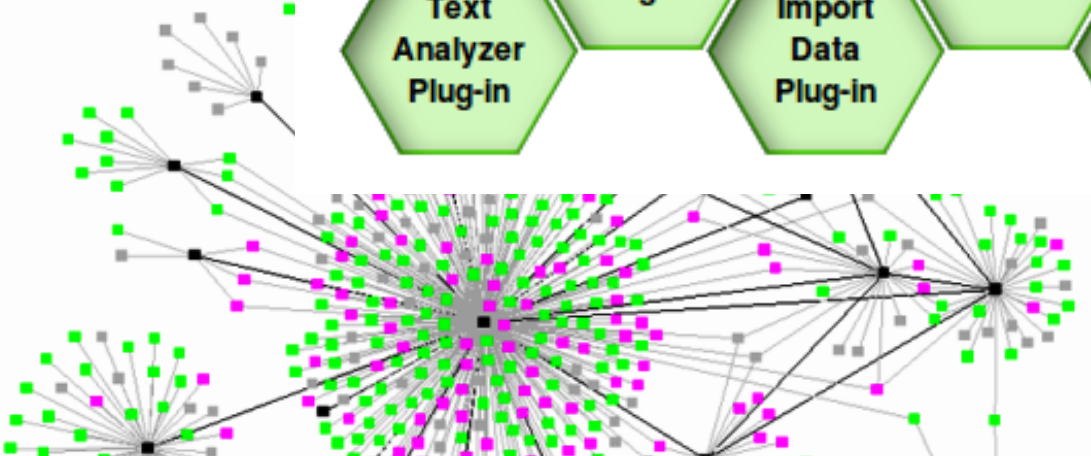


http://ci.columbia.edu/ci/premba_test/c0331/images/s7/7176267017.gif

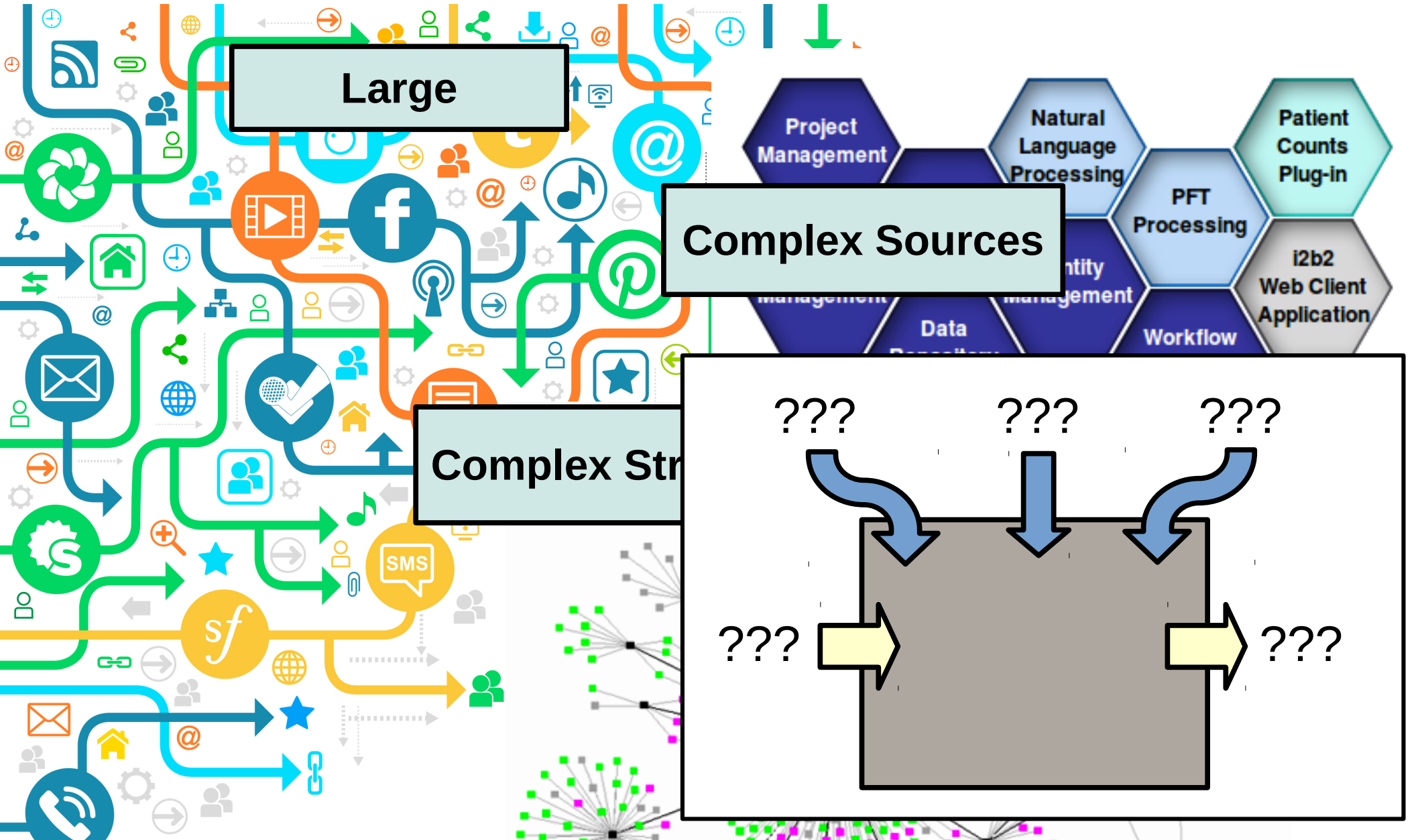
… where we are now







… where we are now



A Story: Treating Depression

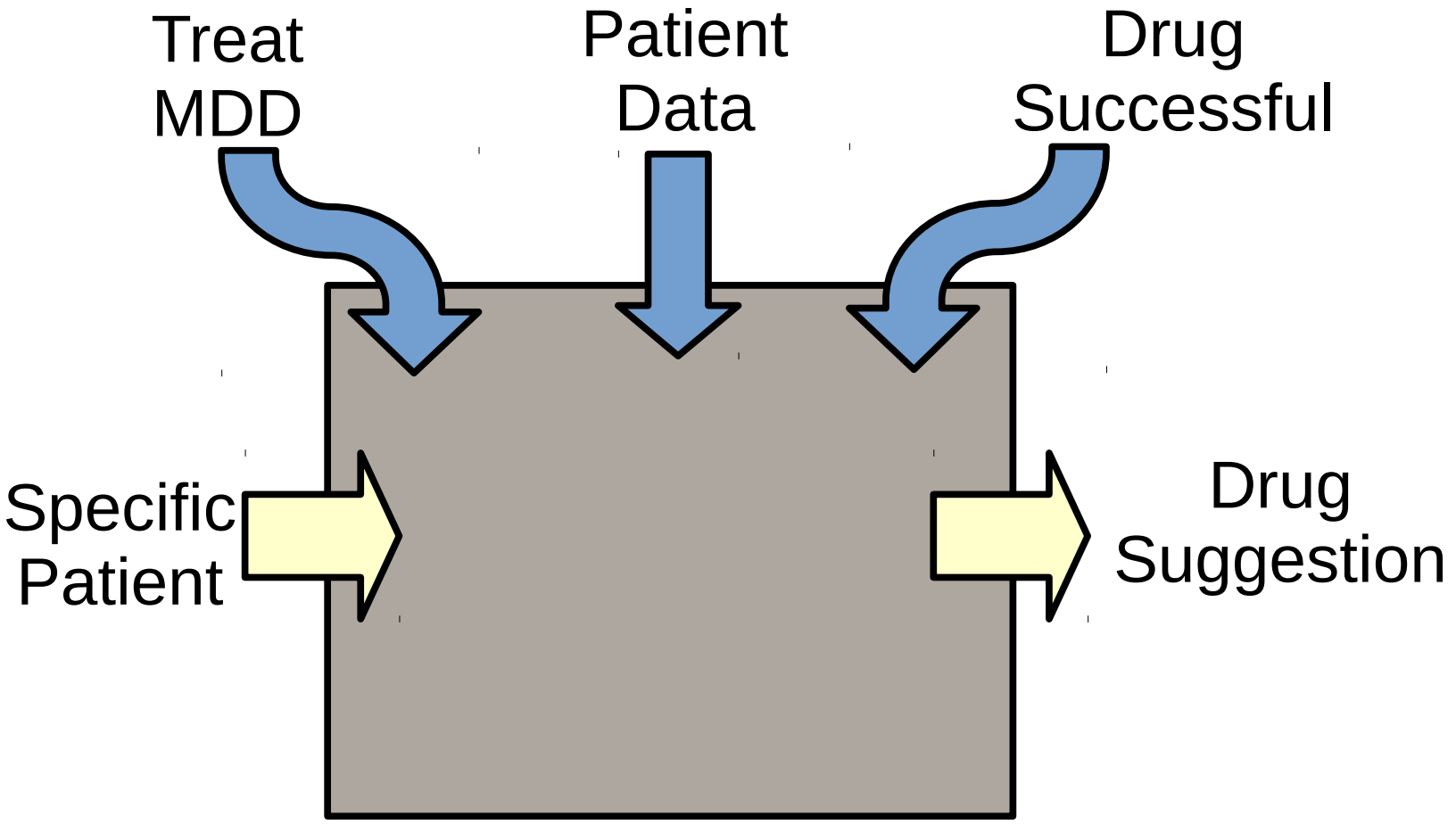
Clinical Question: What meds to give what depression patients?

Clinical Question: What meds to give what depression patients?

Clinical Question: What meds to give what depression patients?

Also: Can we explain who responds to what drug?

Formalizing...



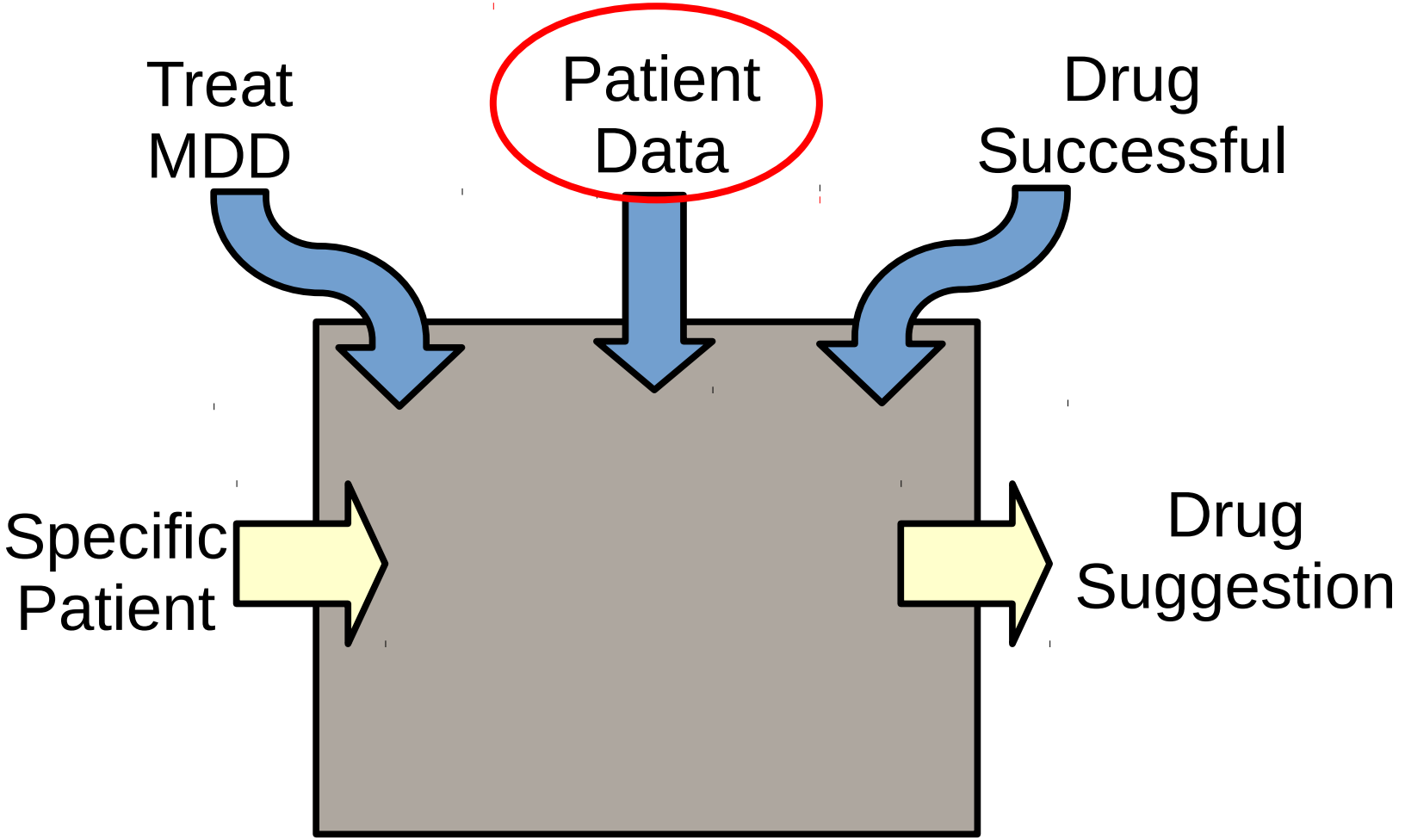
Current Approaches

Most current studies ask narrow questions and require specialized data. Examples:

- Large-scale clinical trial (STAR*D, COMED, iSPOT-D) analyses to decide between certain features or pairs of drugs (e.g. Chekroud et al., Joyce et al., Lavretsky et al.)
- MRIs and biomarkers to determine subtype and treatment choice (e.g. Liston et al., Craighead et al., Breitenstein et al.)

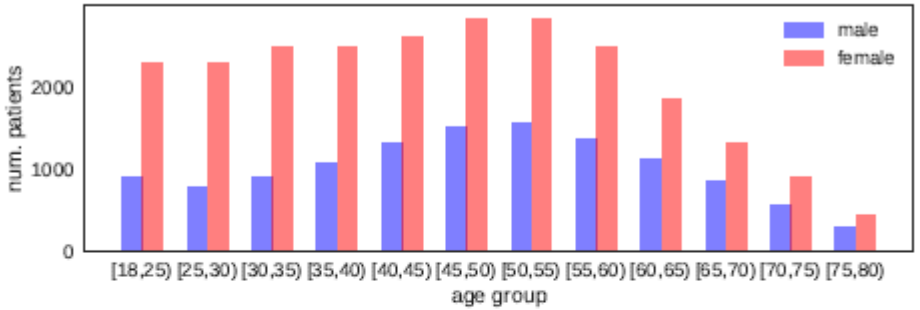
Our goal: We want to be able to recommend all common drugs, for patients with any prior treatment history.

Formalizing...



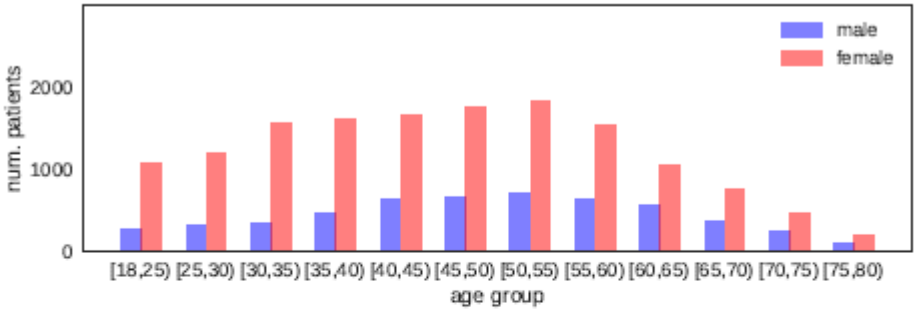
Cohort Statistics

Site A



	female	male	total	frac.
Asian	603	222	825	0.022
Black	959	412	1371	0.037
Hispanic	1006	402	1408	0.038
Other	2153	885	3038	0.082
White	20158	10415	30575	0.822
total	24879	12336	37217	
frac.	0.668	0.331		

Site B



	female	male	total	frac.
Asian	233	56	289	0.014
Black	1604	419	2023	0.100
Hispanic	2264	657	2921	0.145
Other	1026	386	1413	0.070
White	9662	3888	13551	0.671
total	14789	5406	20197	
frac.	0.732	0.268		

Featurization

- Of the 22,000 unique codes (ICDs, CPTs, RXNORM) in the EHR, retain 7,291 which occur in at least 1,000 distinct patients.
- Focus on 10 primary drugs that are prescribed in at least 1000 patients

nortriptyline	sertraline	mirtazapine
amitriptyline	paroxetine	citalopram
bupropion	venlafaxine	escitalopram
fluoxetine		

- Success: require same primary over 90 days, with a visit frequency of at least every 13 months.

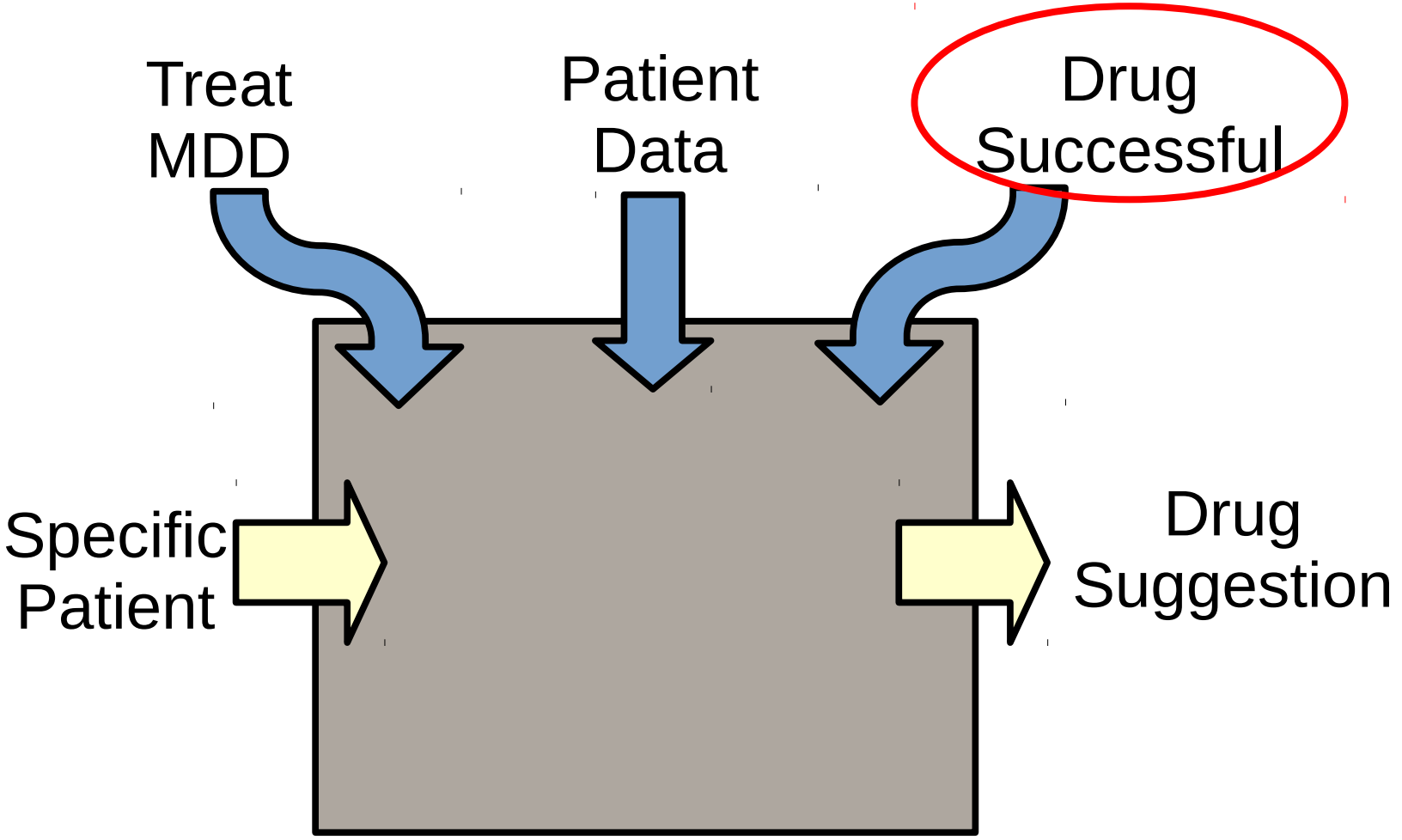
Featurization

- Of the 22,000 unique codes (ICDs, CPTs, RXNORM) in the EHR, retain 7,291 which occur in at least 1,000 distinct patients.
- Focus on 10 primary drugs that are prescribed in at least 1000 patients

nortriptyline	sertraline	mirtazapine
amitriptyline	paroxetine	citalopram
bupropion	venlafaxine	escitalopram
fluoxetine		

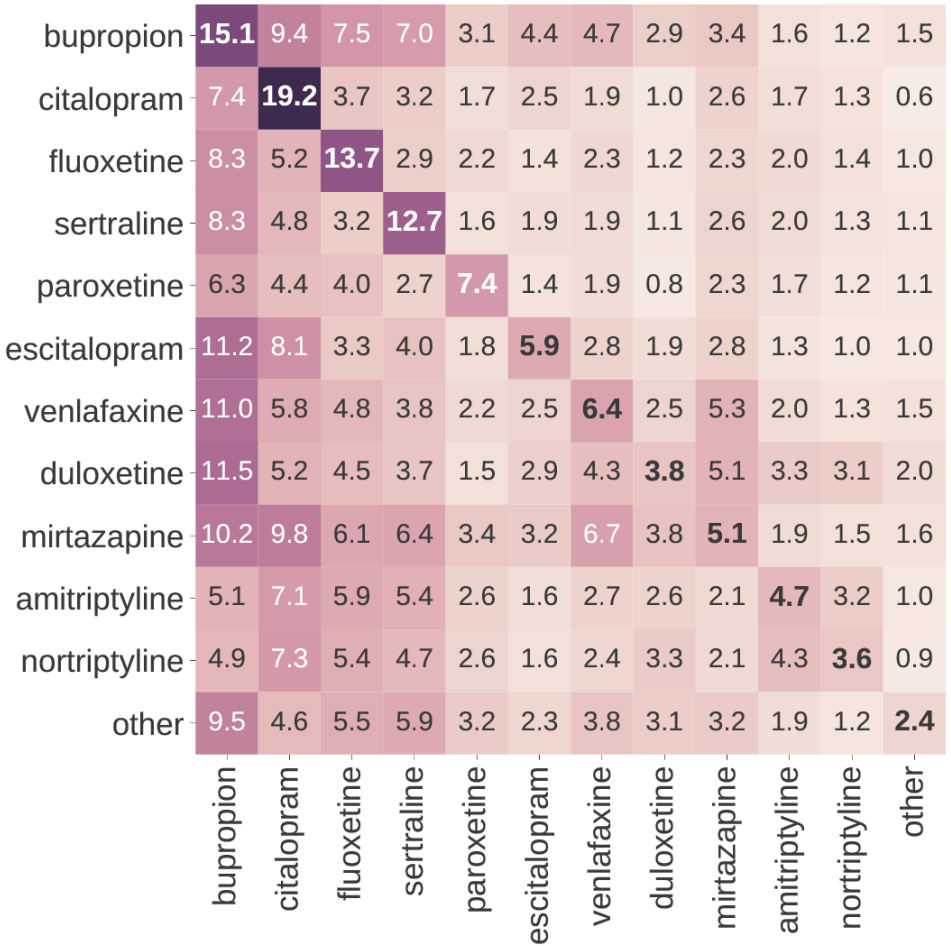
- Success: require same primary over 90 days, with a visit frequency of at least every 13 months.

Formalizing...

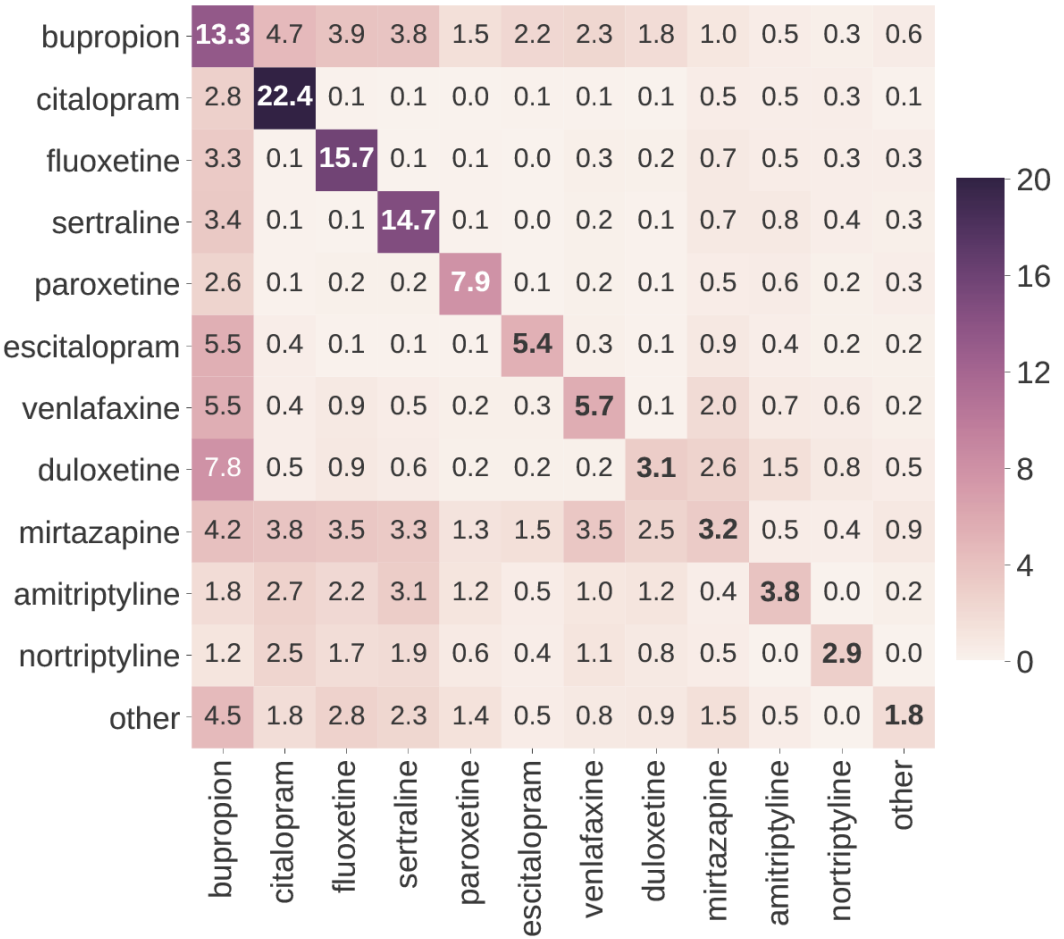


Med Statistics

Observed Combinations

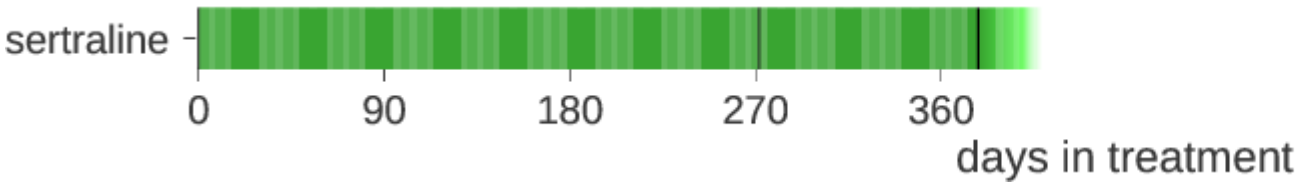


Target Combinations

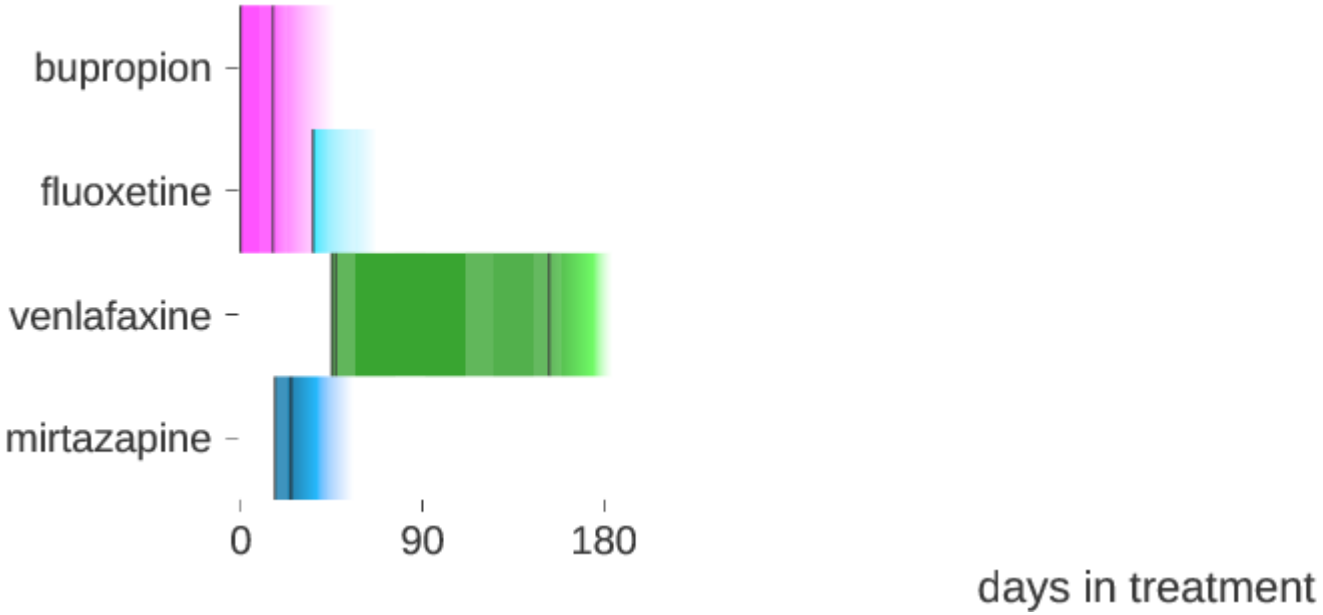


Example Trajectories

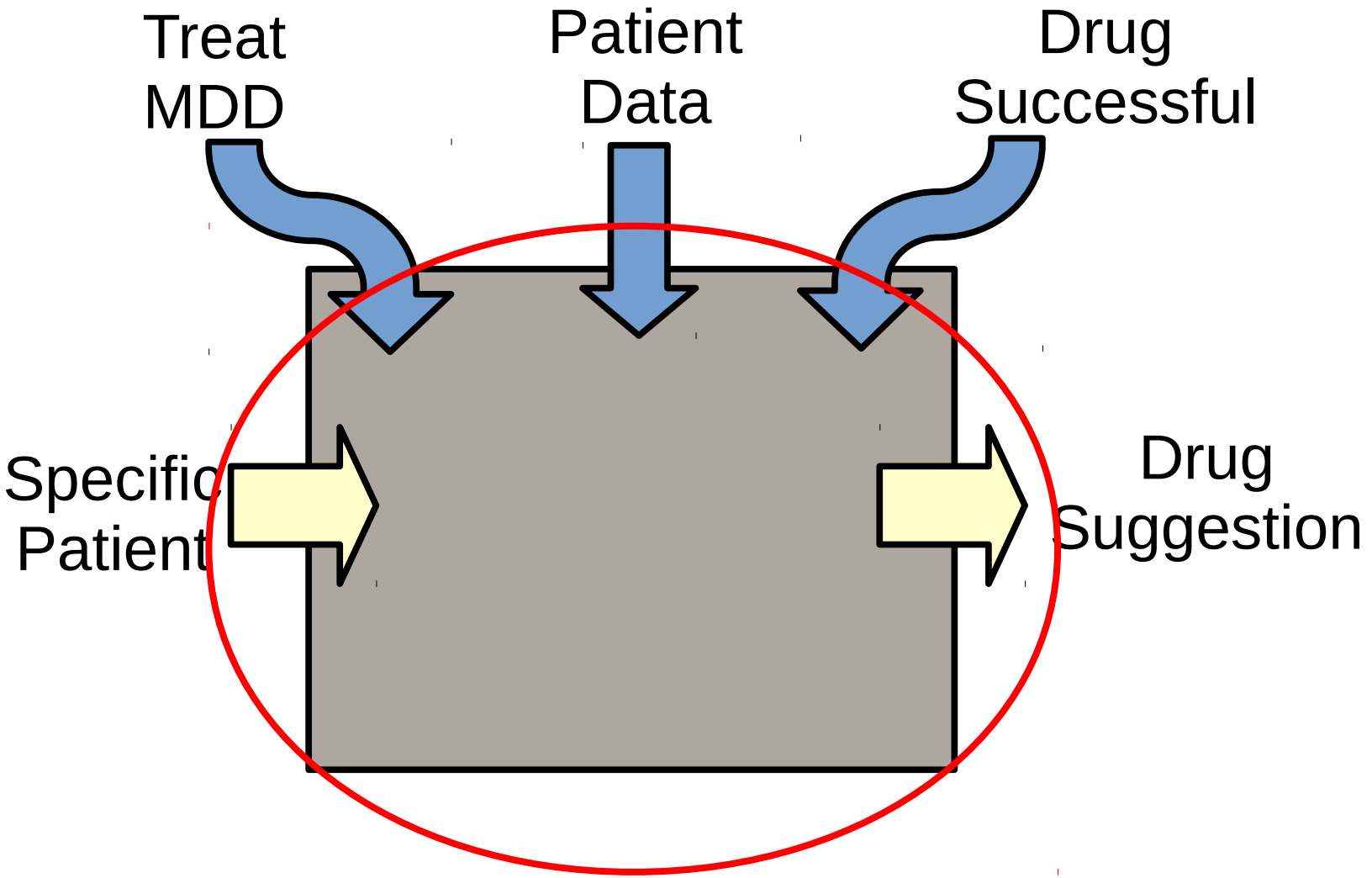
“Instant Success” Patient



“Eventual Success” Patient



Formalizing...



Choices for Classifiers

In high dimensions, standard predictors (logistic regression, random forests, neural nets…) often have a tension between sparsity and interpretability.

- "Hip fracture" could code for "elderly"
- "Pregnant" could code for "female"

Thus, we choose to apply topic models for dimensionality reduction, and then predict based on the doc-topic probabilities.

Topic Models

Topics are distributions over words (codes)

$$\phi_k \sim Dir(\alpha_v)$$

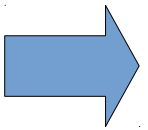
Patients are distributions over topics

$$\pi_d \sim Dir(\alpha_p)$$

$$w_d^i \sim Dir(\pi_d \phi)$$

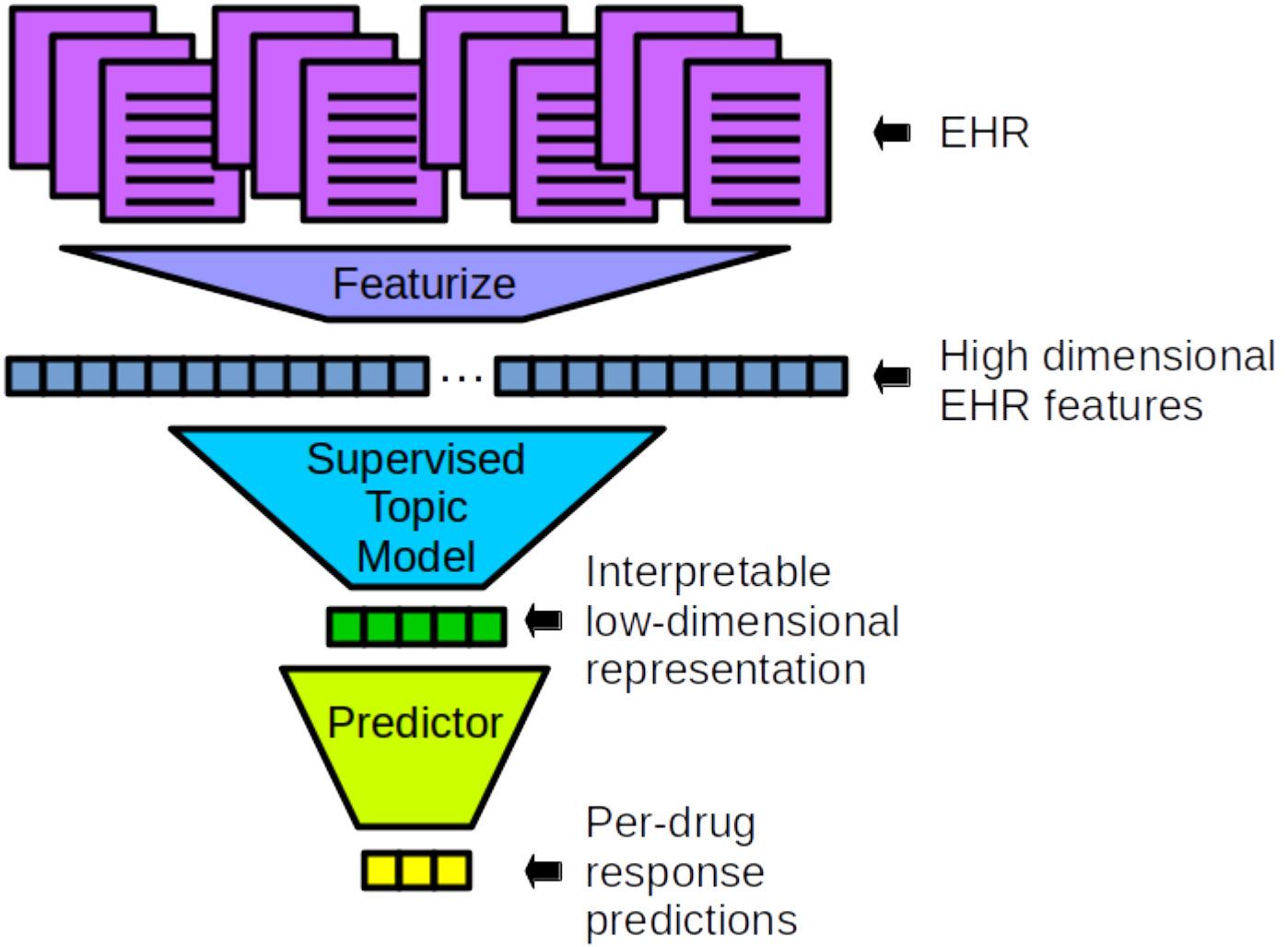
Example topic

```
1.0000   29650:bipolar_affective_disorder,_depres
0.9999    2967:bipolar_affective_disorder,_unspec
0.9999   29570:schizo-affective_type_schizophreni
0.9999   29660:bipolar_affective_disorder,_mixed,
0.9998  c90870:electroconvulsive_therapy_(include
0.9998  c00104:anesthesia_for_electroconvulsive_t
0.9997   29560:residual_schizophrenia,_unspecifie
0.9996   p9427:other_electroshock_therapy
0.9993  d00061:lithium
0.9993   29653:bipolar_affective_disorder,_depres
0.9985   29651:bipolar_affective_disorder,_depres
0.9985  d04825:aripiprazole
```

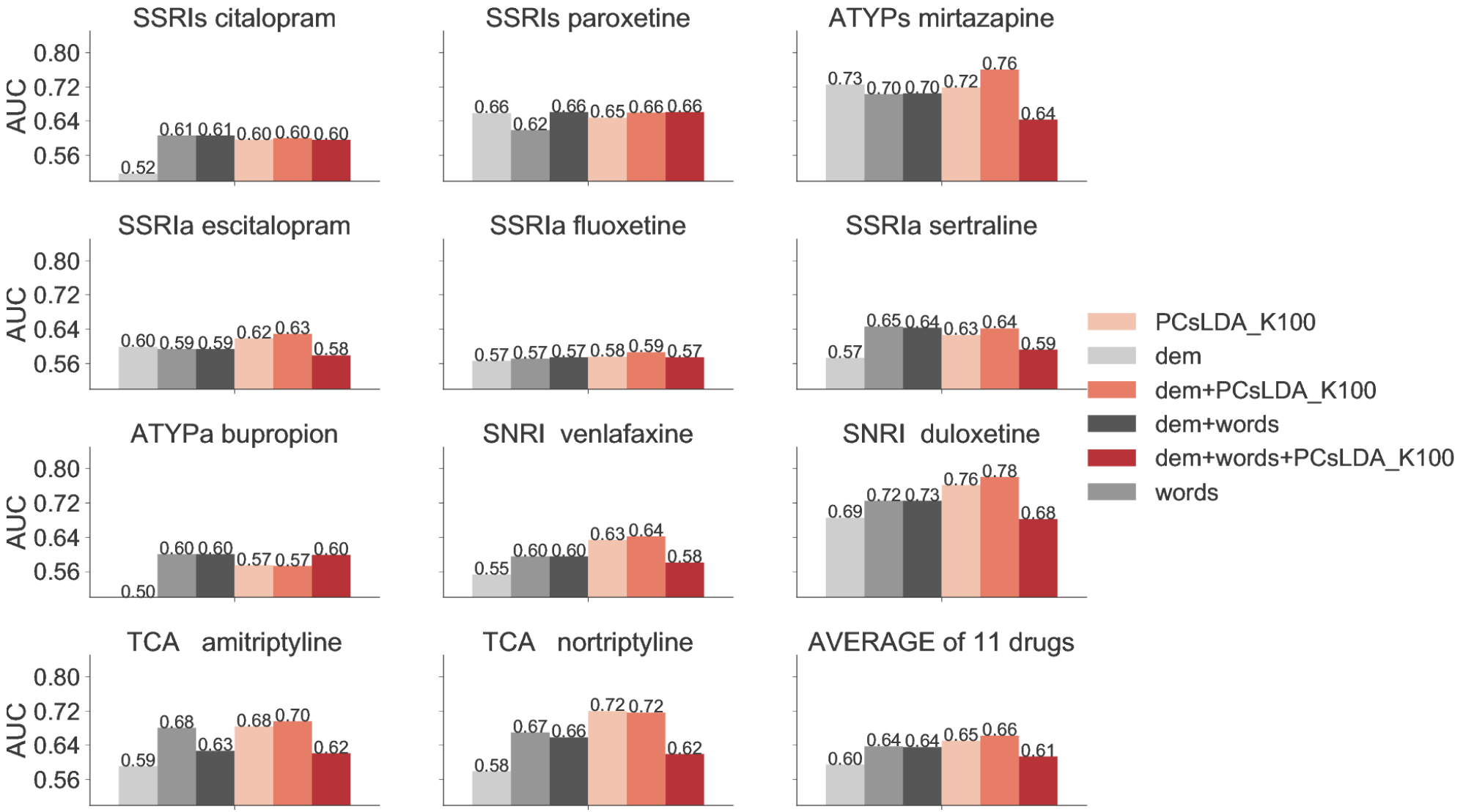


Patient has bipolar disorder

Pipeline


EHR
Featurize
High dimensional
EHR features
Supervised
Topic
Model
Interpretable
low-dimensional
representation
Predictor
Per-drug
response
predictions

Drug-by-Drug Performance



Interpretation: Bupropion

BPsLDA +7.7
0.60 nortriptyline
0.27 nonspecific abnormal findings
0.21 other specified local infection
0.20 embrionic cyst of the fallopian tube
0.18 application of the intervertebrea...
0.16 other malignant neoplasm...
0.15 amoxicilllin/clarithromycin
0.15 need for prophylactic vaccine
0.15 observation or inpatient visit...

Gibbs -0.6
1.0000 bipolar, depressive
0.9999 bipolar, unspecified
0.9999 schizo-affective schizophrenia
0.9999 bipolar, mixed
0.9998 electroconvulsive therapy
0.9998 anesthesia for ECT
0.9997 residual schizophrenia
0.9996 other electroshock therapy
0.9993 lithium

PCLDA +3.8
0.99 migraine, unspecified, without...
0.99 other malaise and fatigue
0.99 common migraine...
0.99 sumatriptan
0.99 asa/butalbital/caffeine
0.99 zolmitriptan
0.99 migraine, unspecified, with..,
0.99 classical migraine, without...
0.99 classical migraine, with...

Interpretation: Bupropiom

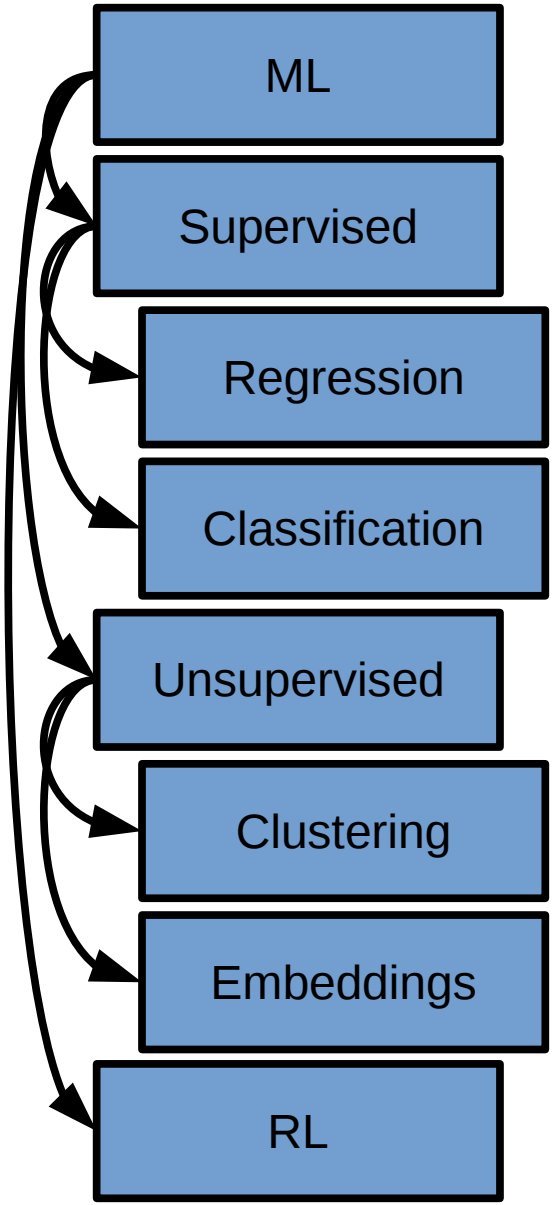
BPsLDA -15.8
0.39 visual field defect, unspecified
0.39 citalopram
0.36 microdissection
0.35 need for prophylactic vaccine
0.31 pet imaging regional or wide area
0.29 visual discomfort
0.29 accident poison by heroin
0.29 personal history of alcohol
0.27 other specified intestinal disorder

Gibbs -2.7
0.9997 respiratory_failure
0.9995 cystic fibrosis...
0.9994 end stage renal disease
0.9992 unspecified septicemia
0.9984 debility unspecified
0.9970 insertion of endotracheal tube
0.9964 sevelamer
0.9964 continuous mech ventilation
0.9963 intubation endotracheal

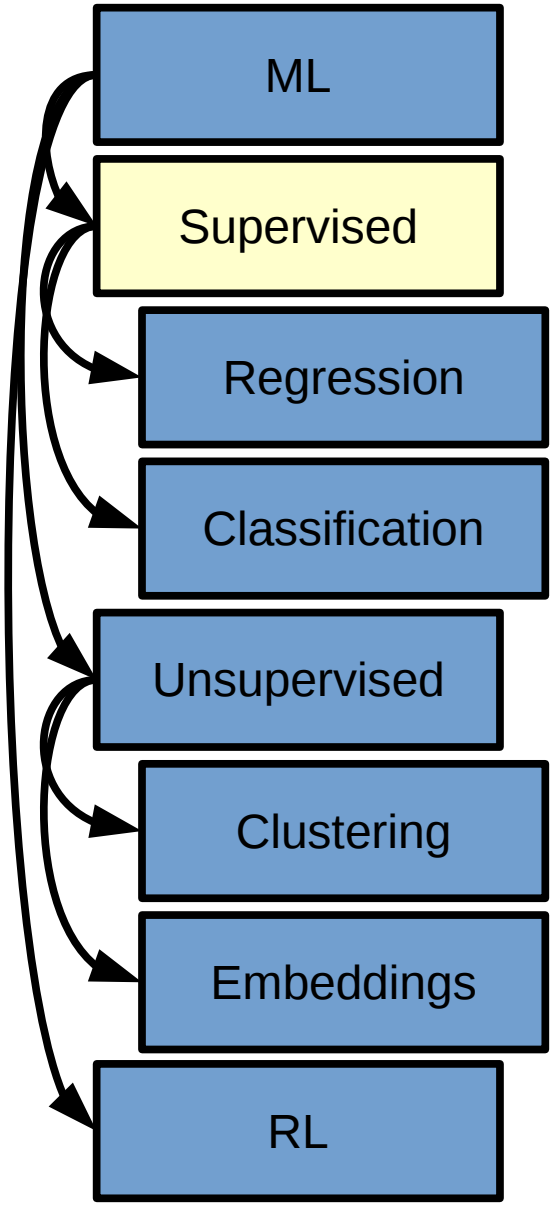
PCLDA -26.4
1.00 semen analysis, complete...
1.00 male infertility, unspecified
1.00 lipoprotein, direct measurement
0.99 sperm isolation, simple…
0.99 tissue culture for non-neoplasm…
0.99 conditions due to anomaly…
0.99 vasectomy, unilateral or…
0.99 arthrocentesis
0.99 scrotal varices

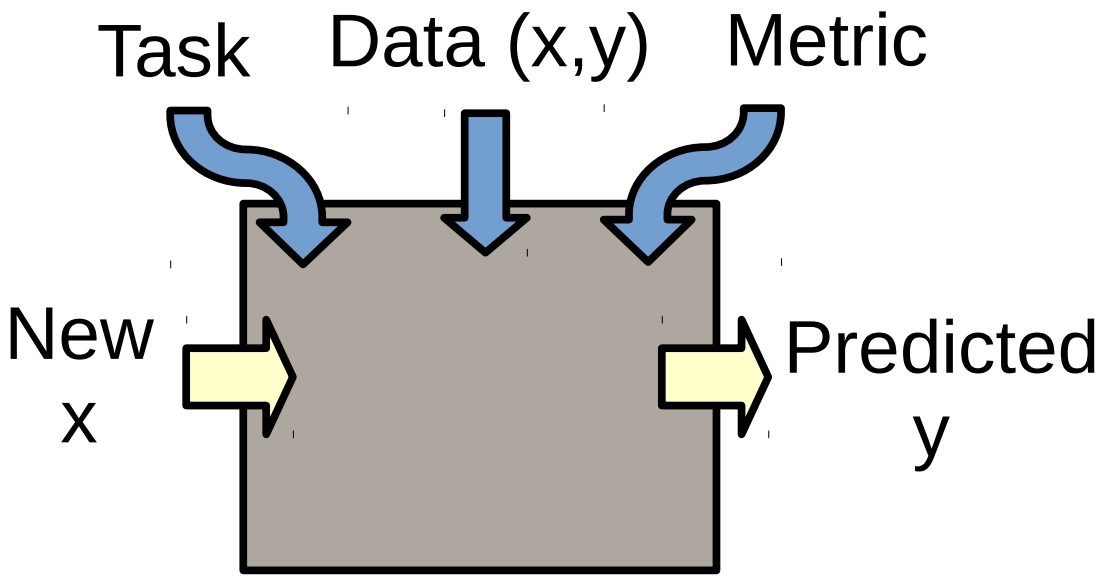
Kinds of Problems We'll Explore

Machine Learning Taxonomy

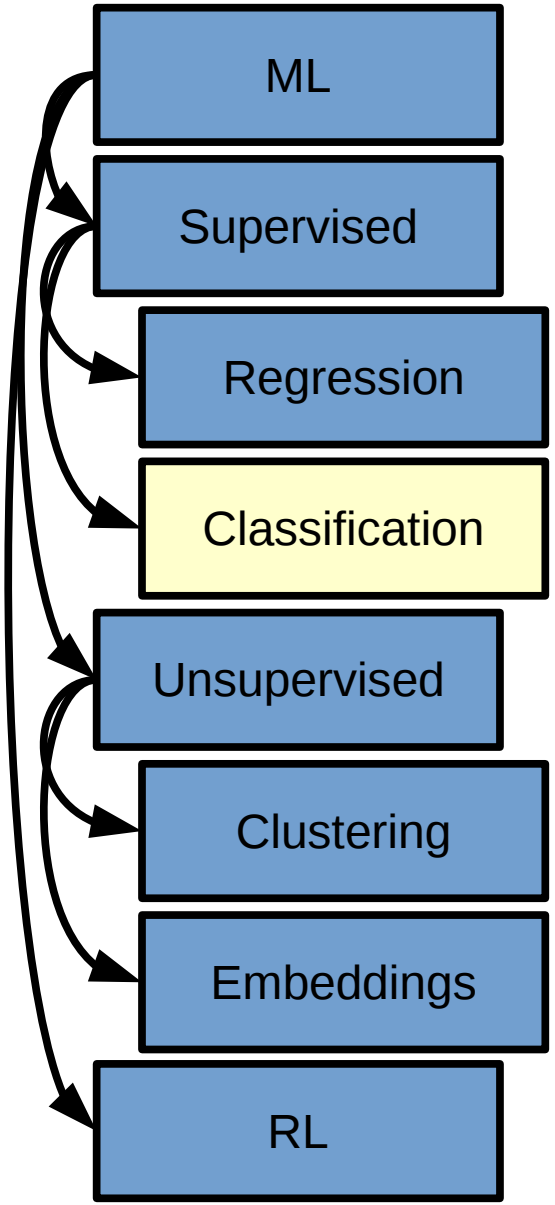


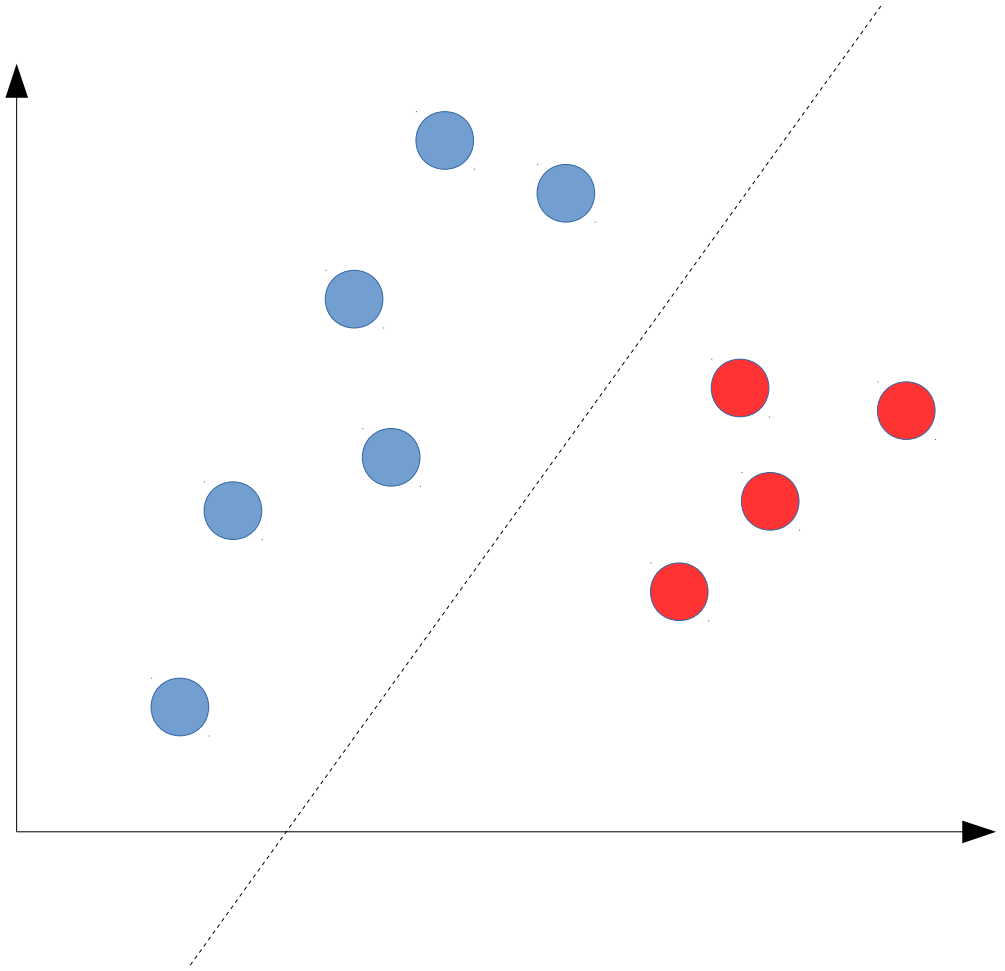
Machine Learning Taxonomy





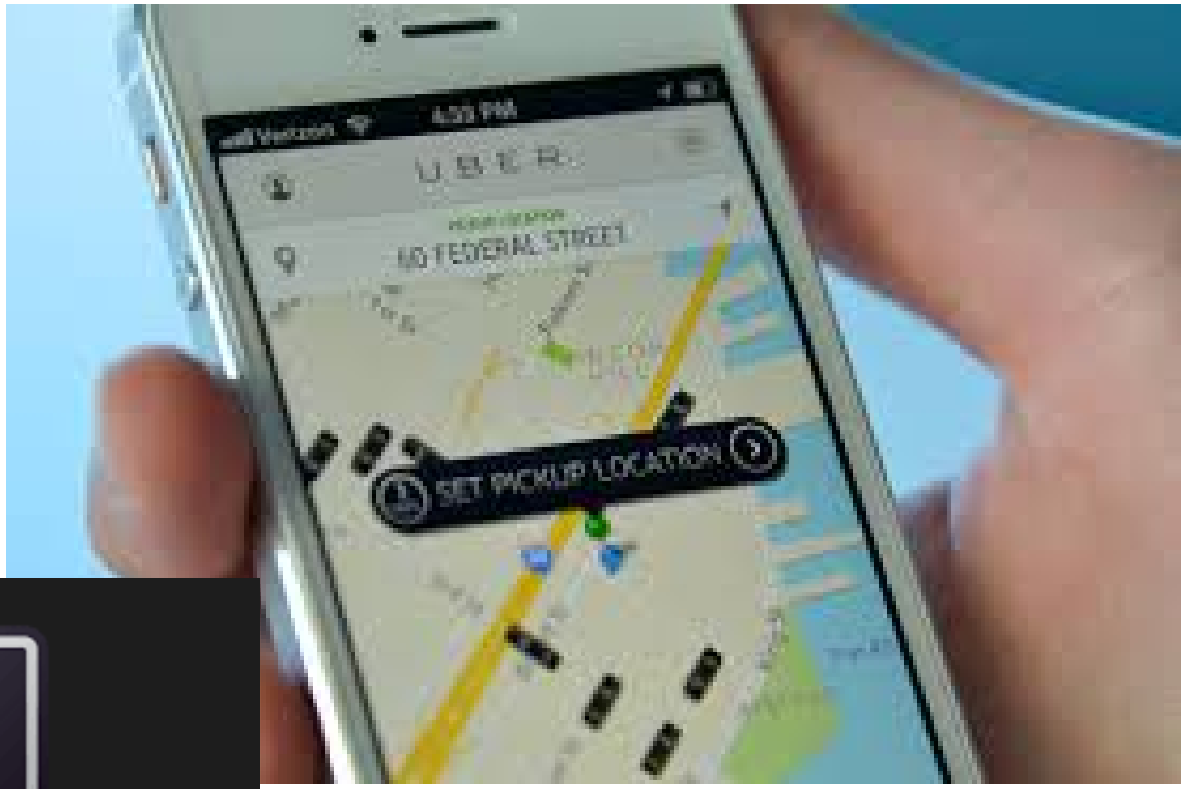
Terminology: Classification





Example: Uber

Improving product, expanding market: Analyzing routes, to pool customers

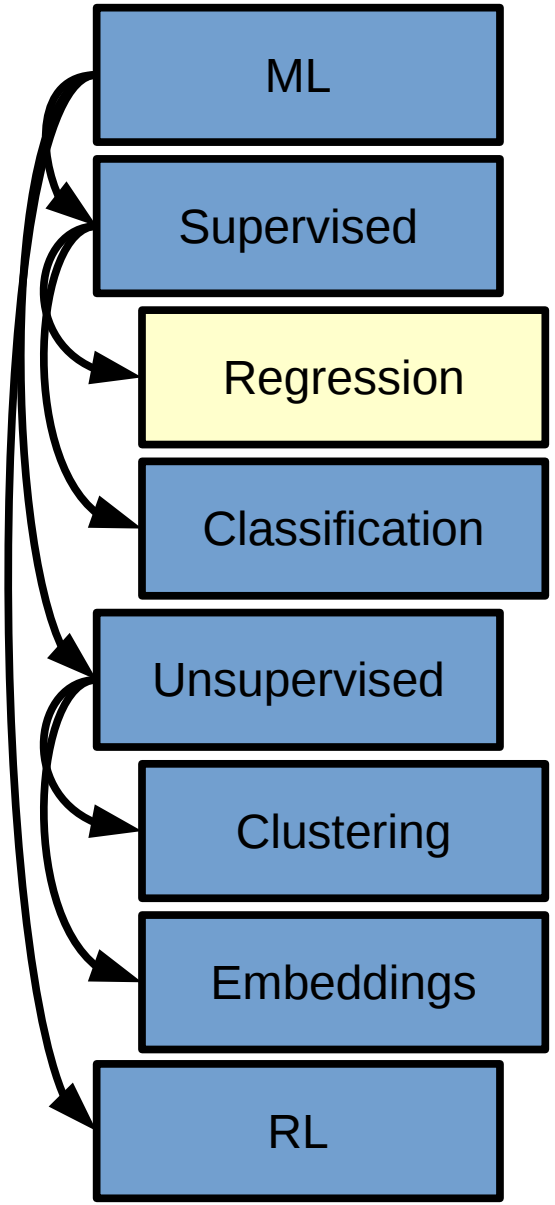


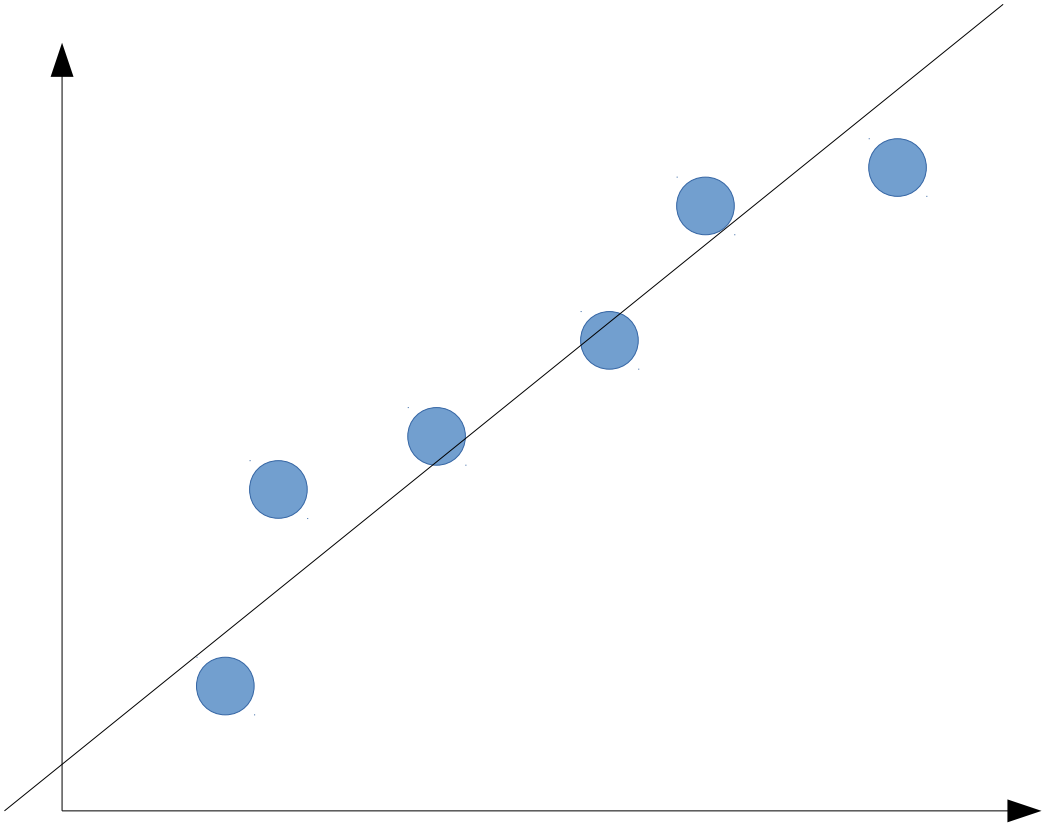
UBER

Example: Swype

Novel Product:
An easier way
to input text on
mobile devices

Terminology: Regression



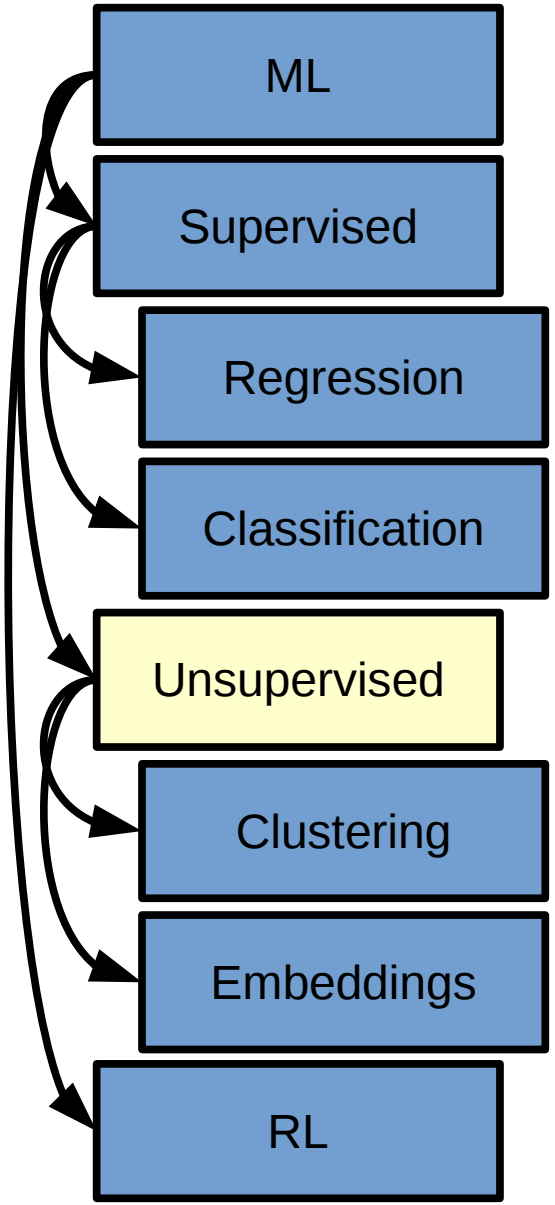


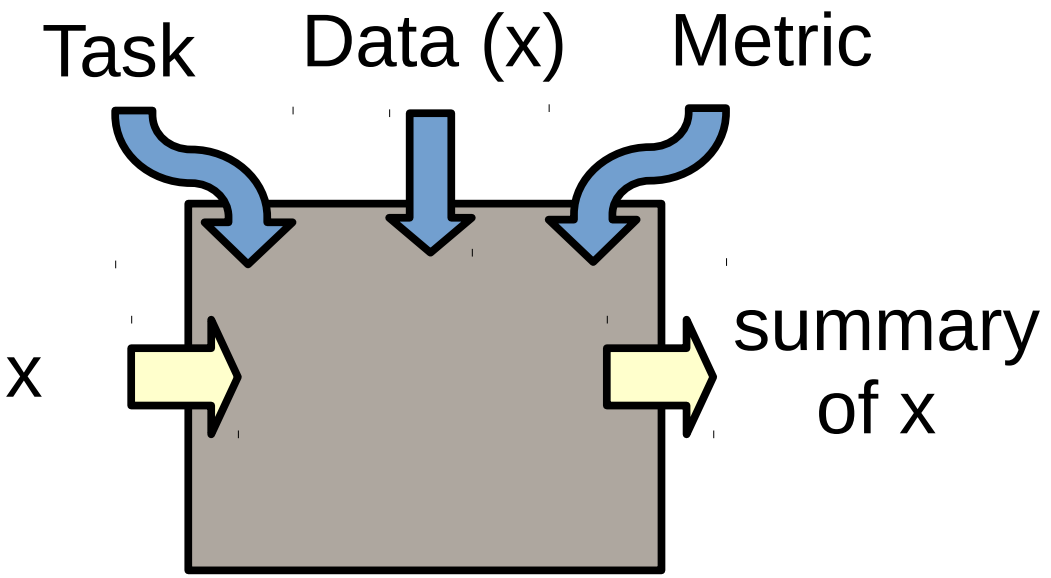
Example: Virtu

VIRTU FINANCIAL

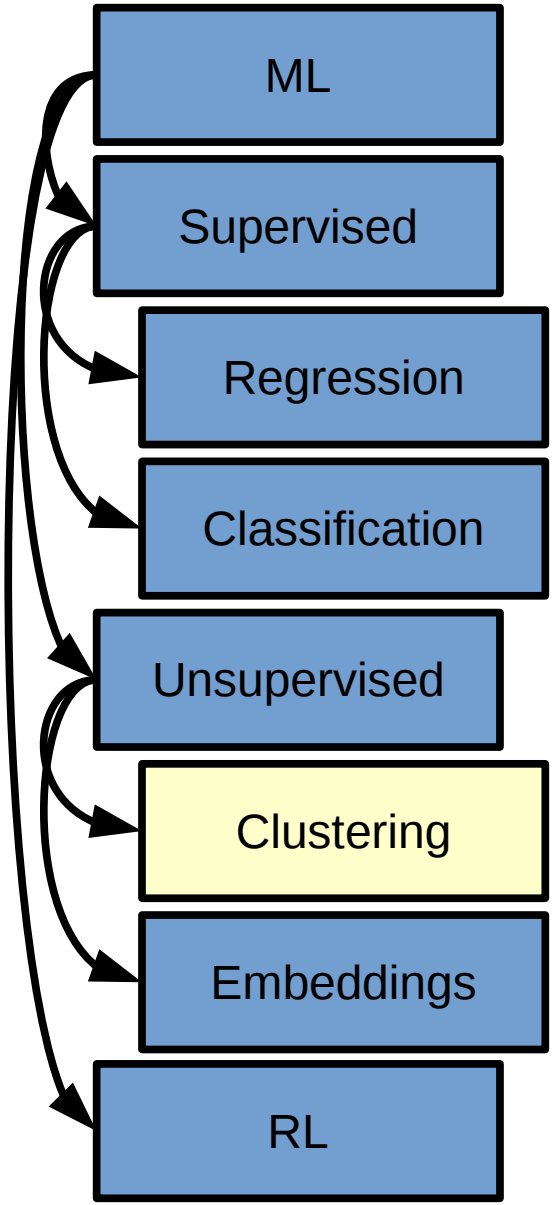
Core technology: Choosing what to trade, and when

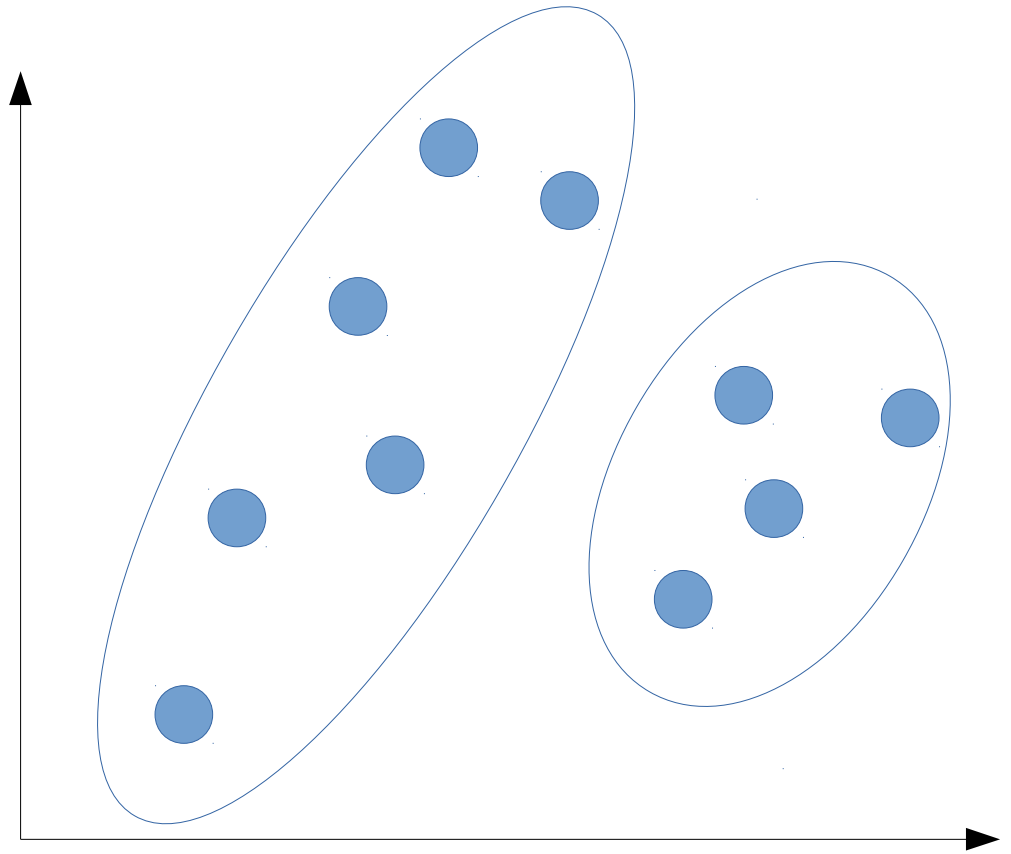
Machine Learning Taxonomy





Terminology: Clustering





Example: News

Top Stories

Fox News

See realtime coverage

Intensive manhunt underway after daring jail escape in California

Fox News - 37 minutes ago

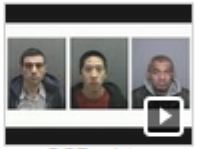
An intensive manhunt was underway Monday for three inmates who pulled a “Shawshank”-style escape through a hole in their California jail cell -- and, who may have ties to notorious Vietnamese street gangs and Iran.

Manhunt Expands For 'Dangerous' Trio After Daring Jailbreak NBCNews.com

Orange County manhunt: Officials suggest violent jail escapees could be hiding nearby Los Angeles Times

Related
California »

In Depth: Authorities struggling to piece together daring jail escape Washington Post



OCRegister

City News Los A...

WOODTV.com

Los Angel...

Philly.com

OCRe

Huffington Post

See realtime coverage

HUFFPOLLSTER: Trump And Clinton Lead, But Iowa Polling Remains Volatile With A Week To Go

Huffington Post - 5 hours ago

Donald Trump has regained the lead in Iowa but things can still change. On the Democratic side, young voters could tip the caucus toward Bernie Sanders, but only if they turn out.

Here's Bernie Sanders's best closing argument against Hillary Clinton in Iowa Washington Post

Bernie Sanders' One Answer on How He Would Get Anything Done ABC News

Related
Hillary Rodham Clinton »
Bernie Sanders »

Opinion: Democratic Iowa Forum: How to Watch the Live Stream Online Daily Beast

Wikipedia: Statewide opinion polling for the Democratic Party presidential primaries, 2016

CNN

Kansas City Star

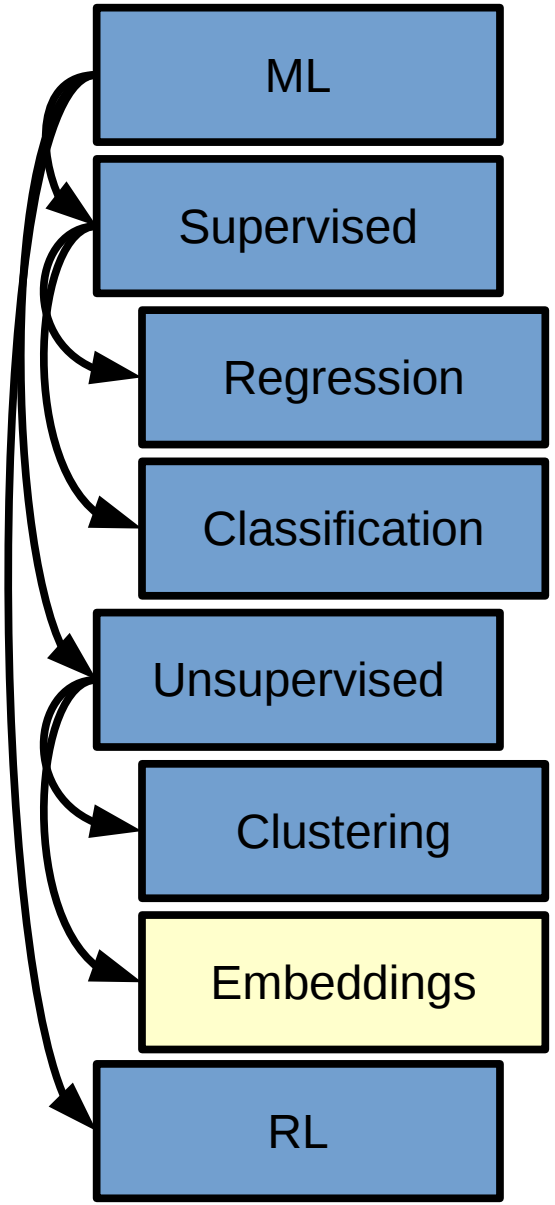
CNN

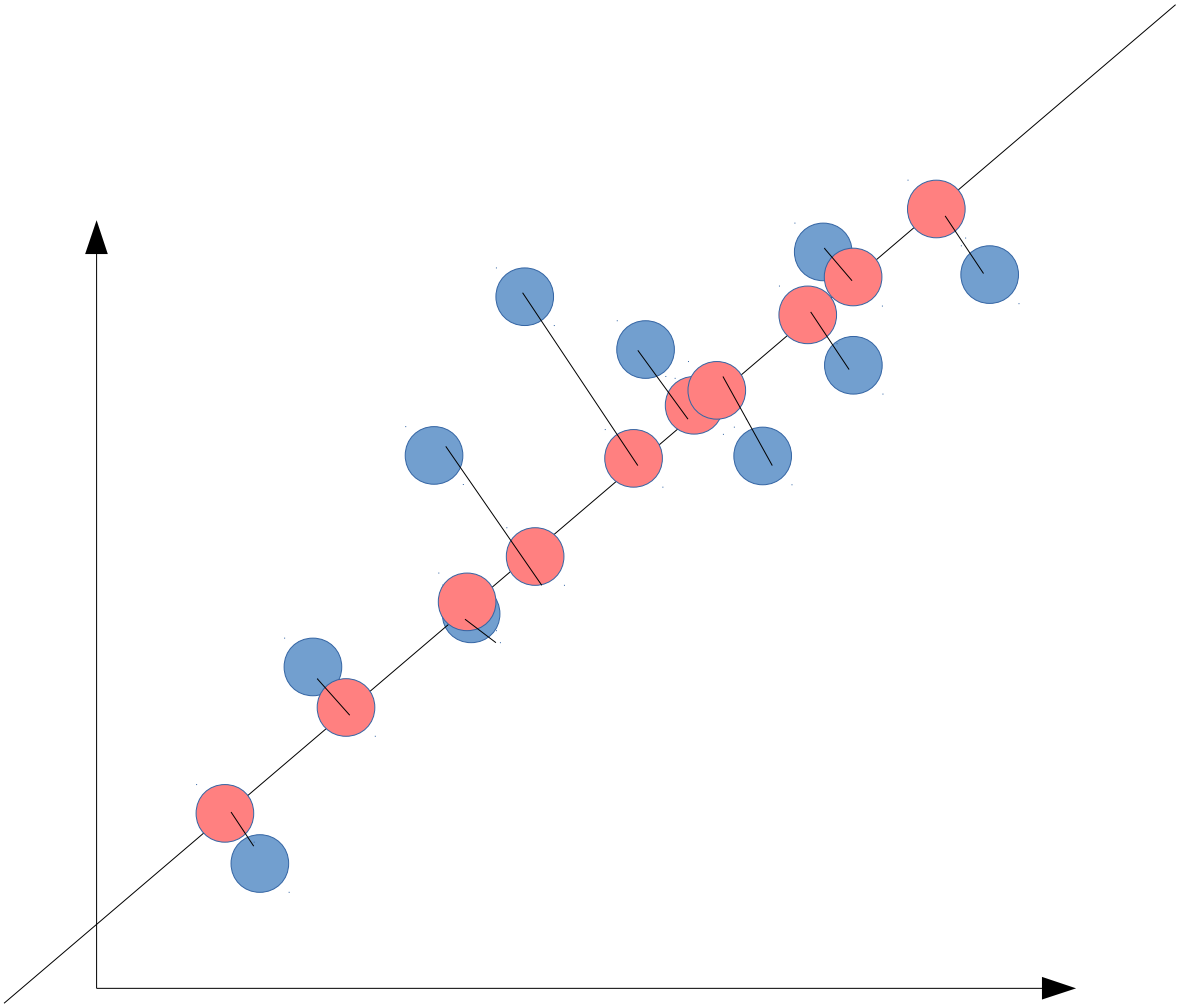
Reuters

Washingto...

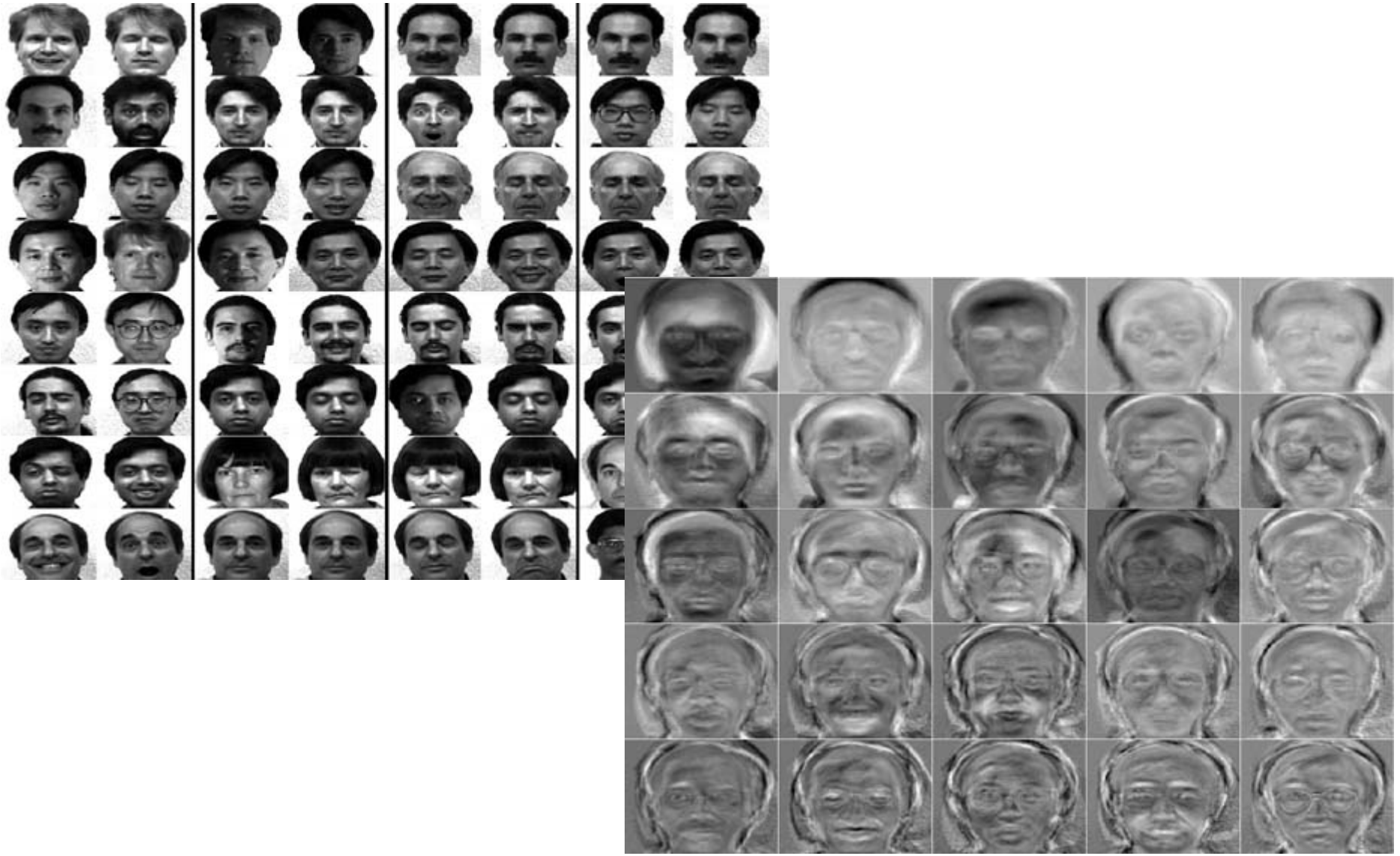
New Yc

Terminology: Embedding



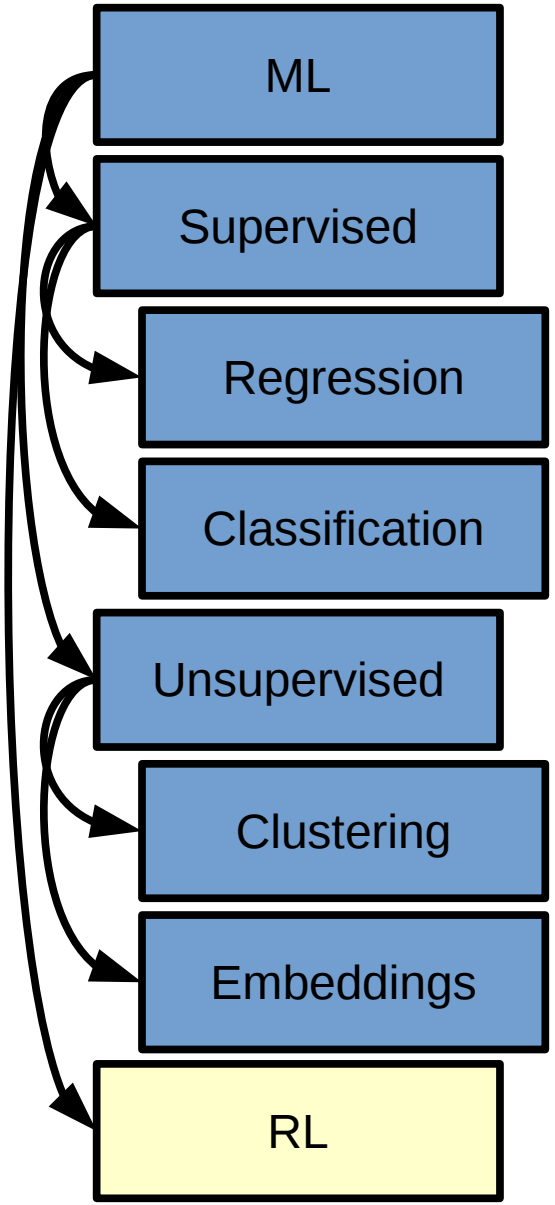


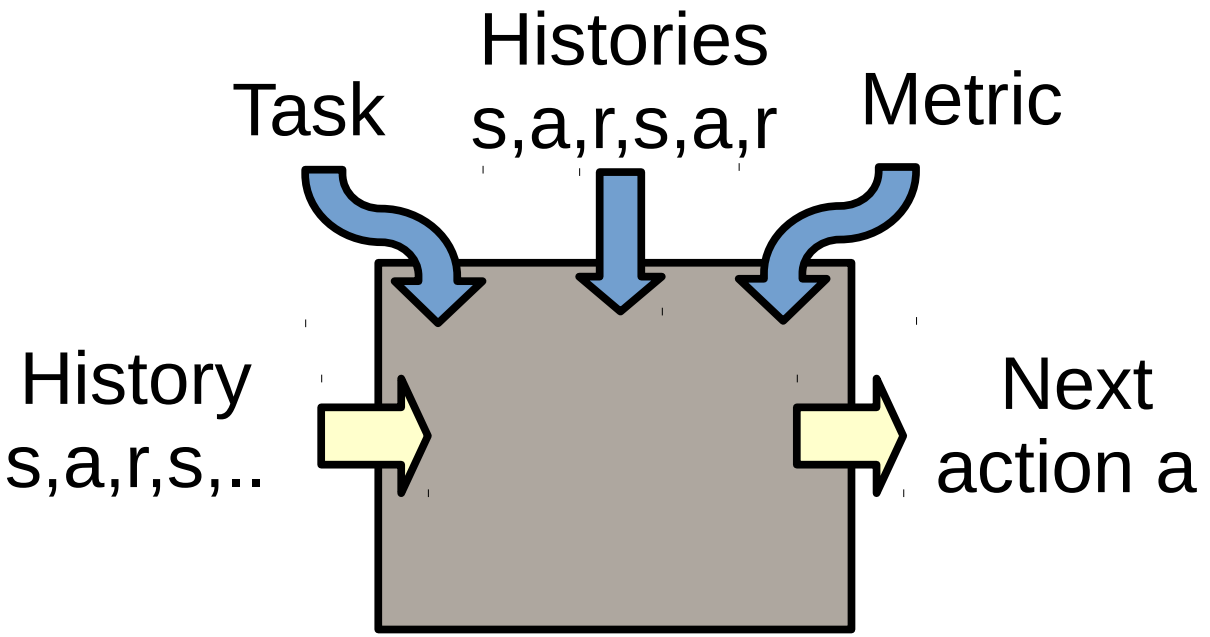
Example: Eigenfaces



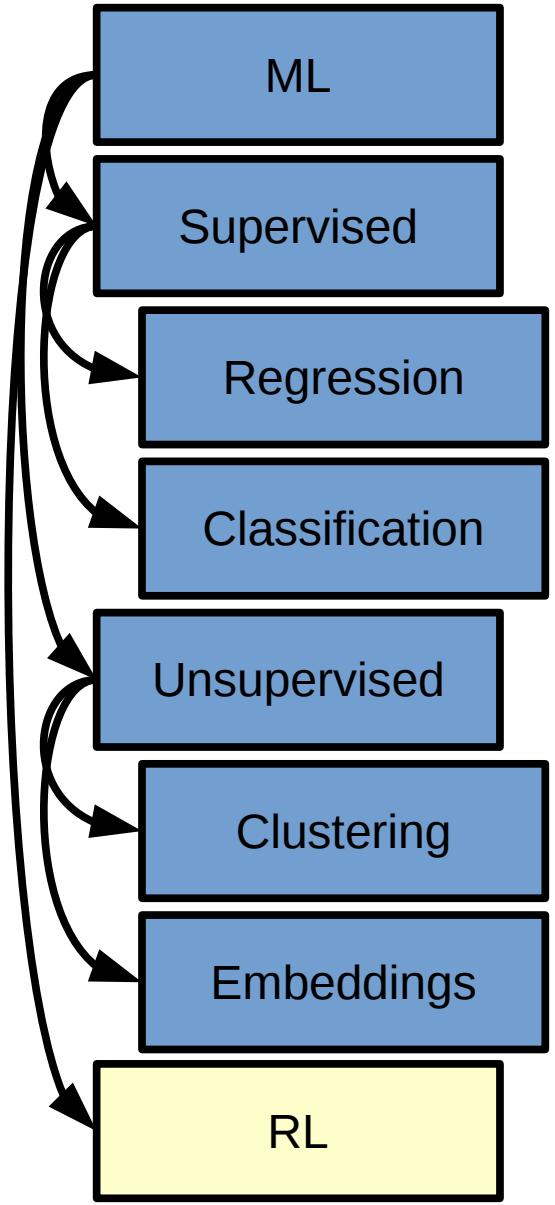
https://www.cs.princeton.edu/~cdecoro/eigenfaces/

Terminology: Reinforcement Learning





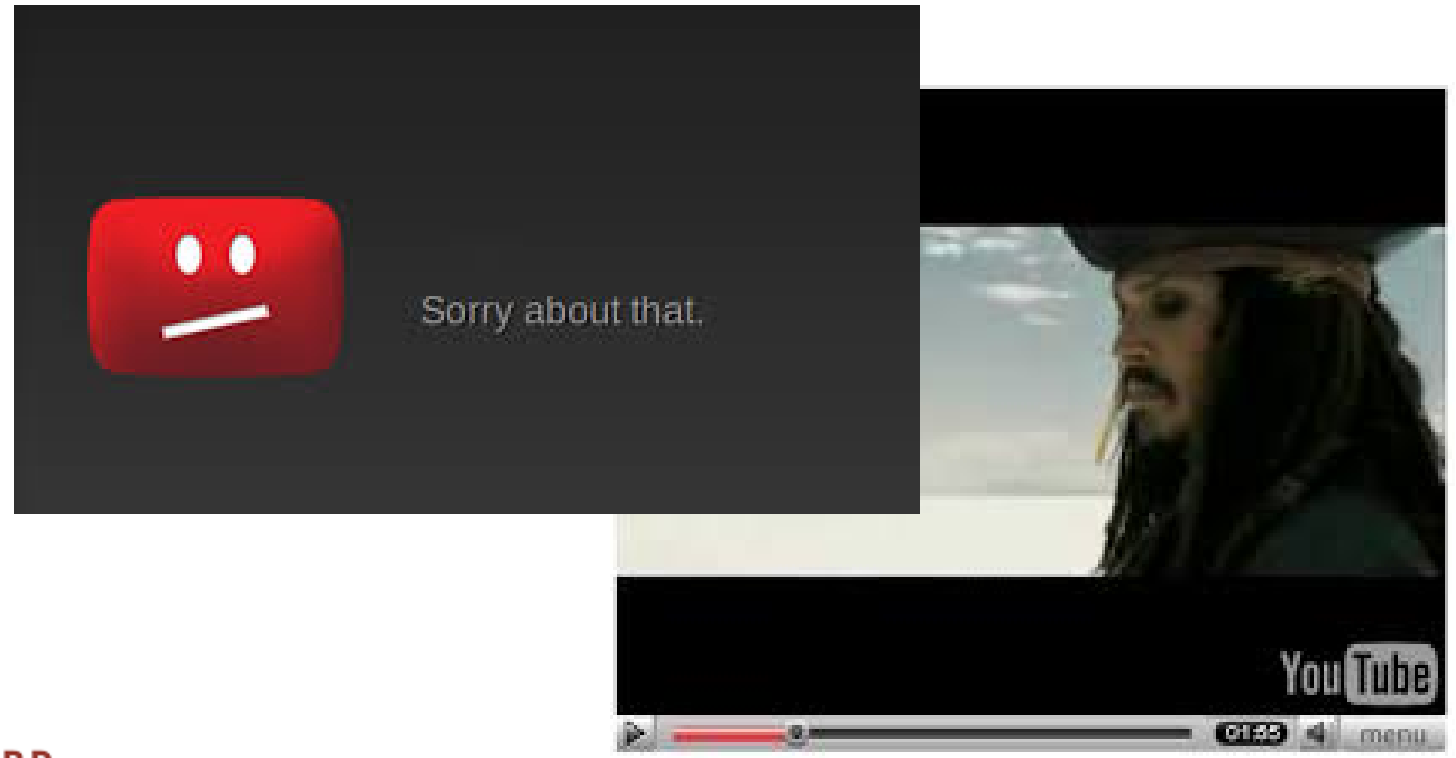
Terminology: Reinforcement Learning



Peter Kormushev, Imperial College

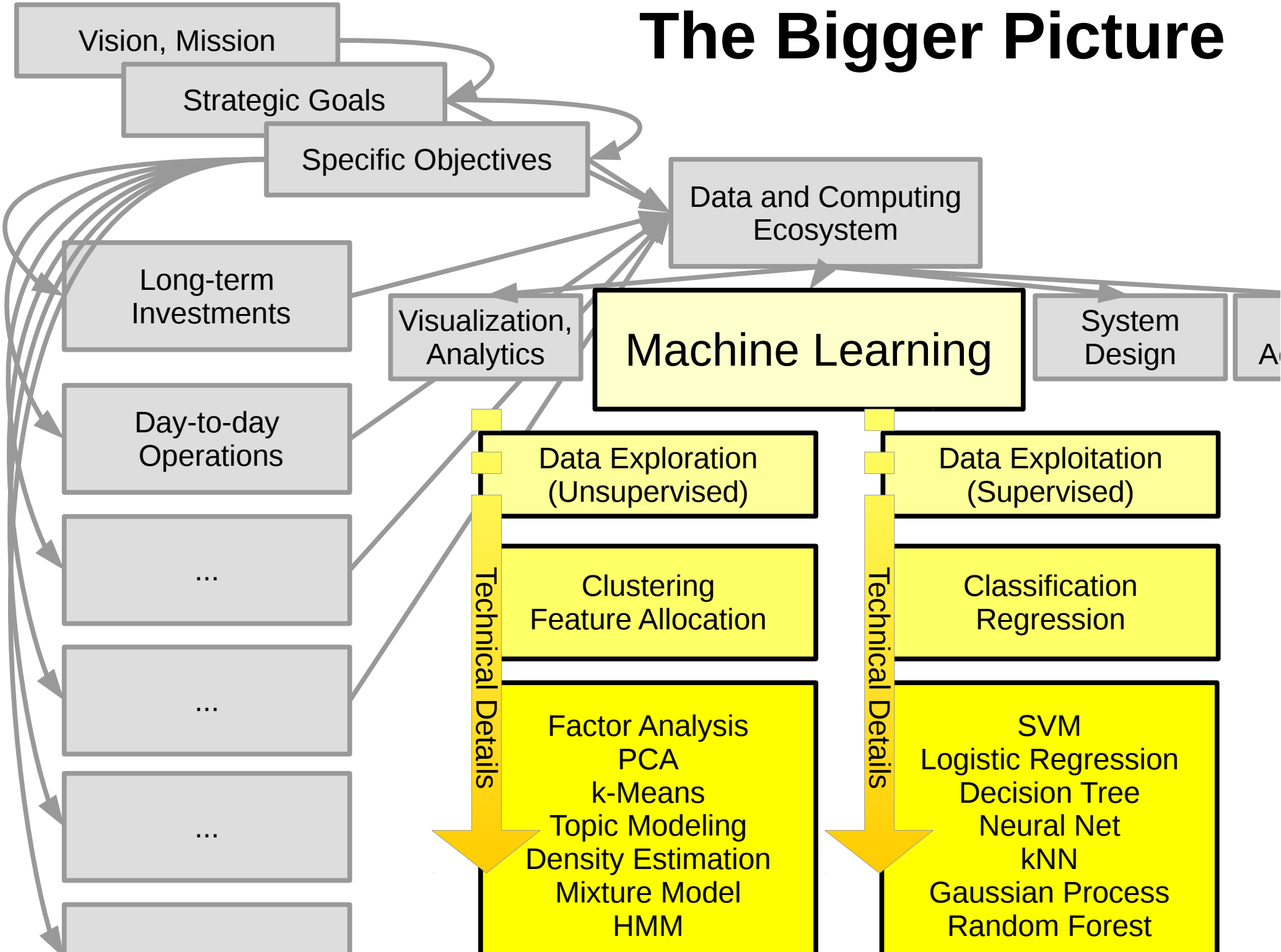
Your Turn: Detecting Copyright Violations on Youtube...

(~300 hours of new video per minute)



HARVARD
School of Engineering and Applied Sciences

The Bigger Picture



The Bigger Picture

Vision, Mission

Strategic Goals

Specific Objectives

Data and Computing Ecosystem

Long-term Inve

System Design

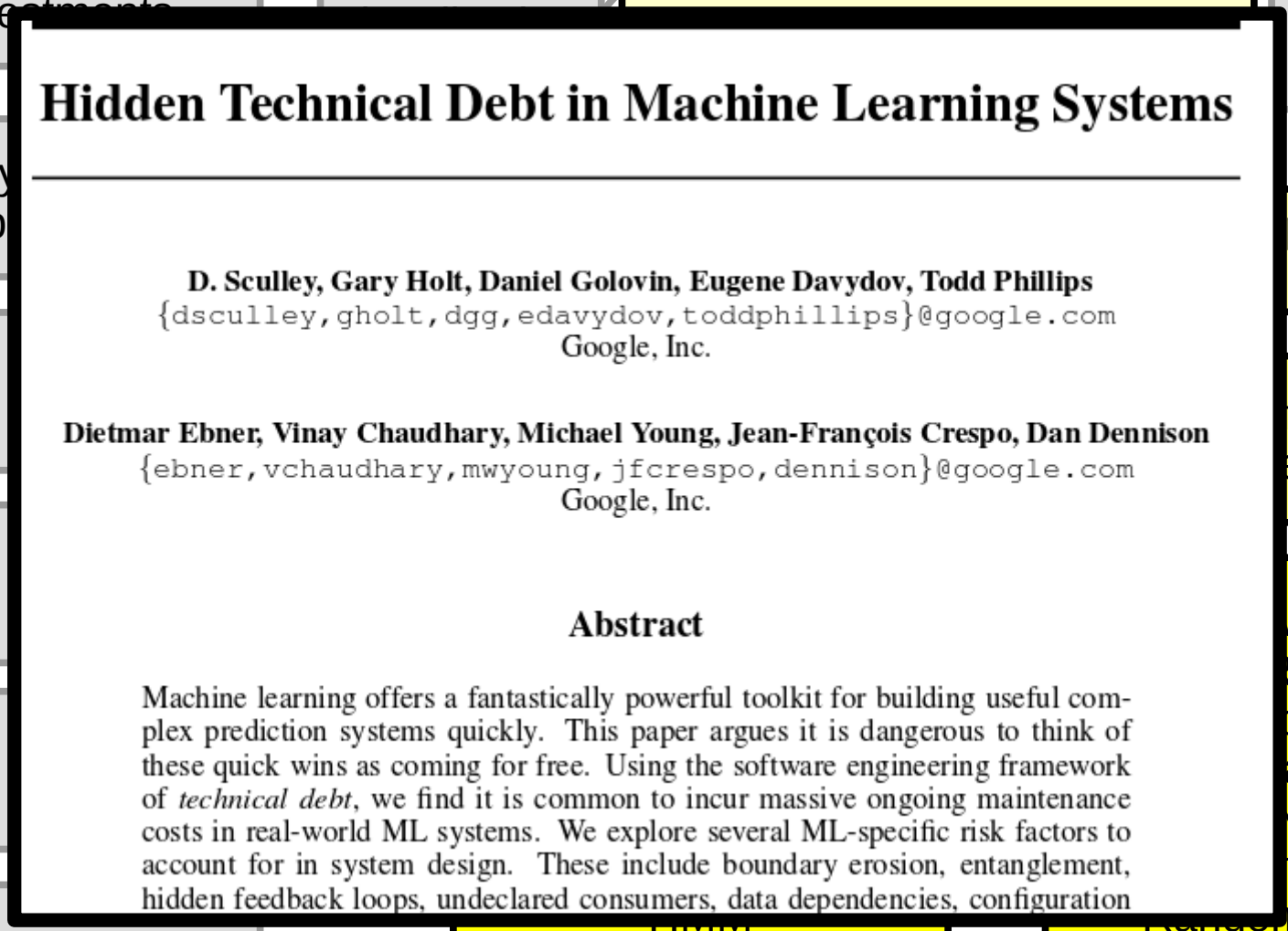
A

Day
Op

ploitation
rvised)

fication
ession

VM
Regression
on Tree
al Net
NN
n Process
n Forest



Hidden Technical Debt in Machine Learning Systems

D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips
{dsculley,gholt,dgg,edavydov,toddphillips}@google.com
Google, Inc.

Dietmar Ebner, Vinay Chaudhary, Michael Young, Jean-François Crespo, Dan Dennison
{ebner,vchaudhary,mwyoung,jfcrespo,dennison}@google.com
Google, Inc.

Abstract

Machine learning offers a fantastically powerful toolkit for building useful complex prediction systems quickly. This paper argues it is dangerous to think of these quick wins as coming for free. Using the software engineering framework of *technical debt*, we find it is common to incur massive ongoing maintenance costs in real-world ML systems. We explore several ML-specific risk factors to account for in system design. These include boundary erosion, entanglement, hidden feedback loops, undeclared consumers, data dependencies, configuration

Even Bigger

https://static1.squarespace.com/static/528152c0e4b009d7ee781896/t/57a34bb5be6594bf8c897856/1470319621526/taking+photos+cell+phone

Structure of the Course

- Start with some basic regression problems.
- Discuss model selection and evaluation.
- Continue with the different kinds of problems:
  - Supervised: regression, classification
  - Unsupervised: classification, embeddings
  - Reinforcement Learning
- Technical material interspaced with realworld stories, concept exercises, and ethics lecture.

Logistics

- Homeworks: smaller exercises, toy problems
  - MUST be in LaTeX
- Practicals: real data, compete in teams
- Sections: math review, then flipped classroom
- Piazza: Clarification and Content Tags
- Canvas: Announcements, Gradebook
- Github: Homework, Code, LaTeX submissions

Office hours to be posted very soon!

https://harvard-ml-courses.github.io/cs181-web

FAQs

- Can I sit in/audit/take the course from some other school?
- Can I do simultaneous enrollment?
- Do I need all the prereqs?
  - Programming
  - Statistics
  - Calculus and linear algebra

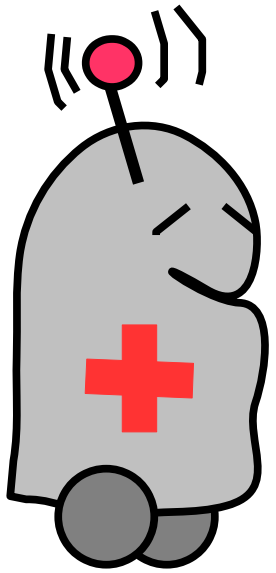
FAQs

- Can I sit in/audit/take the course from some other school?
- Can I do simultaneous enrollment?
- Do I need all the prereqs?
  - Programming
  - Statistics
  - Calculus and linear algebra

Problem 1.1:
Take the gradient with respect to the vector of weights *w*
...

Optimizing HIV Treatment

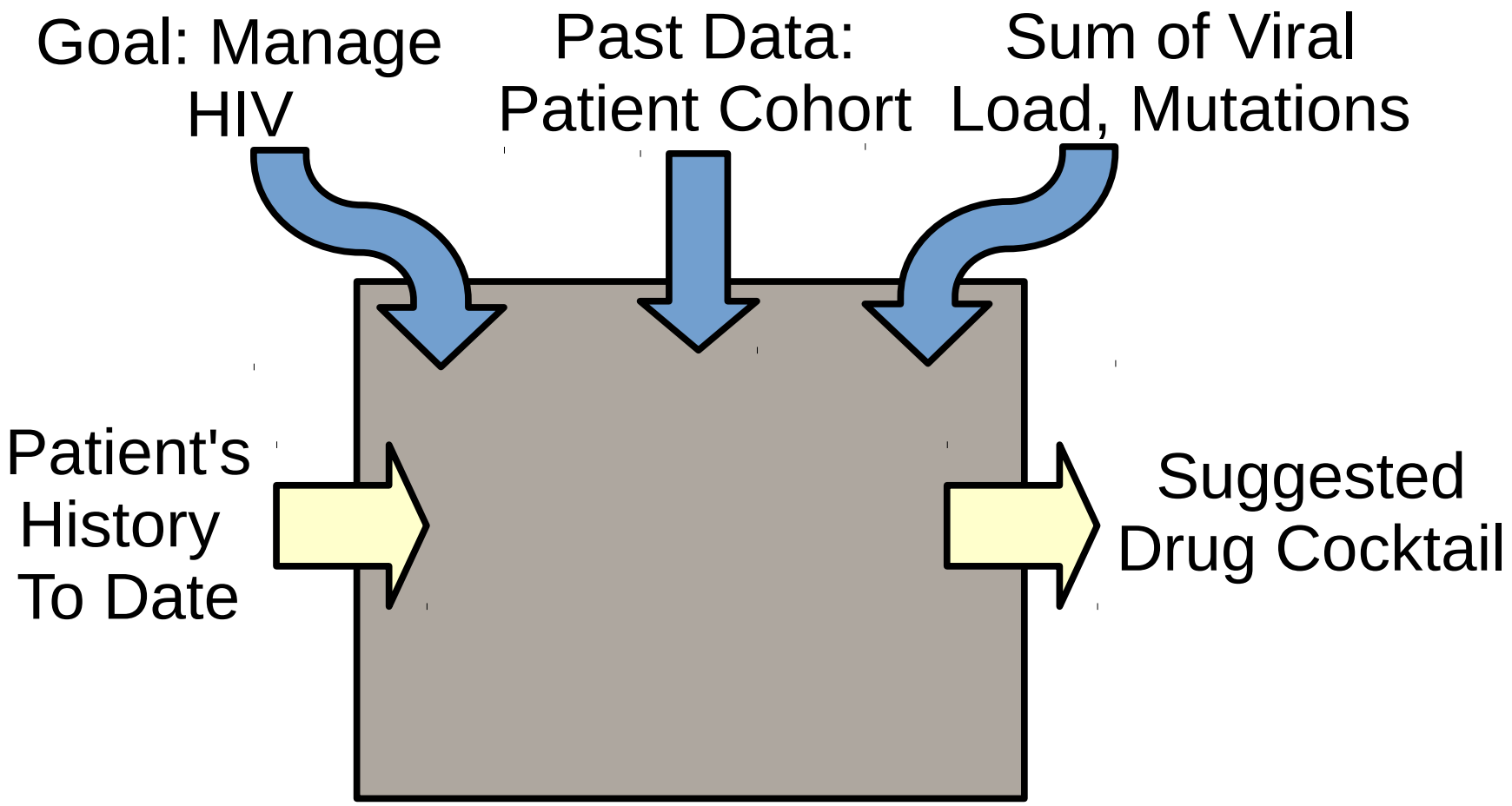
- Over 36 million people suffer from HIV worldwide.
- Requires life-long treatment with antiretrovirals.
- Rapidly mutating virus results in drug resistance.



We must reason about the

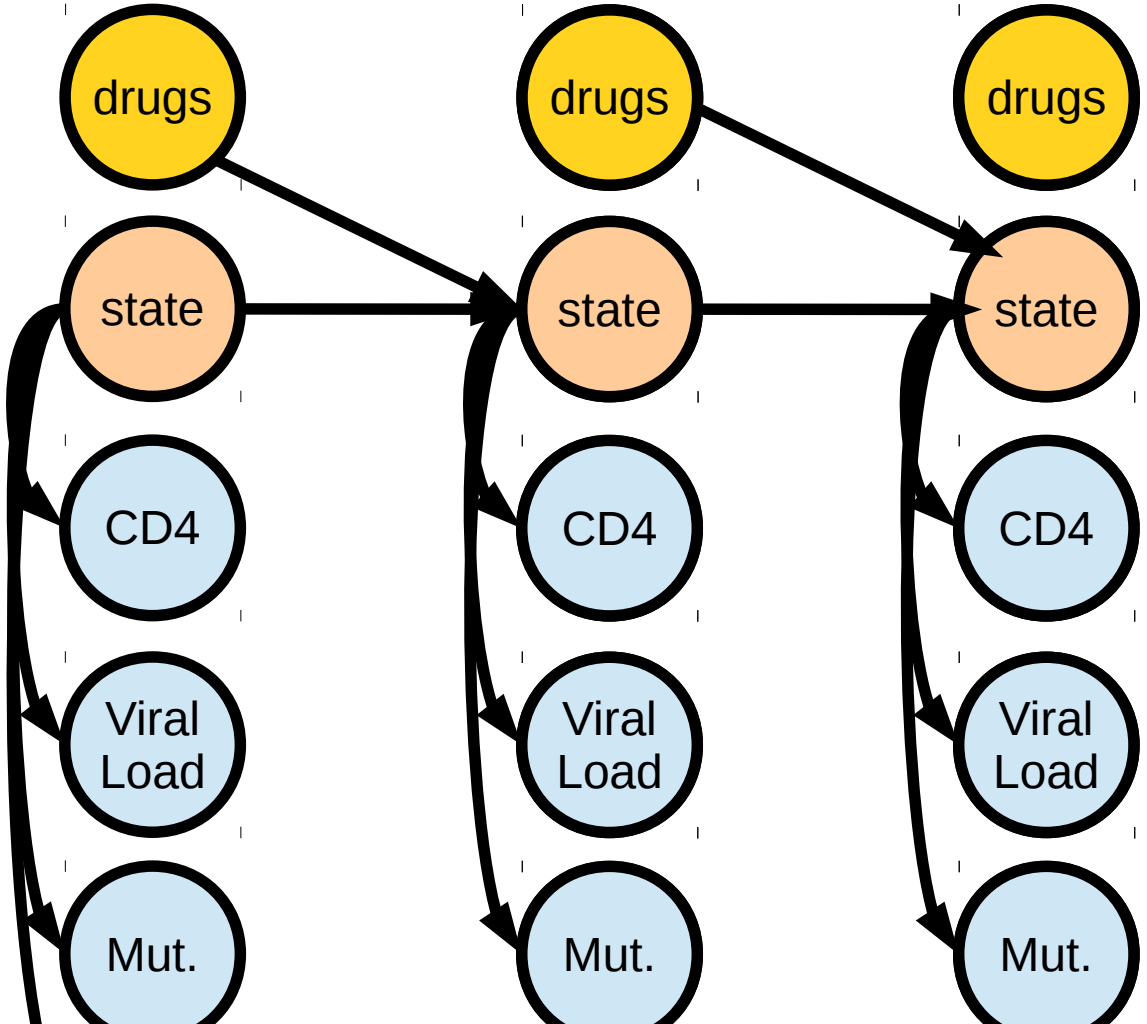
sequence of treatments.

In our framework:



Model-based Reinforcement Learning

Solve the long-term problem (e.g. Ernst 2005; Parbhoo 2014; Marivate 2015), but often in simulation/simplified settings.



Rewards:

If $V_t > 40$:

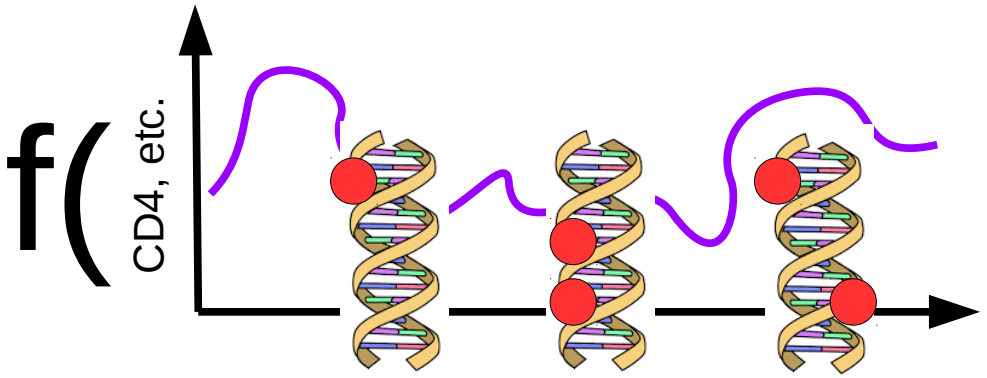
$r_t = -.7 \log V_t + .6 \log T_t - .2|M_t|$

Else:

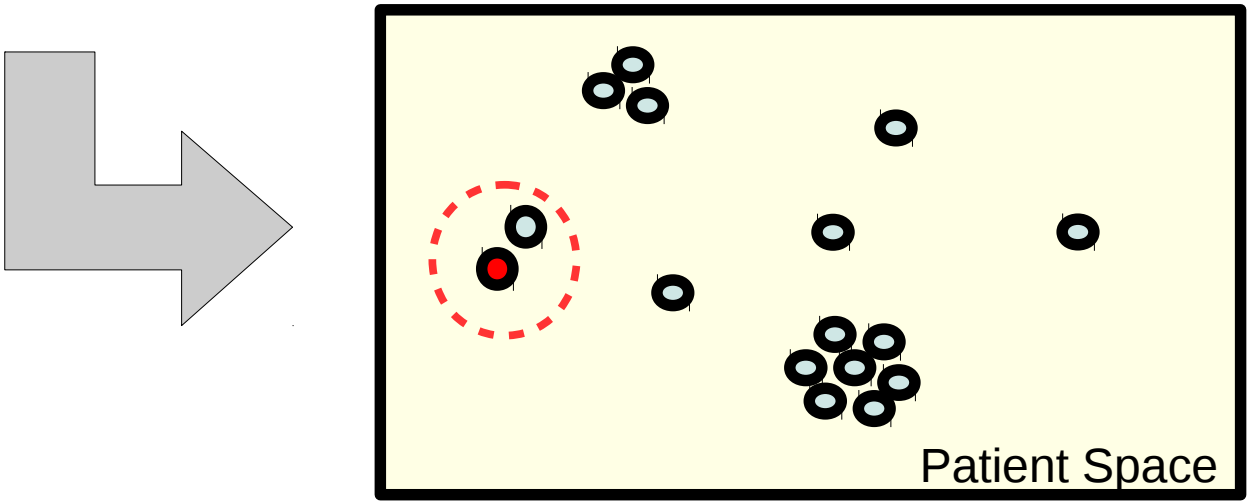
$r_t = 5 + .6 \log T_t - .2|M_t|$

Neighborhood-based Predictions

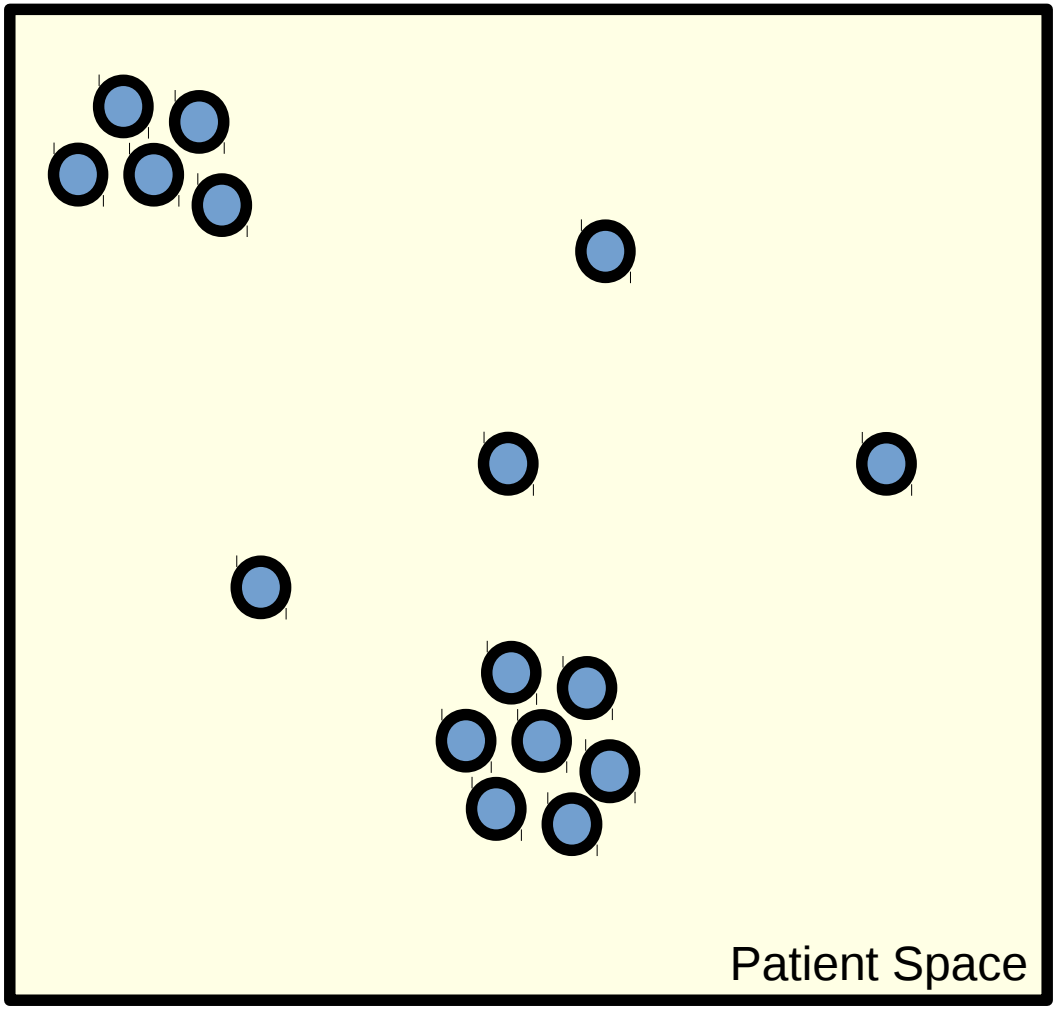
Use the full patient history to predict immediate outcomes (e.g. Bogojeska 2012), but often ignore long term effects.



, drug)= I(v<400)
in 21 days

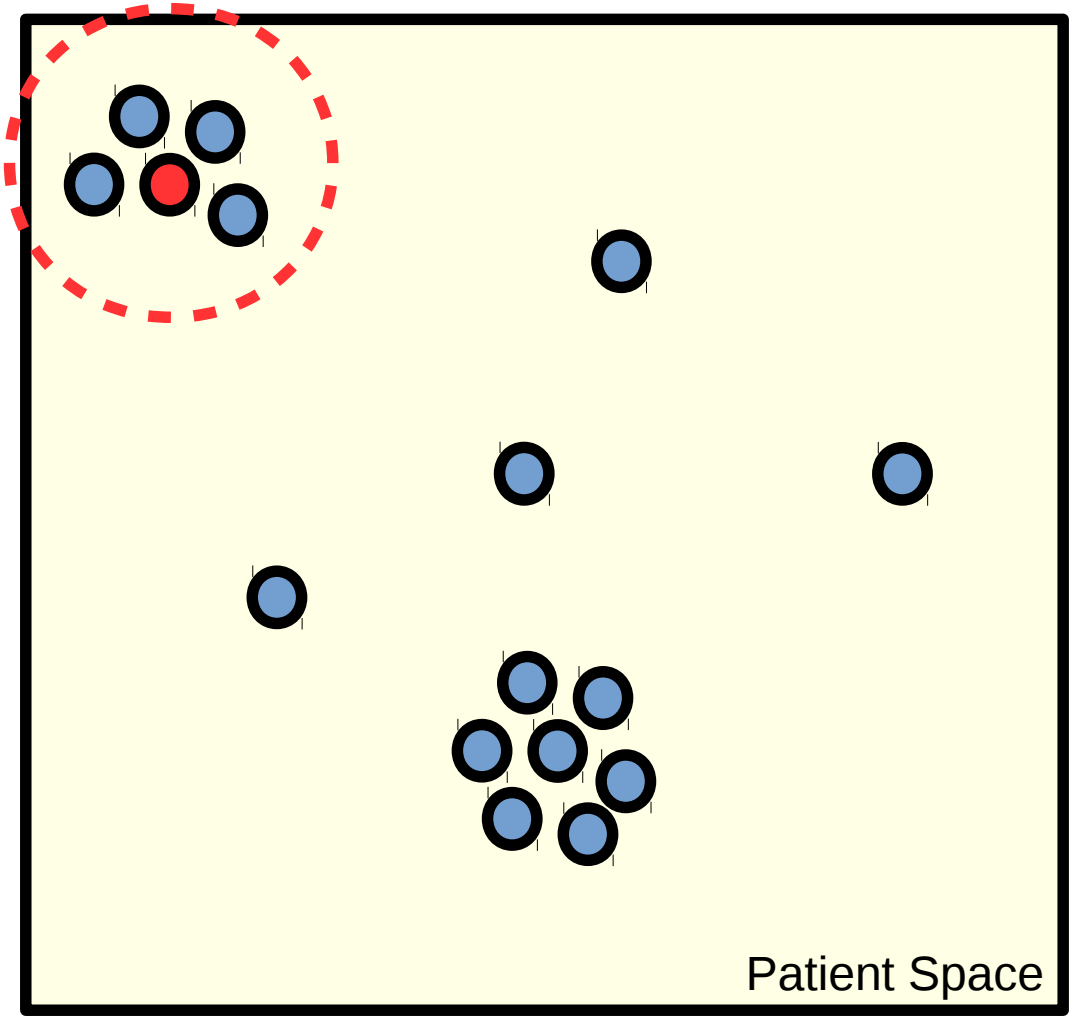


Models and neighborhoods have complementary strengths!

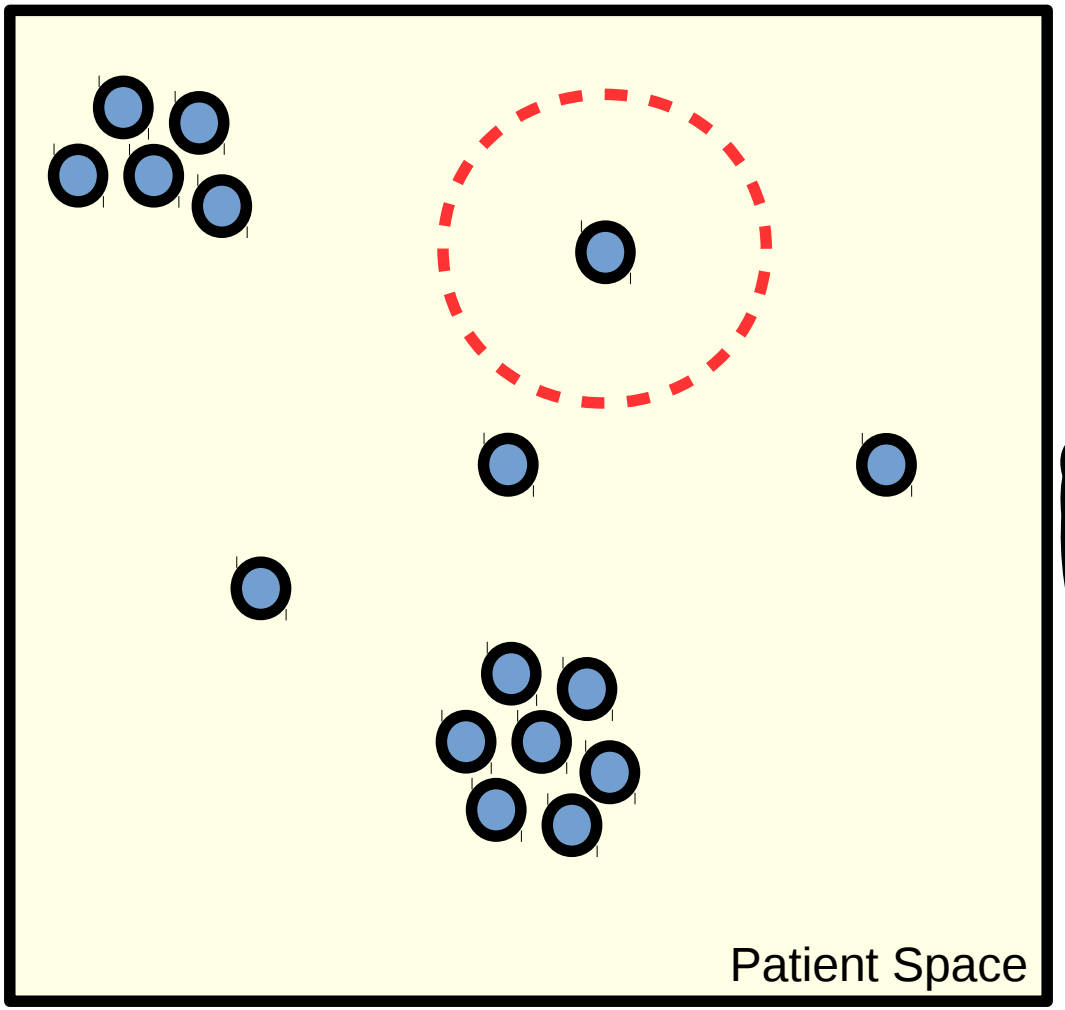


Models and neighborhoods have complementary strengths!

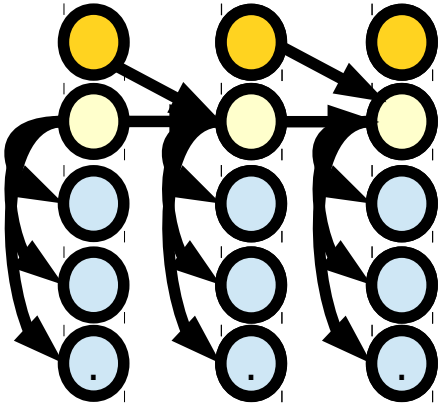
Patients in clusters may be best modeled by their neighbors



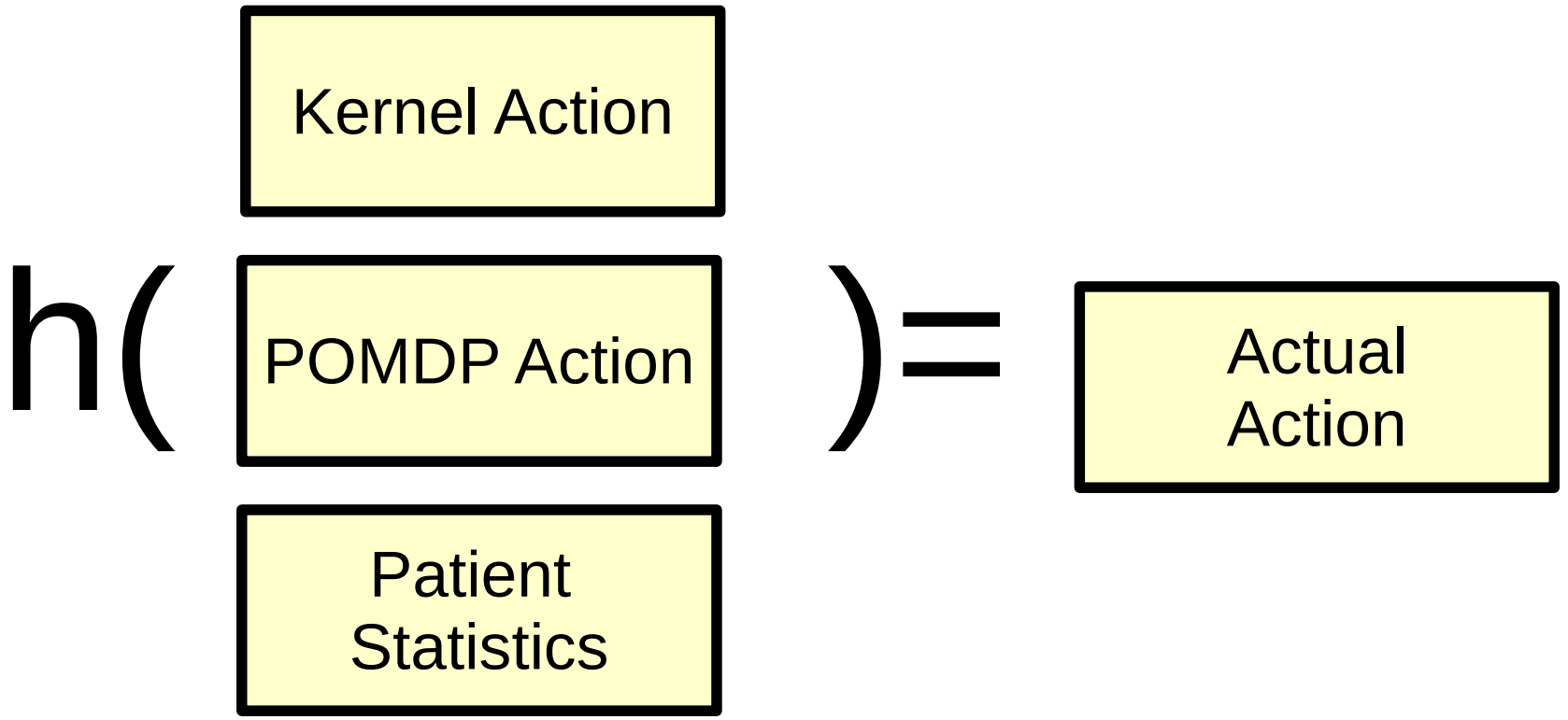
Models and neighborhoods have complementary strengths!



Patients without neighbors may be better modeled with a model



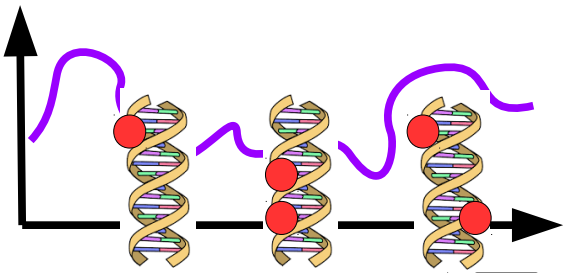
Key Idea: Combine!

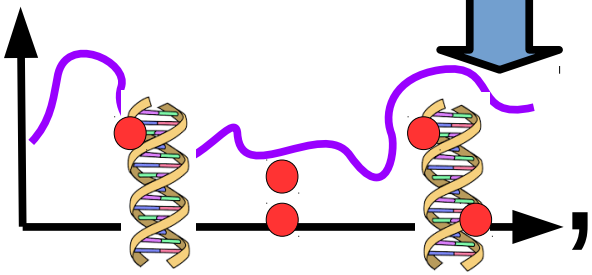


Key Idea: Combine!

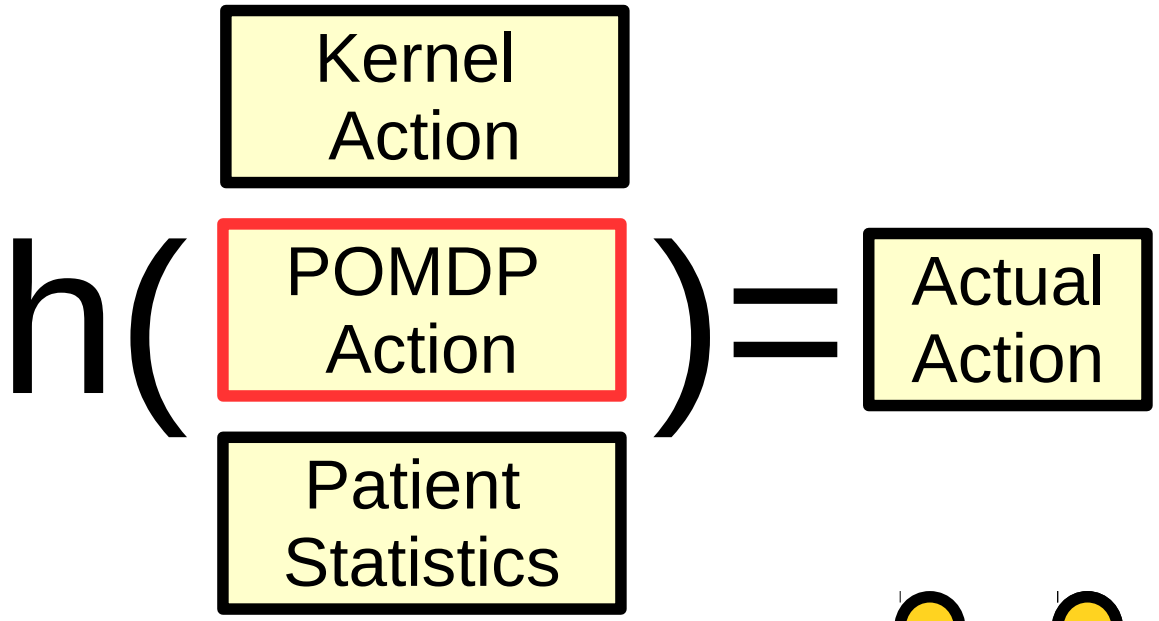
Kernel Action

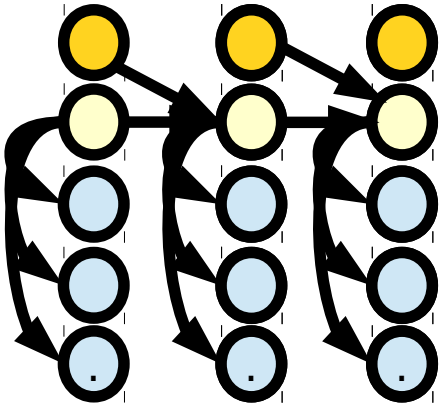
h(POMDP Action , Patient Statistics) = Actual Action

f( , drug) = v<400

g( , drug) = #mut +Σv

Key Idea: Combine!





Build POMDP and solve for a policy that minimizes viral load + mutations over 5 years.

Does it work?

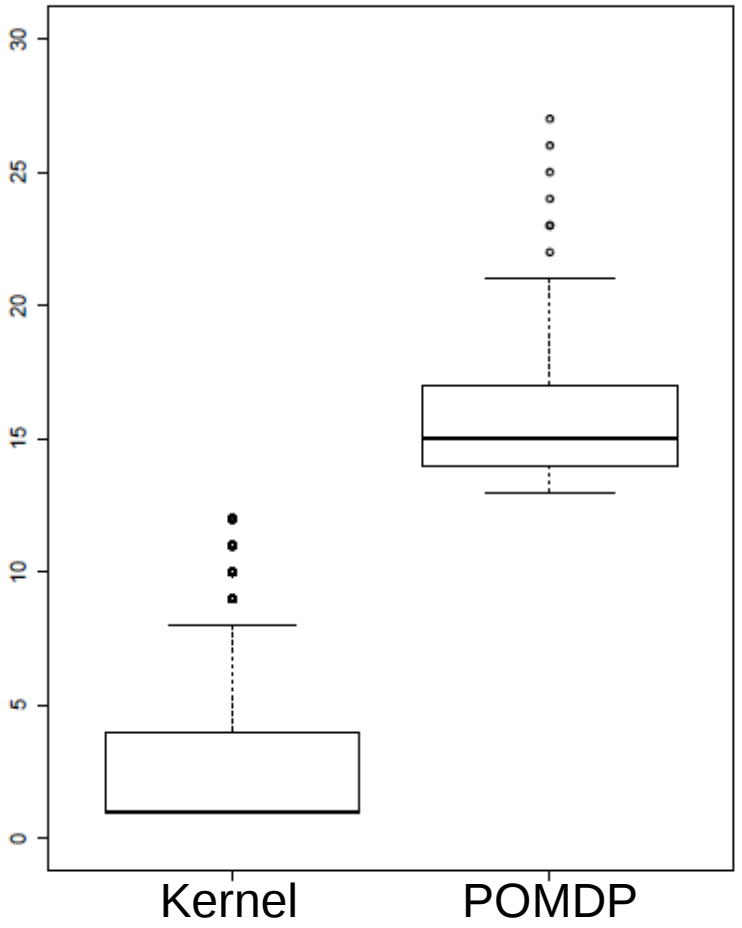
- 32,960 patients from EU Resist Database; hold out 3,000 for testing.
- Observations: CD4s, viral loads, mutations
- Actions: 312 common drug combinations (from 20 drugs)

Approach	DR Reward
Random Policy	-7.31 ± 3.72
Neighbor Policy	9.35 ± 2.61
Model-Based Policy	3.37 ± 2.15
Policy-Mixture Policy	11.52 ± 1.31

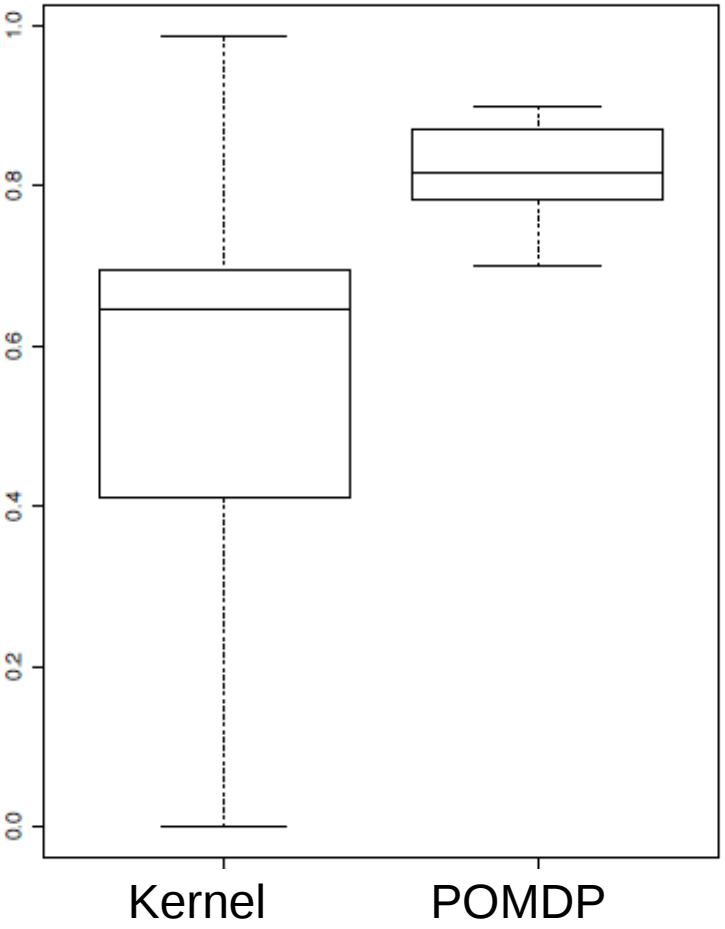
*Mixture chooses POMDP about 30% of the time.

Do choices make sense?

History Length

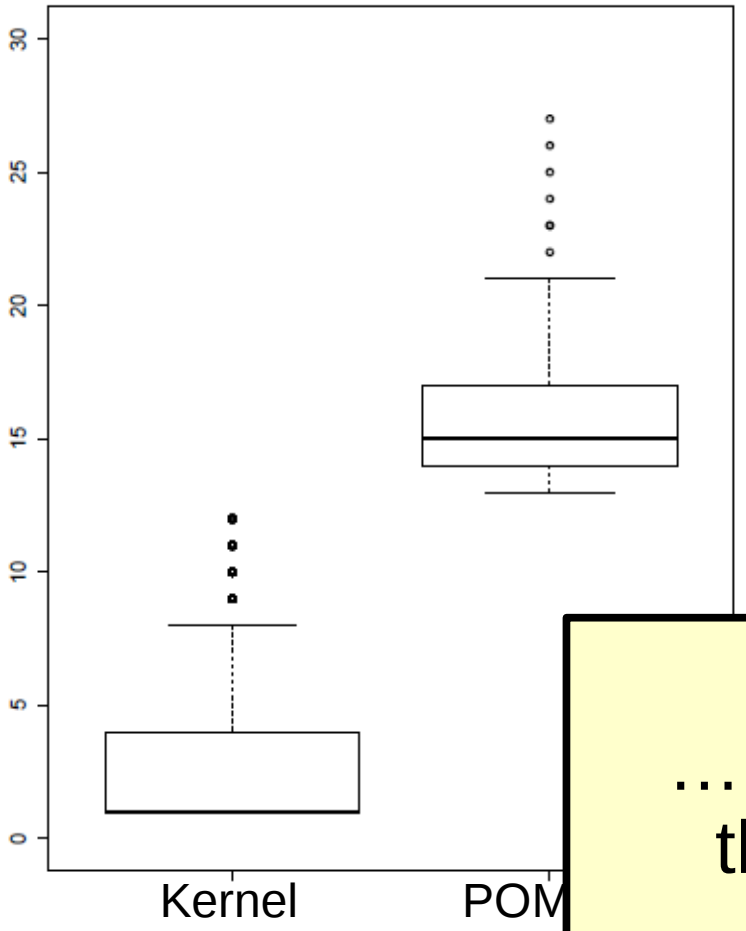


2nd Quantile Distance

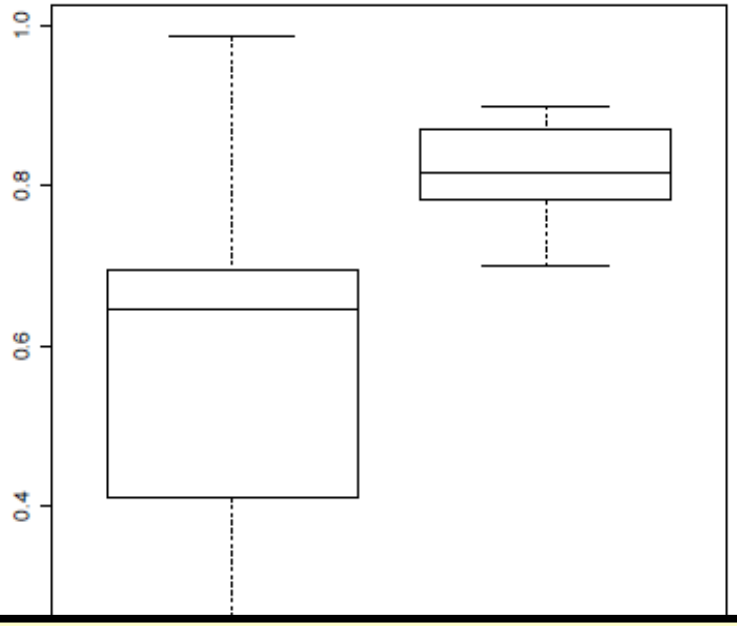


Do choices make sense?

History Length

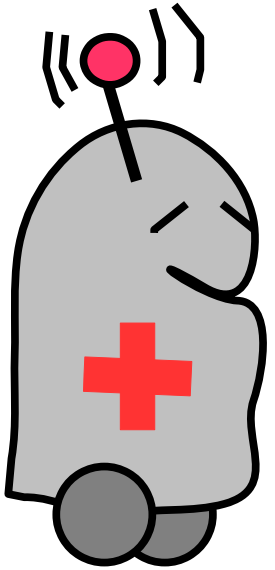


2nd Quantile Distance



… clinicians also agree that the policies make sense!

Where else can we use this idea?



Sepsis Management in the ICU

Sepsis Management in the ICU

Much harder!
We care about mortality, not just maintaining blood pressure… and it turns out, that's a lot trickier to evaluate!

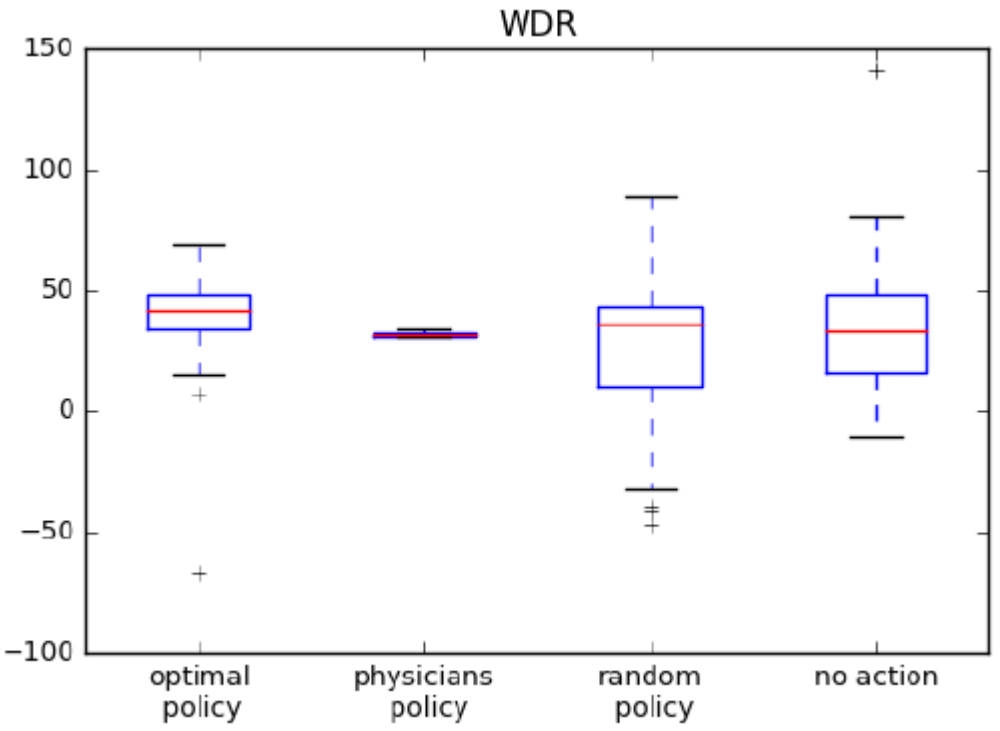
Evaluation Challenges

- Choosing actions to include: (When all you've got is a hammer…)
- Statistical methods have high variance
- Simpler evaluations can fail in unexpected ways

https://i0.wp.com/lipstickandlifesaving.com/wp-content/uploads/2017/03/img_9064.jpg

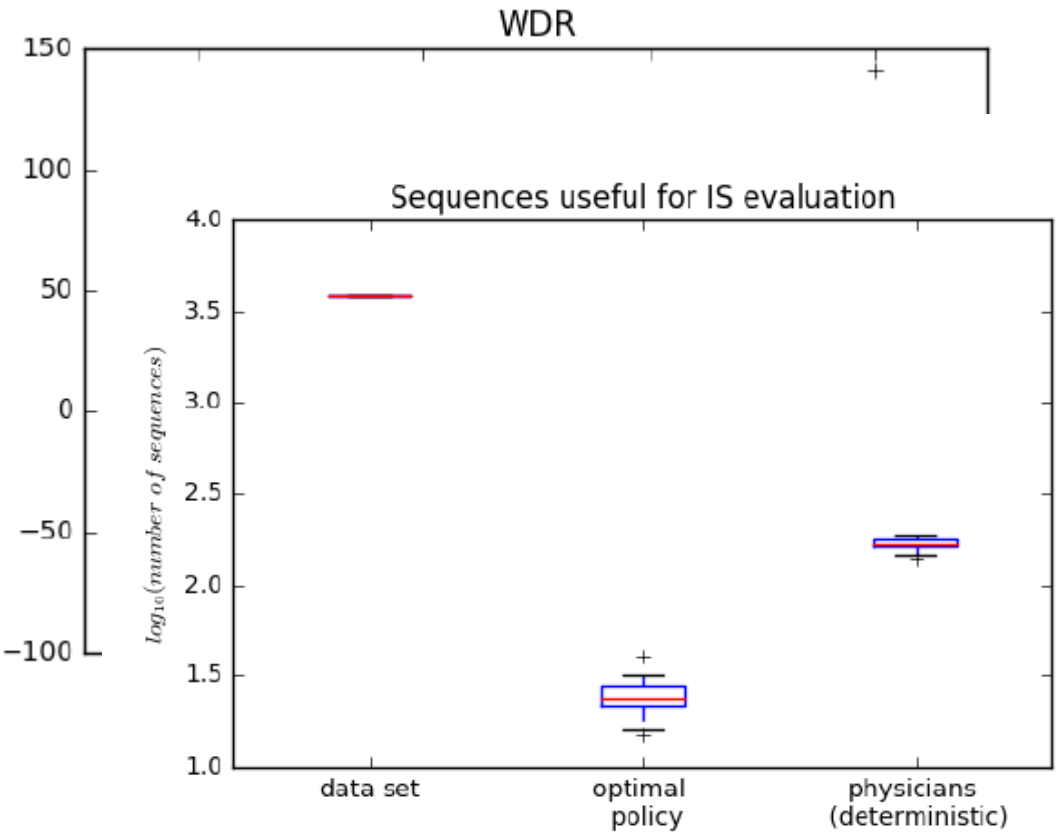
Evaluation Challenges

- Choosing actions to include: (When all you've got is a hammer…)
- Statistical methods have high variance
- Simpler evaluations can fail in unexpected ways



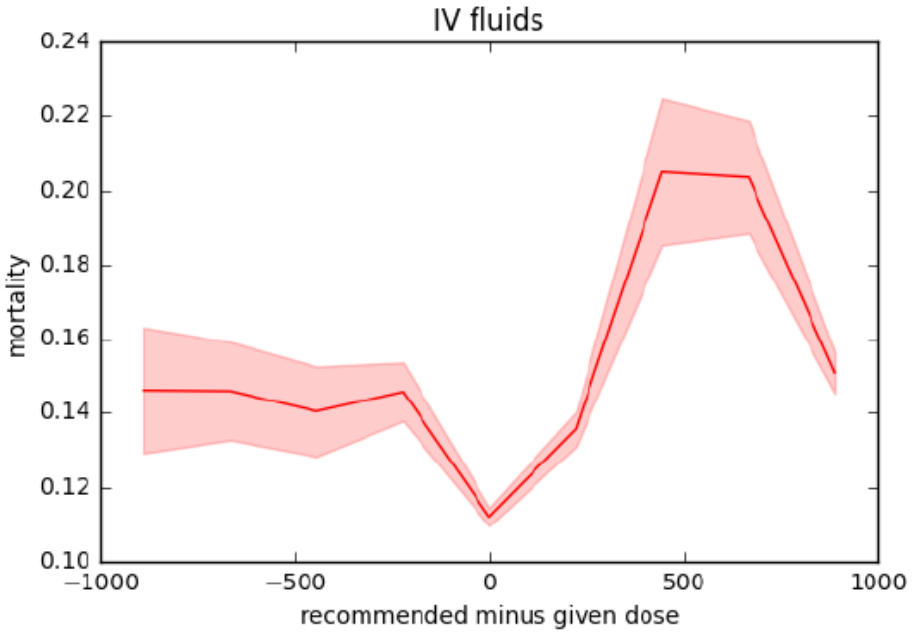
Evaluation Challenges

- Choosing actions to include: (When all you've got is a hammer…)
- Statistical methods have high variance
- Simpler evaluations can fail in unexpected ways



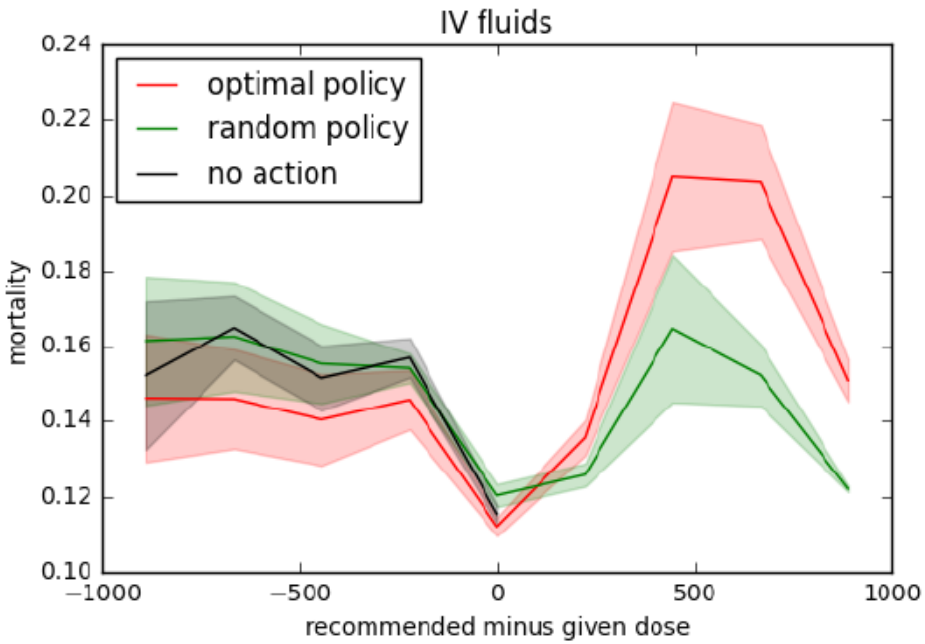
Evaluation Challenges

- Choosing actions to include: (When all you've got is a hammer…)
- Statistical methods have high variance
- Simpler evaluations can fail in unexpected ways

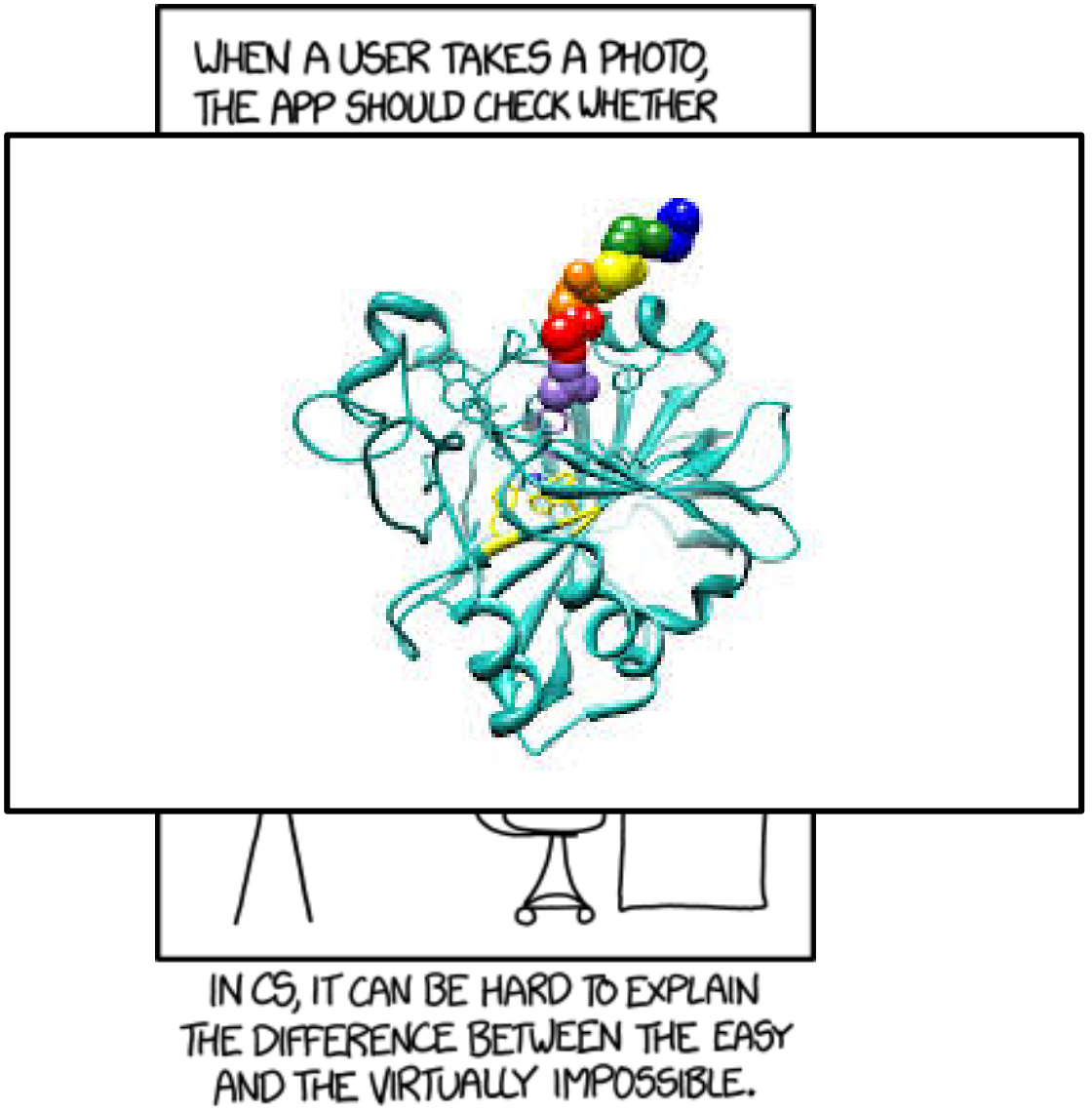


Evaluation Challenges

- Choosing actions to include: (When all you've got is a hammer…)
- Statistical methods have high variance
- Simpler evaluations can fail in unexpected ways



Small changes can make a problem a lot harder (or easier)

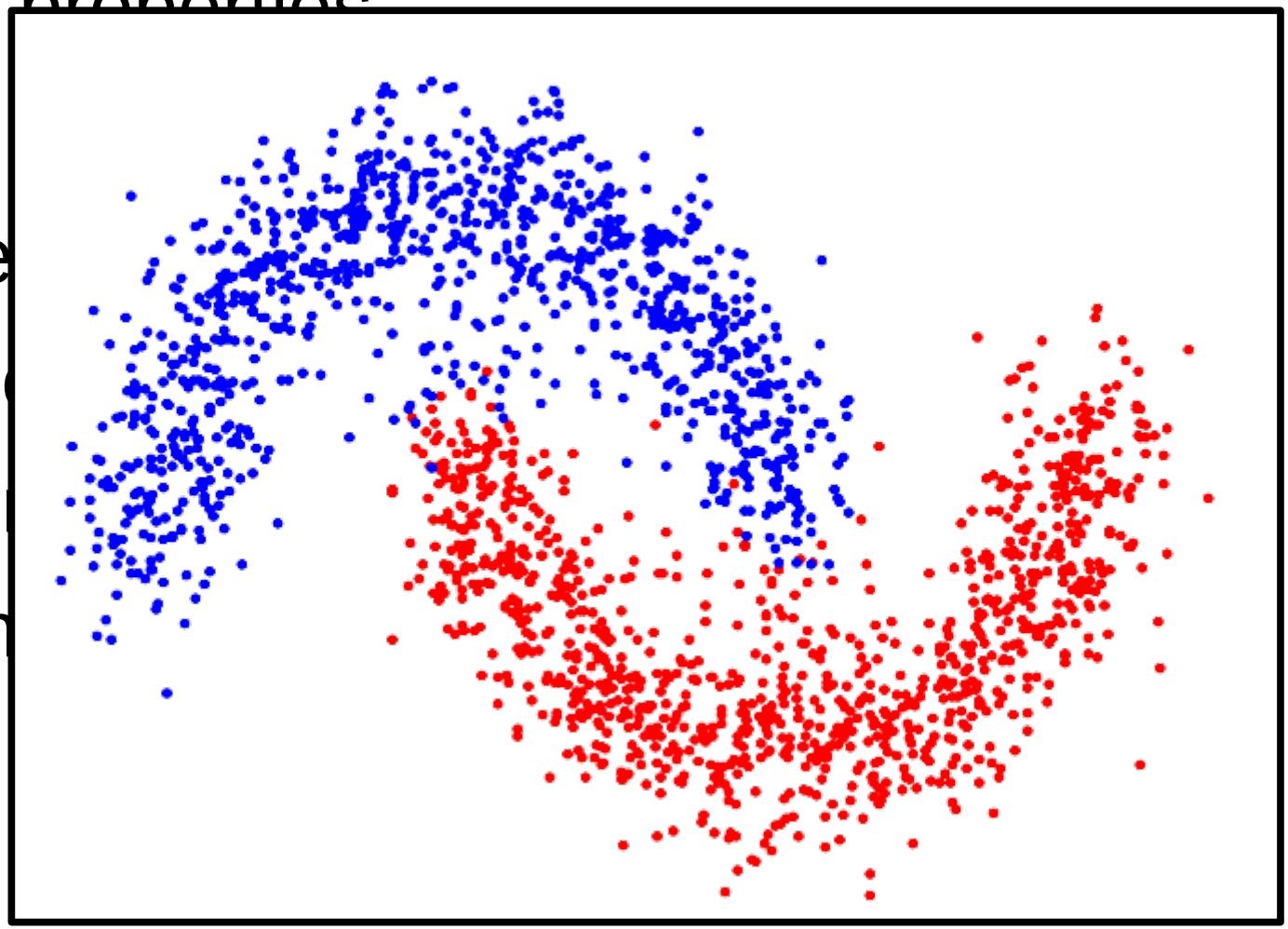


Hidden Assumptions

- Parameters, hyperparameters
- Distributional properties
- Cluster sizes and proportions
- Number of iterations
- Presence of local optima
- Approximations to distributions
- Conditional independence, Markovianity

Hidden Assumptions

- Parameters, hyperparameters
- Distributional properties
- Cluster sizes
- Number of ite
- Presence of l
- Approximatio
- Conditional in



Solution: Evaluate Carefully

- What data should be collected?
- How should it be processed?
- How should we separate a test set?
- What algorithms and parameters are used?
- How do we define success?

Fancy algorithms are no substitute for good science!!

Malaria (Bad Air) in Rome...