Introduction to Machine Learning in (Astro)Physics

Joshua Bloom
Astronomy

ML4FP School Lecture
Aug 12, 2024
Lawrence Berkeley National Laboratory

Agenda

- Introduction
- Brief Introduction to Machine Learning
- Examples of real-world ML uses
- Lessons Learned for Practical Machine Learning

Bio

Education

Harvard 92-96 AB
Cambridge 96-97 MPhil
Caltech 97-02 PhD

Work

Los Alamos 94, 95
Harvard Society of Fellows 02-05
UC Berkeley / LBNL
Asst. Prof 05-08
Assoc. Prof 08-12
Prof 12-
Chair 20-23
Wise.io CTO 13-16

Recognition

Hertz Fellow
Sloan Fellow
Pierce Prize American Astronomical Society
Gordon & Betty Moore Foundation Data-Drive Investigator
TwoSigma Fac. Fellow

Teaching

Fork me on GitHub

- **Python Bootcamps**
  200+ undergrad/grad yr^{-1}
- **Python for Data Science (AY 250); Data Lab (AY 128/256)**
- **Radiative processes, high-energy astrophysics**

Professor, UC Berkeley, LBNL

Research

- **Extragalactic Transients**
- **Automated Data-driven Discovery & Inference in the Time Domain**
- **300+ refereed articles**

Industry

- **ML Applications Company**

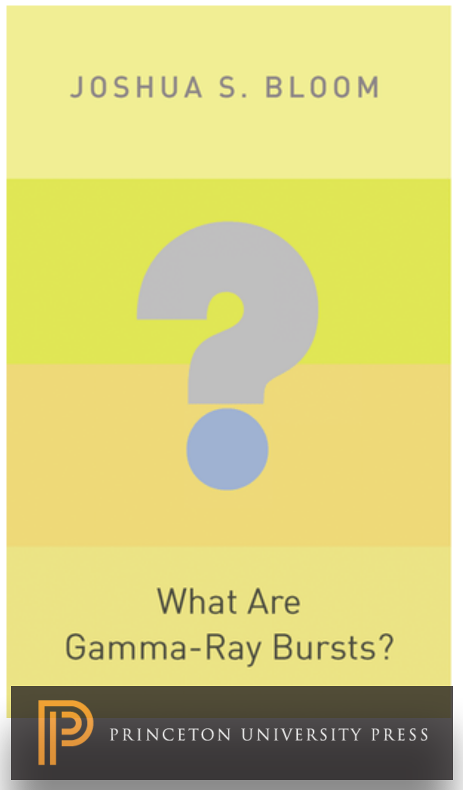
Wise.io

Q4'16

CTO, Co-founder

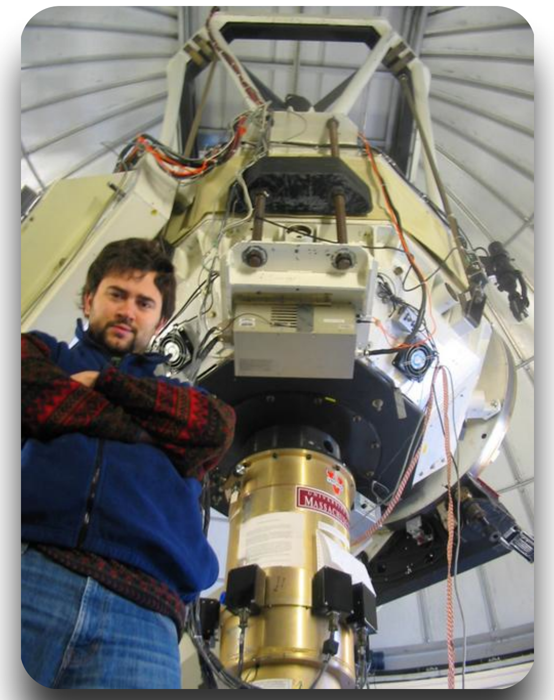
Research Themes in *Time-Domain* Astrophysics

(Explosive) Rarities & Extreme Objects



Origin & Nature of gamma-ray bursts & supernovae 1104.2274

Discoverer of relativistic tidal disruption around black holes 1104.3257



Intelligent Data Collection/Action Agents

PI of Peters Automated Infrared Imaging Telescope astro-ph/0511842

Co-PI of RATIR, Exec Committee LS4

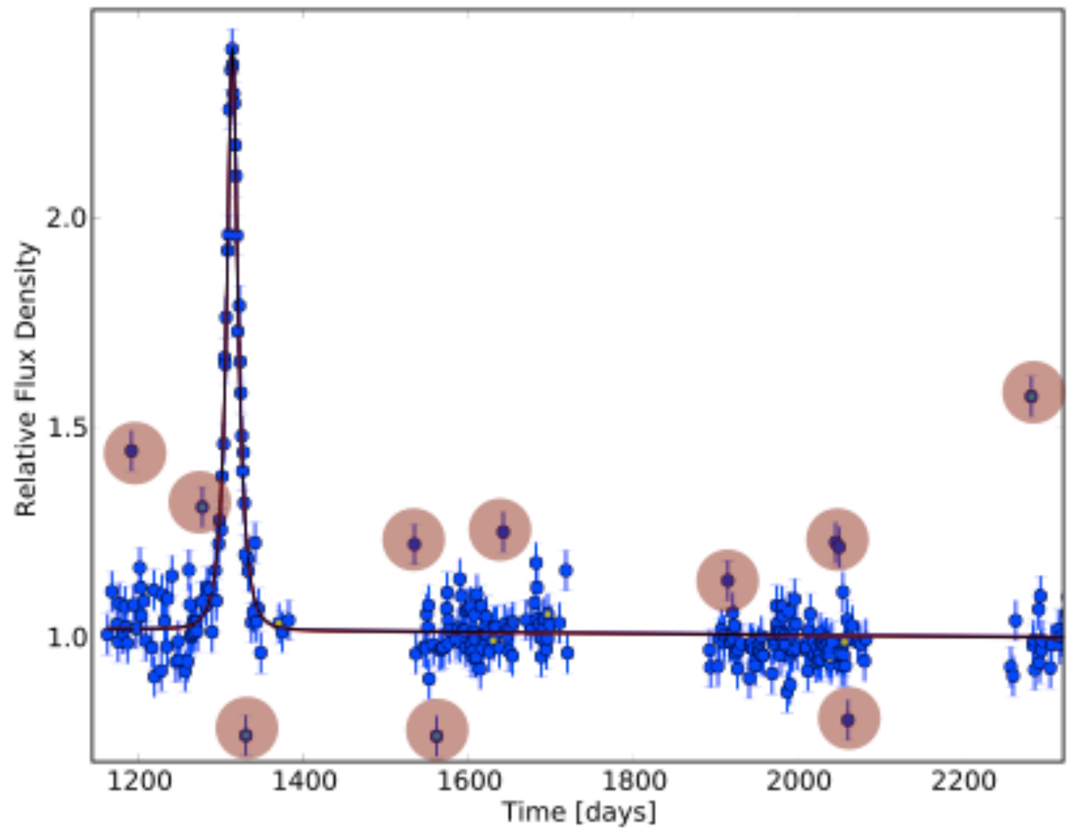
Creator of VOEvent/Net - messaging standard for transients http://voeventnet.caltech.edu/about/

Time-Domain Informatics

Novel Discovery & Inference Frameworks 1104.3142, 1106.5491

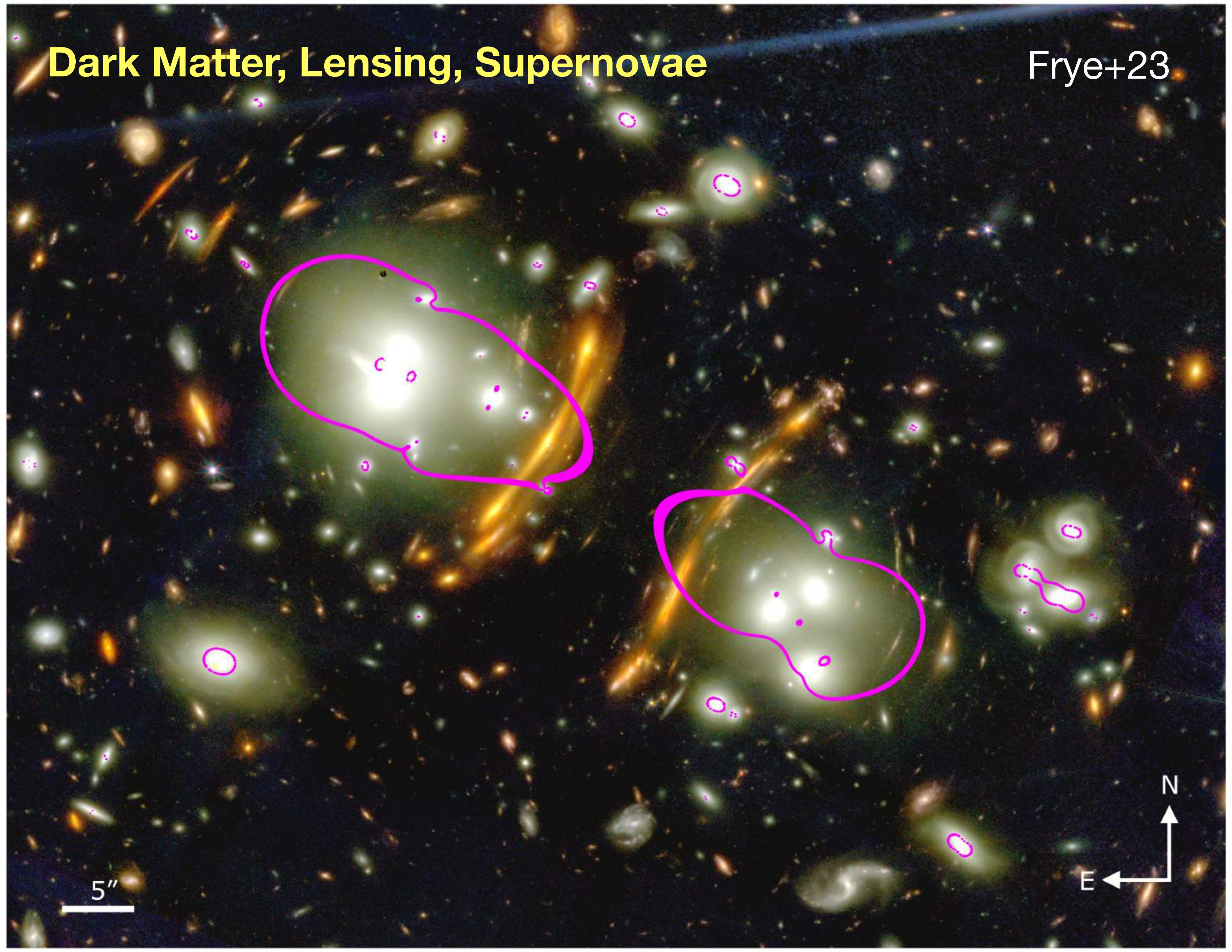
→ Precision near-field cosmology 1405.1035

PI Center for Time-Domain Informatics (Berkeley)



Dark Matter, Lensing, Supernovae

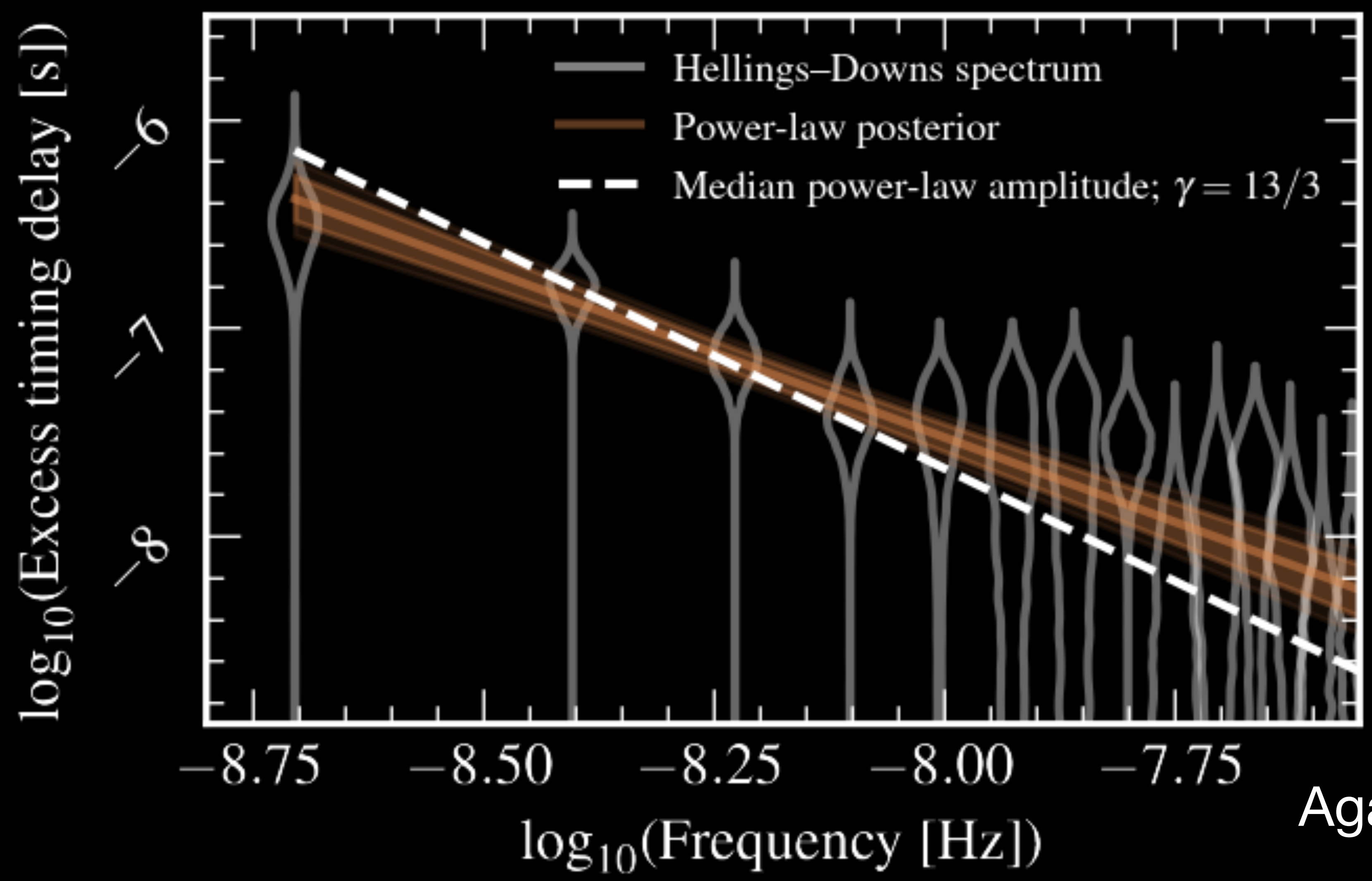
Frye+23



Black Hole Energetics, Tests of General Relativity

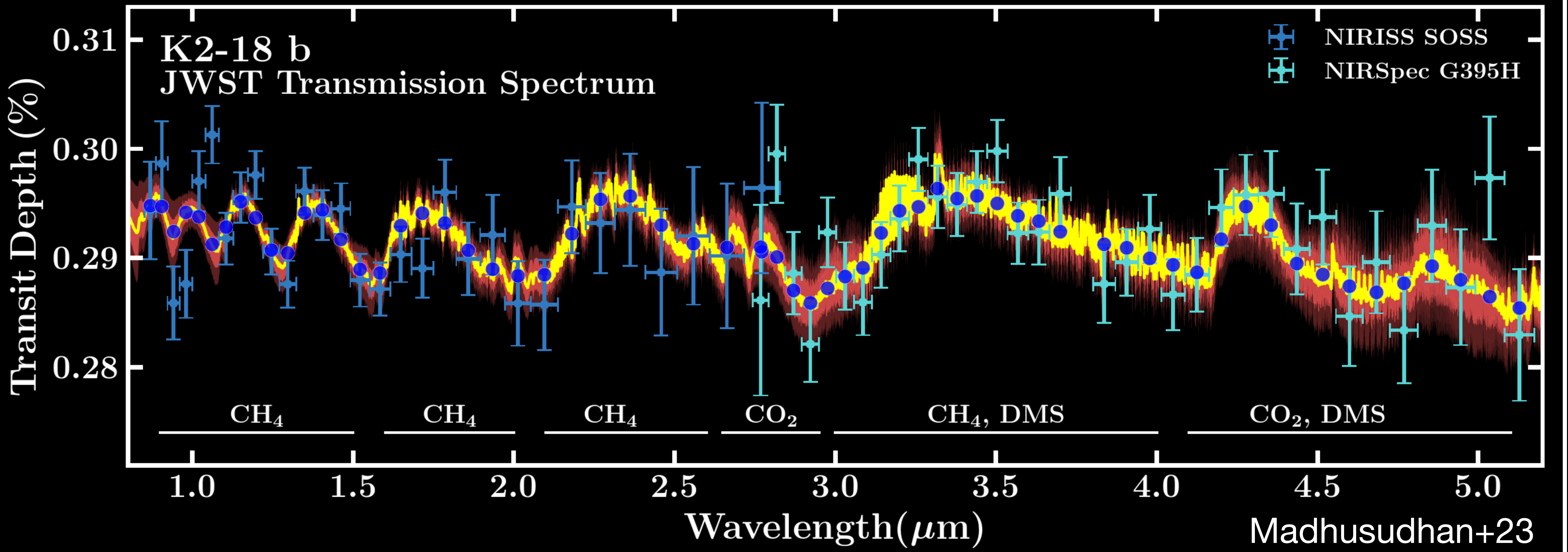
EHT, Akiyama+21

Gravitational Wave Background, Massive Binary Black Holes



Agazie+23

Exoplanet Atmospheres



Madhusudhan+23

Machine Learning:
A Brief Primer

Forms of Practical/Practiced Machine Learning Tasks

- **classification** - event discovery, sorting
- **regression** - weather forecasting, model based inference,
  time-series prediction
- **imputation** - data cleaning, inference of missing information
- **recommendation** - ranking, product recommendation,
  “Netflix prize”
- **clustering** - event segmentation, structure discovery
- **outlier detection** - anomaly identification, process control
- **dimensionality reduction** - visualization, manual insight
- **information retrieval** - search, indexing, document retreival
- **generative** - text-to-image, code creation
- **navigation & planning** - interacting systems (e.g., self-driving cars, robotics)

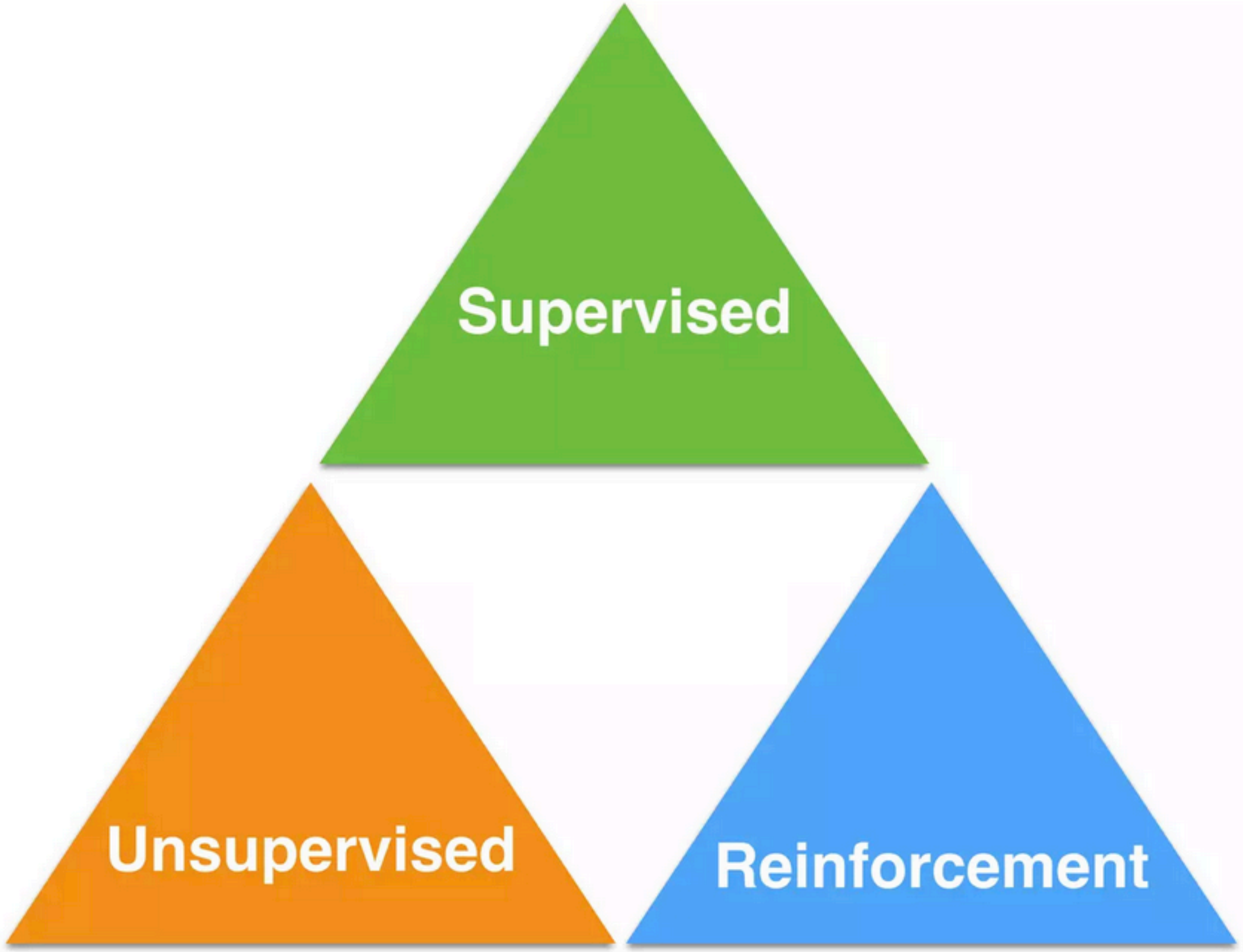
AI/ML Defination

“Field of study that give computers the ability to learn without being explicitly programmed.”

-Arthur Samuel, 1959

Machine Learning Approaches

- Labeled data
- Direct feedback
- Predict outcome/future



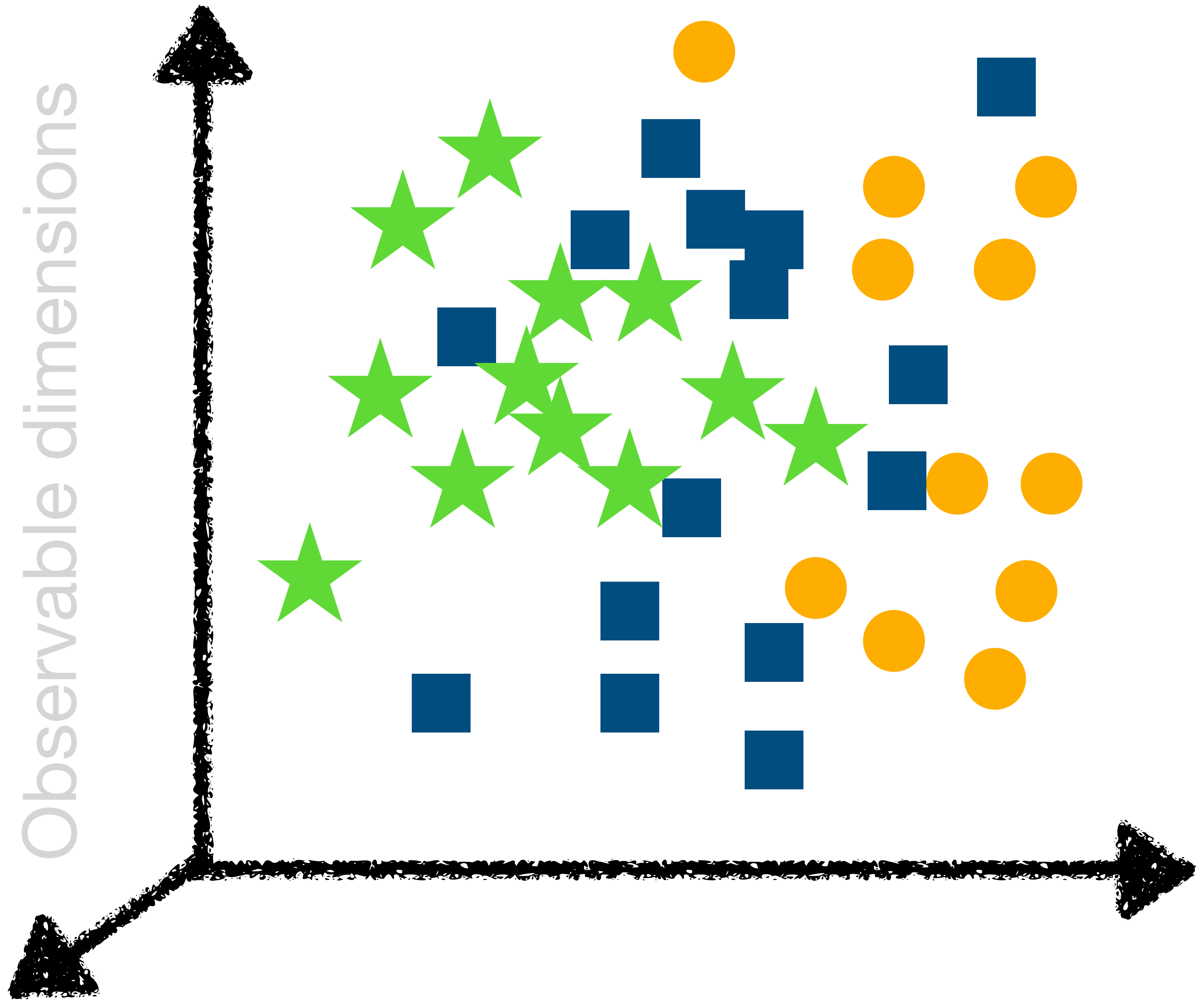
- No labels
- No feedback
- "Find hidden structure"

- Decision process
- Reward system
- Learn series of actions

[S. Raschka (2015)]

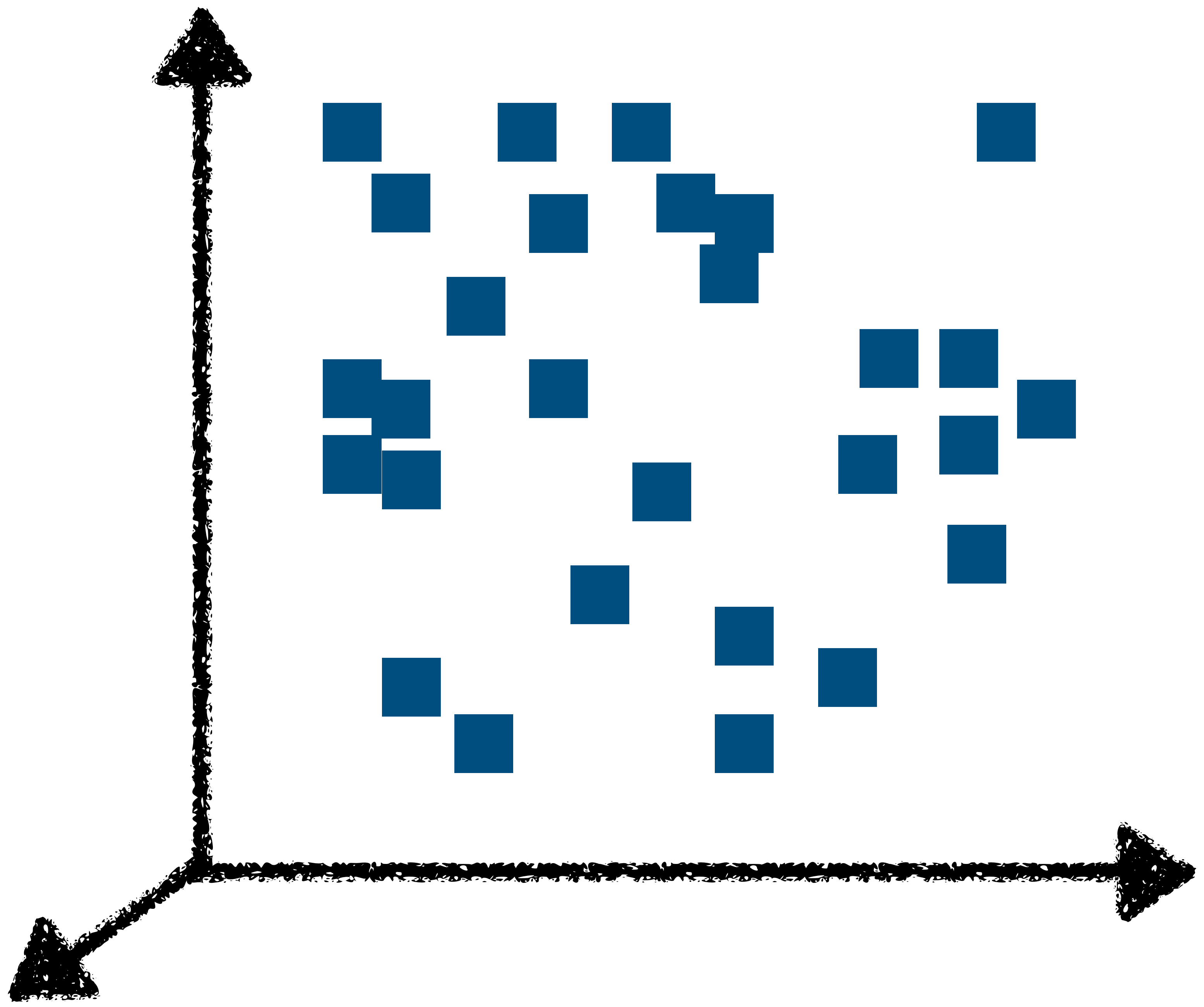
Machine Learning Approaches

Supervised



Labelled (outcome) data
Direct/quantifiable metrics
on learning efficacy (score)

Unsupervised



No labels
No explicit feedback

Also, self-supervised, semi-supervised

Regression (Supervised)

Goal: predict a *continuous* outcome y variable from a vector of observable input features $\vec{x}$. Use **training set** of $(\vec{x}, y)$ pairs to learn this mapping: $f(\vec{x}) = y$

Theory-driven vs. **data-driven** approach…

Some non-neural algorithms that are still *very* useful and performant:

- Linear Regression: $f_w(\vec{x}) = w_0 + w_1 x_1 + \ldots + w_p x_p$
- Lasso & Ridge
- Gaussian Process Regression
- k-Nearest Neighbor Regression
- Regression Forests

Classification (Supervised)

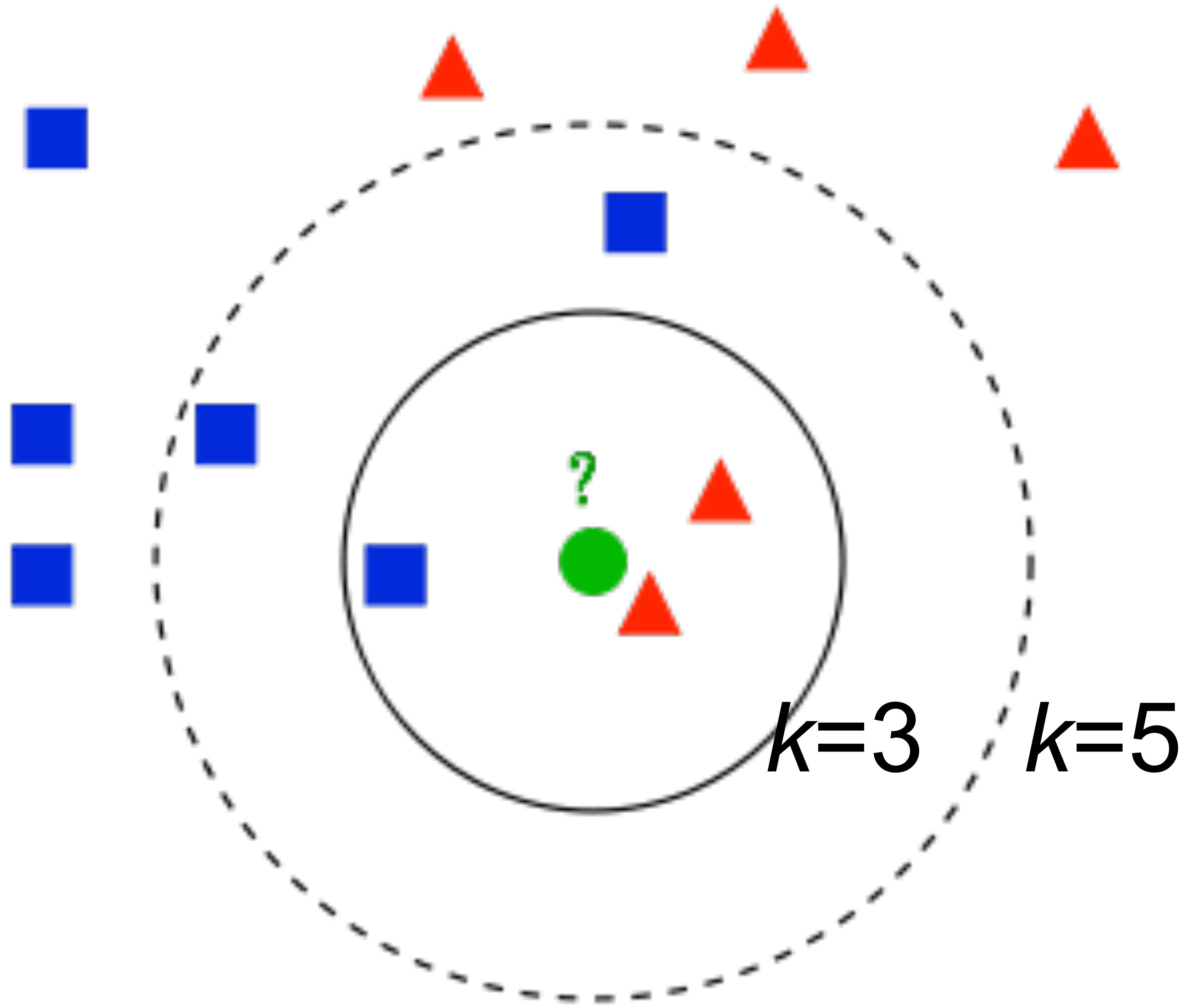
Goal: predict a *discrete* class y_n (n classes) from a vector of observable input features $\vec{x}$. Use **training set** of $(\vec{x}, y_n)$ pairs to learn this mapping: $f(\vec{x}) = y_n$

Some of the non-neural algorithms that are still useful:

- Logistic Regression
- KNN Classification
- LDA / QDA
- Naive Bayes
- Random Forest & boosted trees

Classification (Supervised)

kNearestNeighbors (kNN)

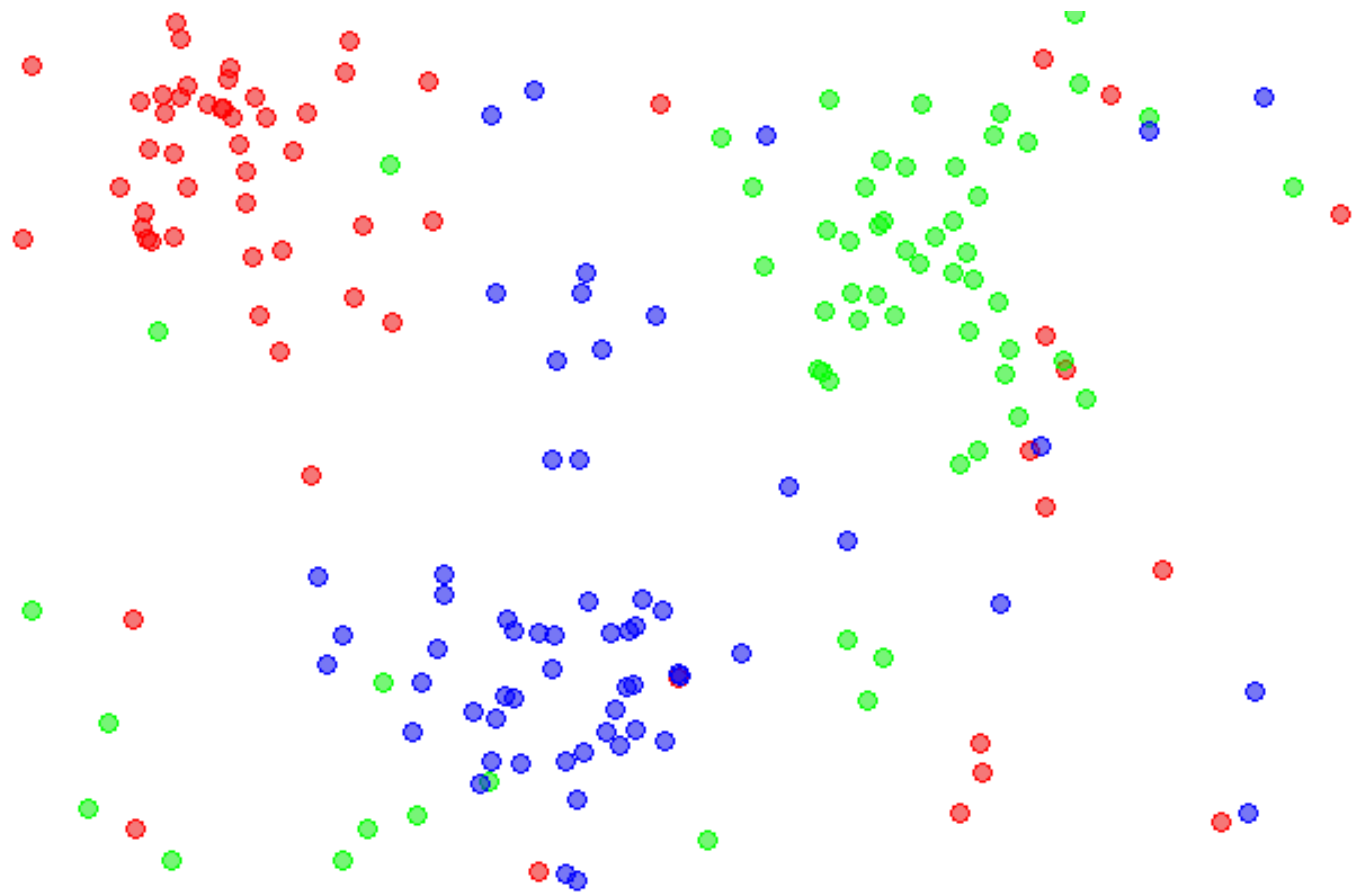


For each test point, $\vec{x}$ find the *k*-nearest instances in the training data

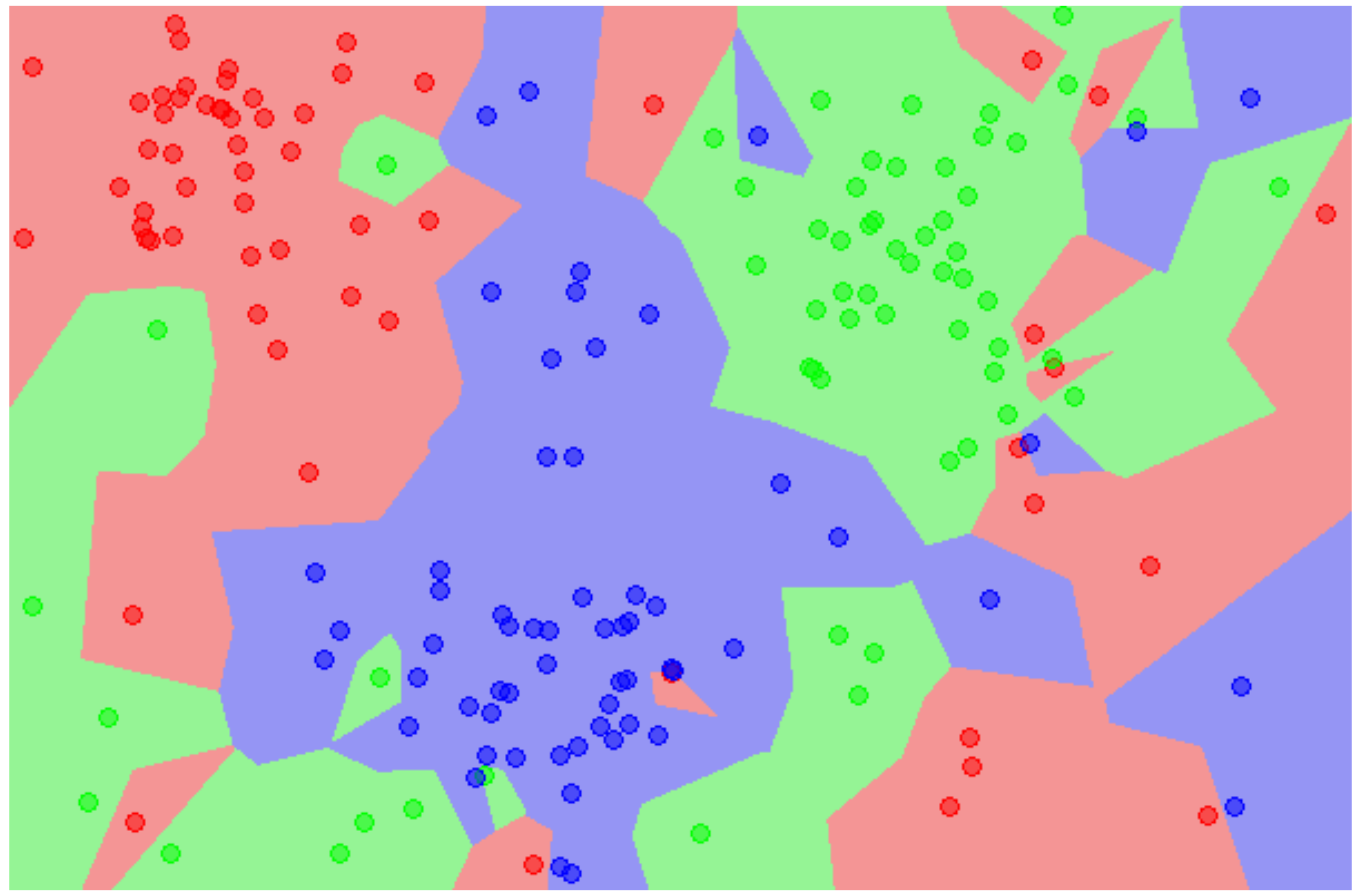
Classify the point according to the majority vote of their class labels

Classification (Supervised)

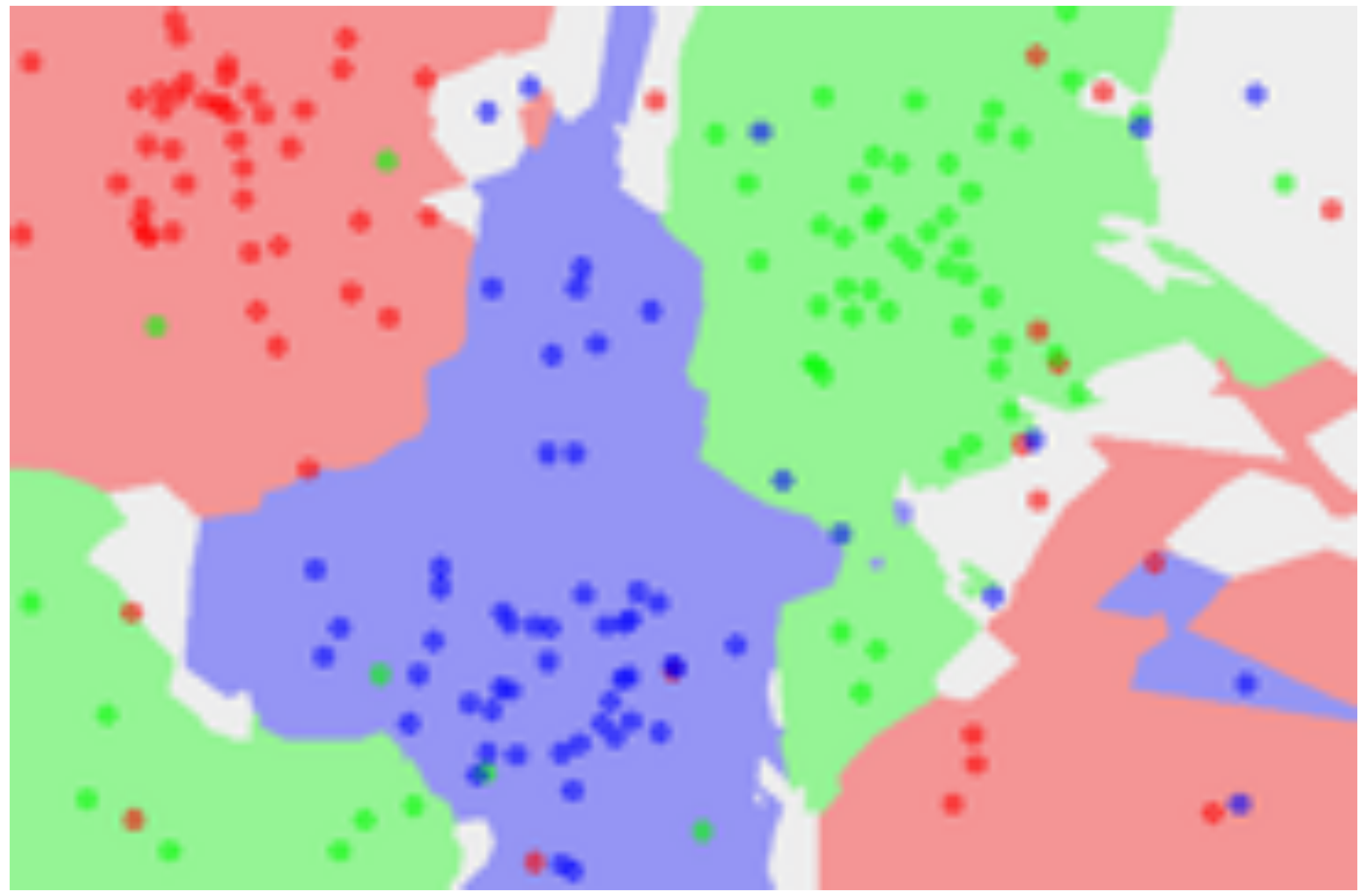
kNearestNeighbors (kNN)



Dataset



k=1



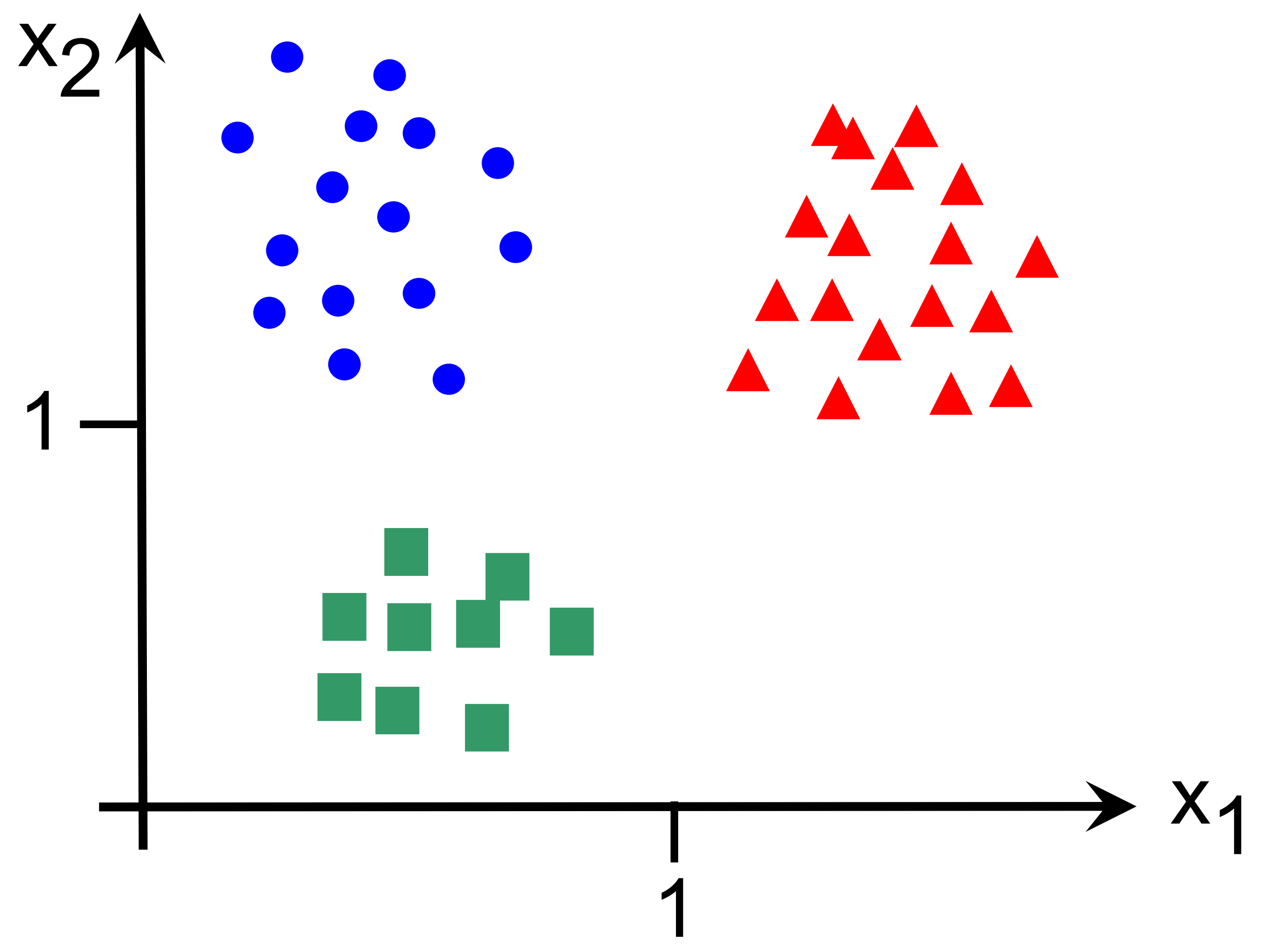
k=5

k is a **hyperparameter** that must be learned/tuned

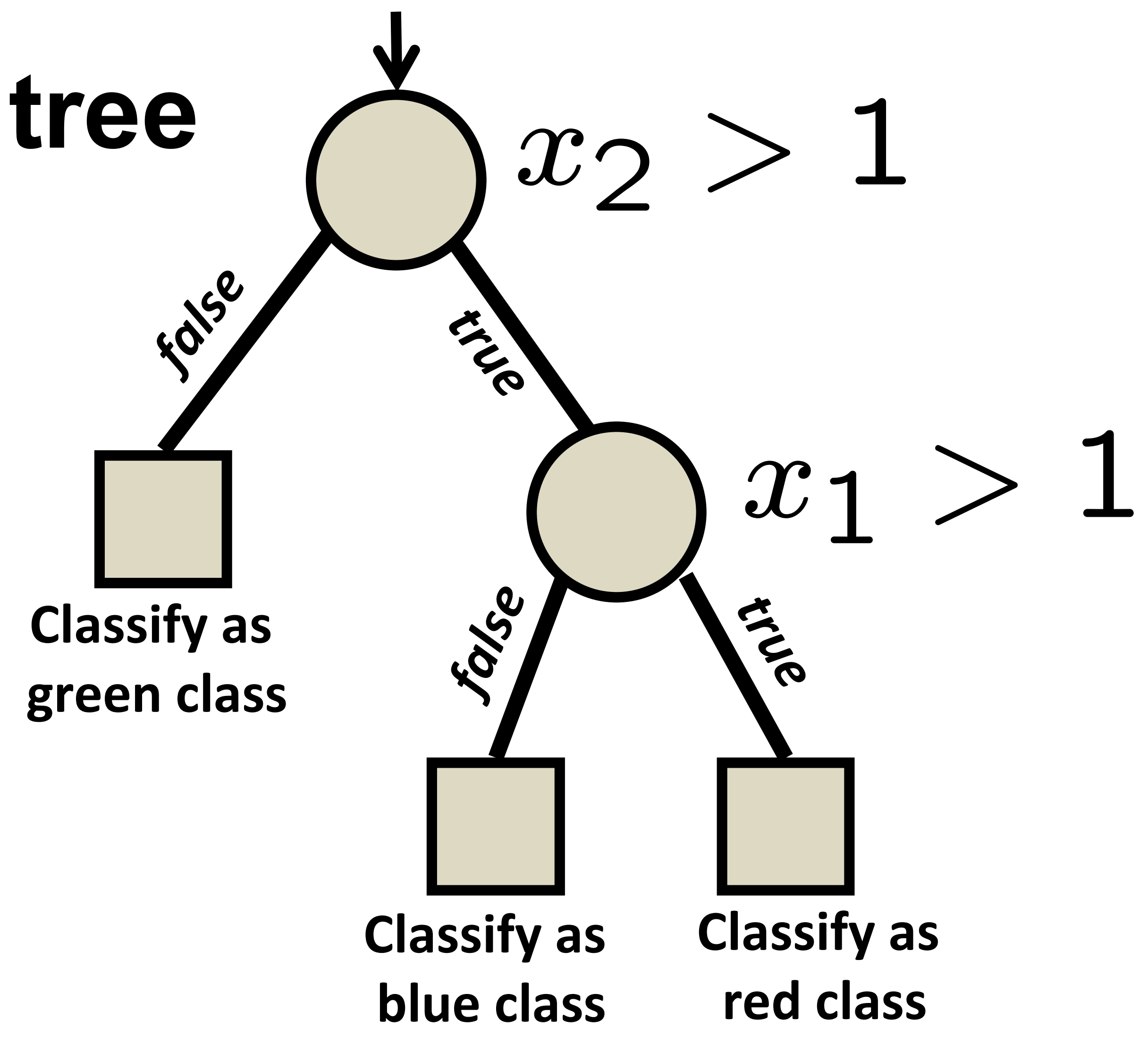
All ML models have their own set of hyperparameters

Classification

Decision Trees



A decision tree



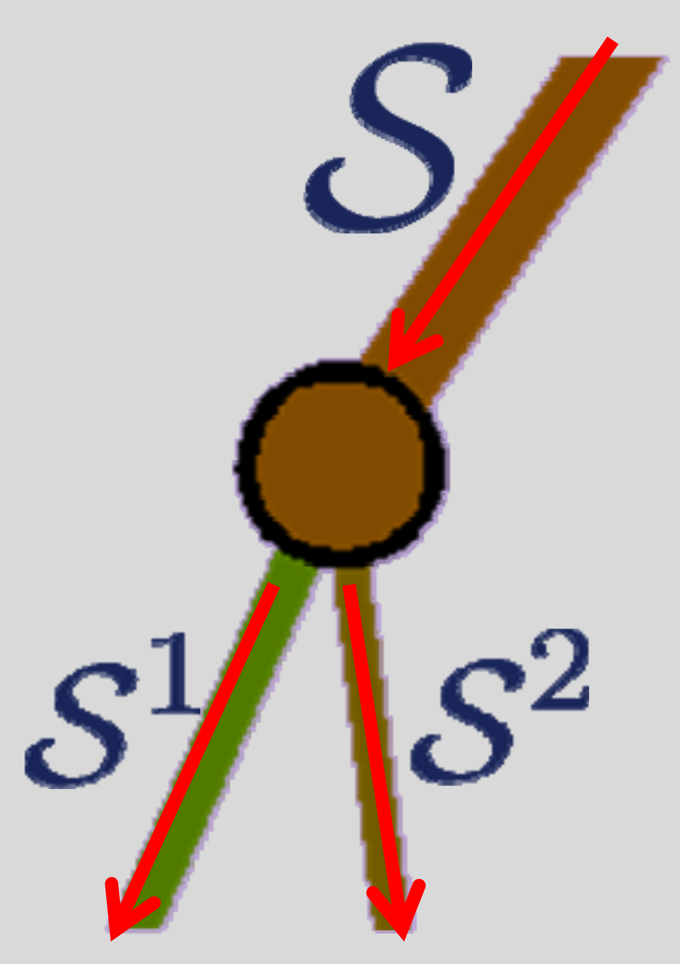
Classification and regression trees (CART)

http://www.robots.ox.ac.uk/~az/lectures/ml/lect5.pdf

Classification

Building Trees Rigorously (Node Splitting Criteria)

Decision Trees



Information gain

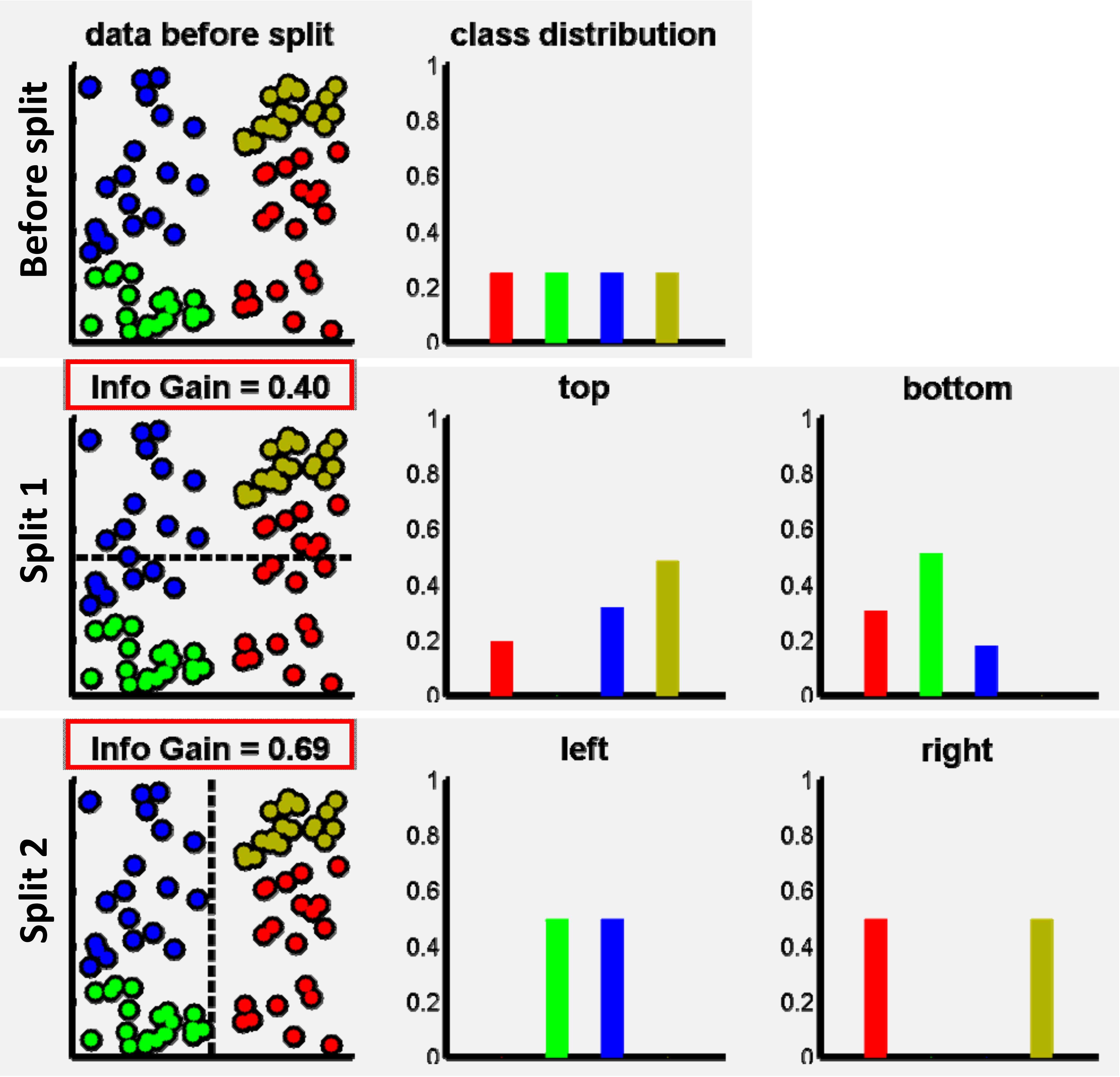
$$I = H(\mathcal{S}) - \sum_{i \in \{1,2\}} \frac{|\mathcal{S}^i|}{|\mathcal{S}|} H(\mathcal{S}^i)$$

Shannon's entropy

$$H(\mathcal{S}) = -\sum_{c \in \mathcal{C}} p(c) \log(p(c))$$

Node training

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta} \in \mathcal{T}} I$$



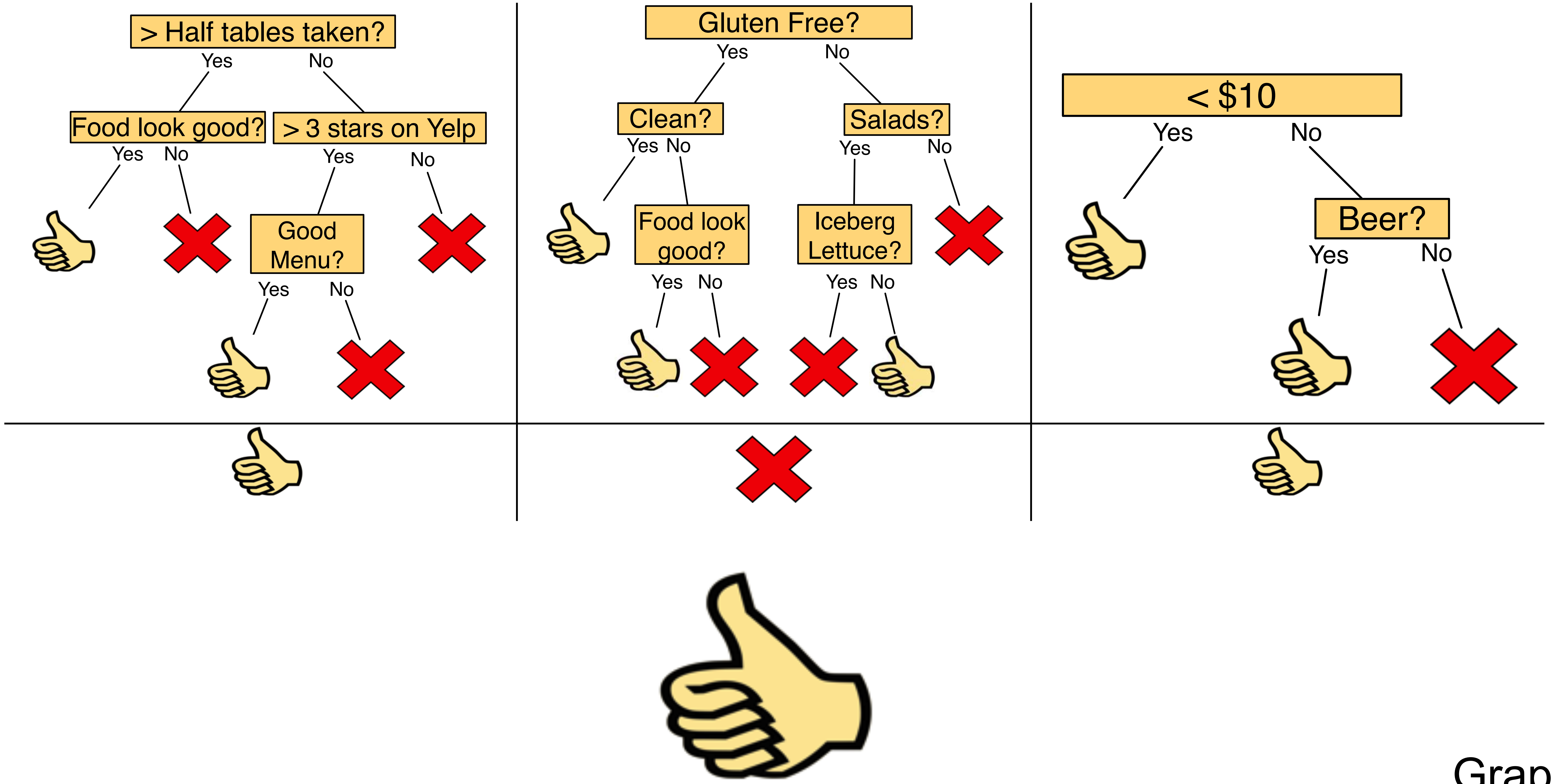
A. Criminisi, J. Shotton and E. Konukoglu (2012). *Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning*- pp 97.

Classification

Decision Trees

Random Forests

ensembles generally increase robustness



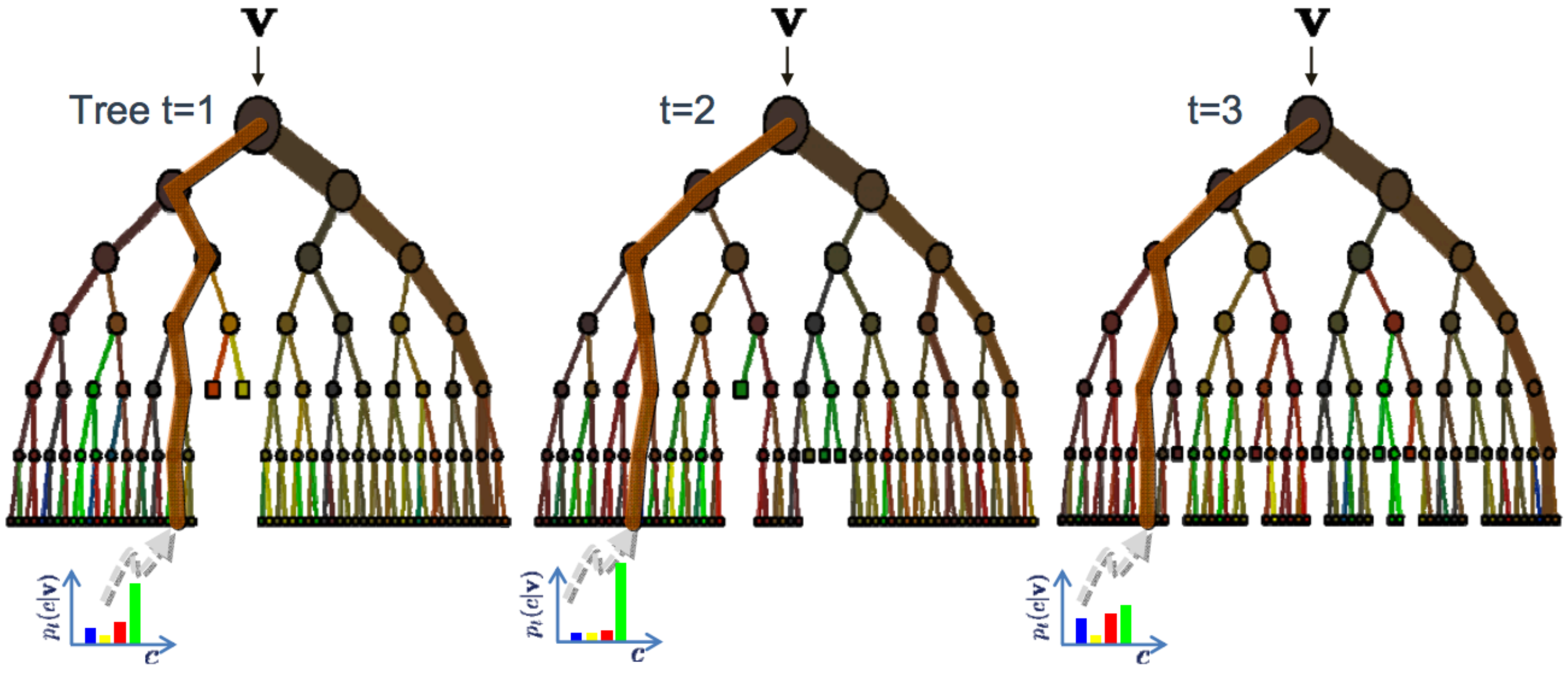
Graphic via D. Eads

Classification

Decision Trees

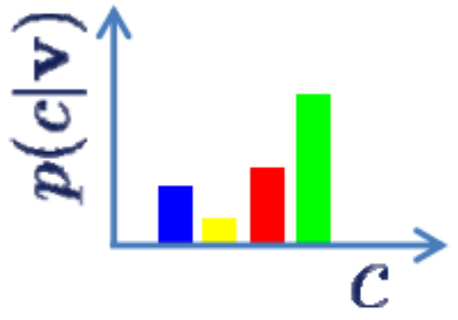
Random Forests

ensembles generally increase robustness



The ensemble model

Forest output probability $p(c|\mathbf{v}) = \frac{1}{T}\sum_{t}^{T} p_t(c|\mathbf{v})$



Classification

Decision Trees

Random Forests

ensembles generally increase robustness

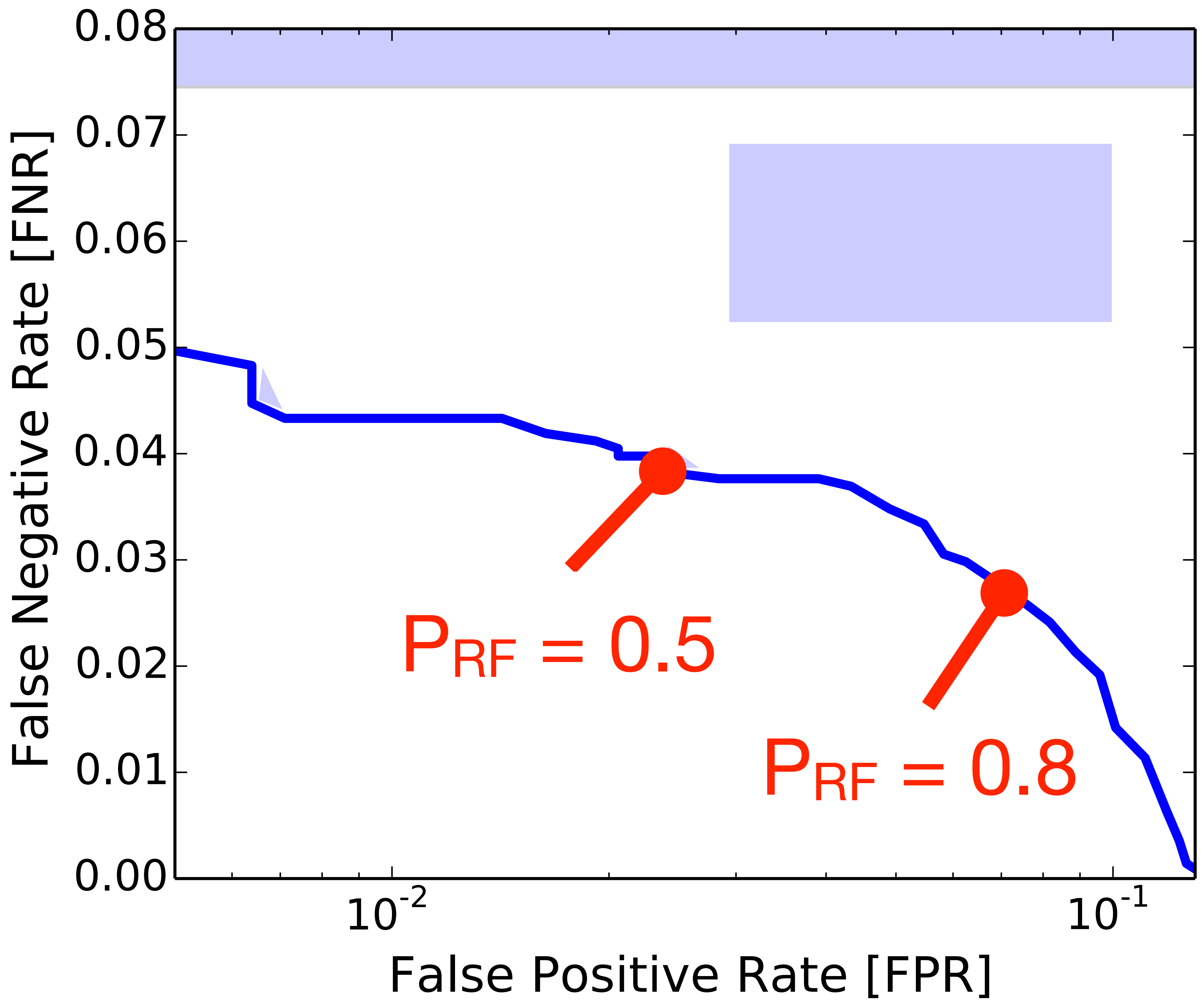
😎 Tree ensembles (RF, xgboost, lightGBM, …) are natural and usually SOTA approaches for **tabular data**

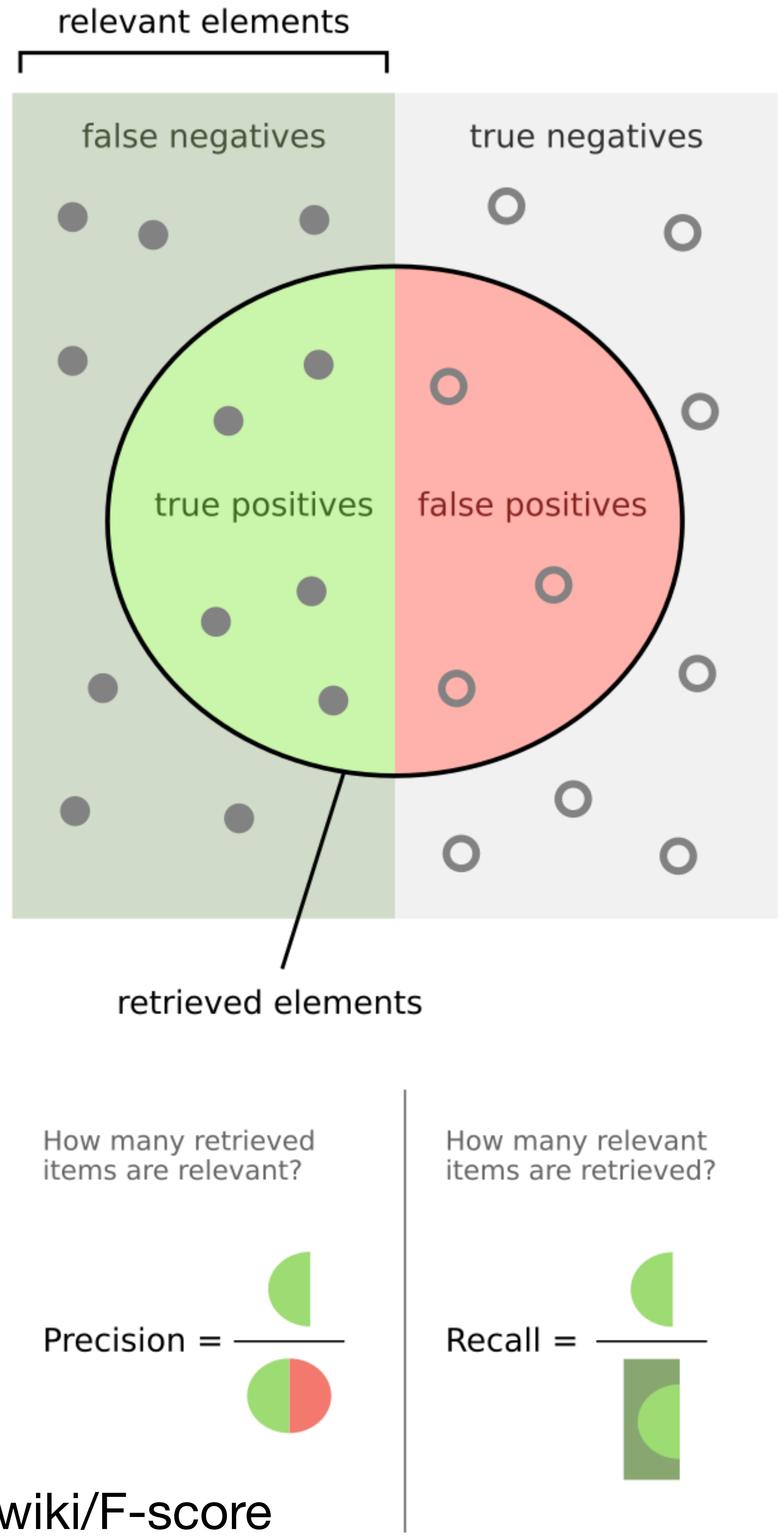
- Splits are performed in the natural units of each feature (as opposed to ad hoc normalization & weighting)
- Feature importance and (“out of bag”) error estimatation are natural
- (Frequentist) probabilities are natural

Classification

Decision Trees

Mapping P(class) to other evaluation metrics

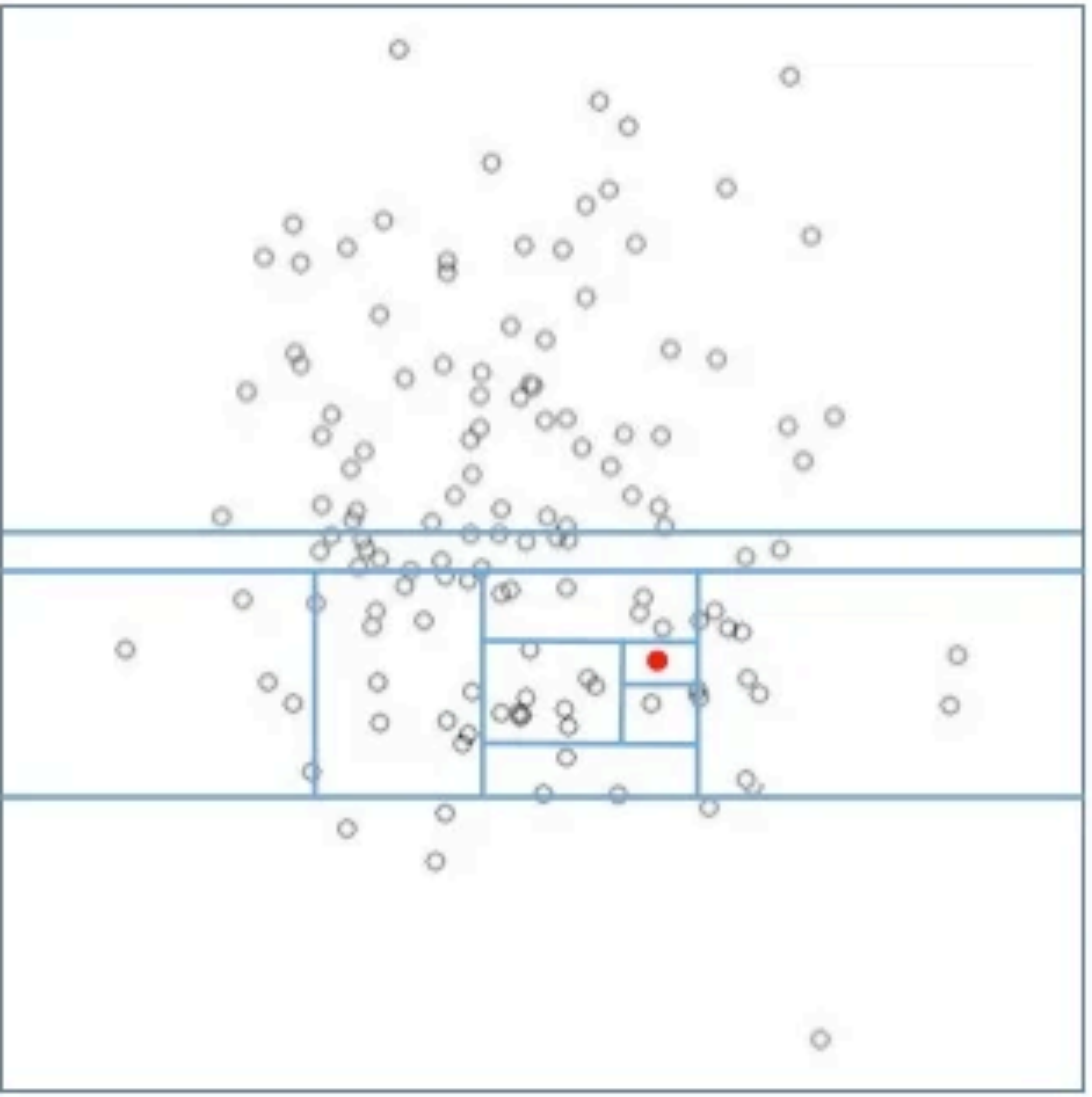




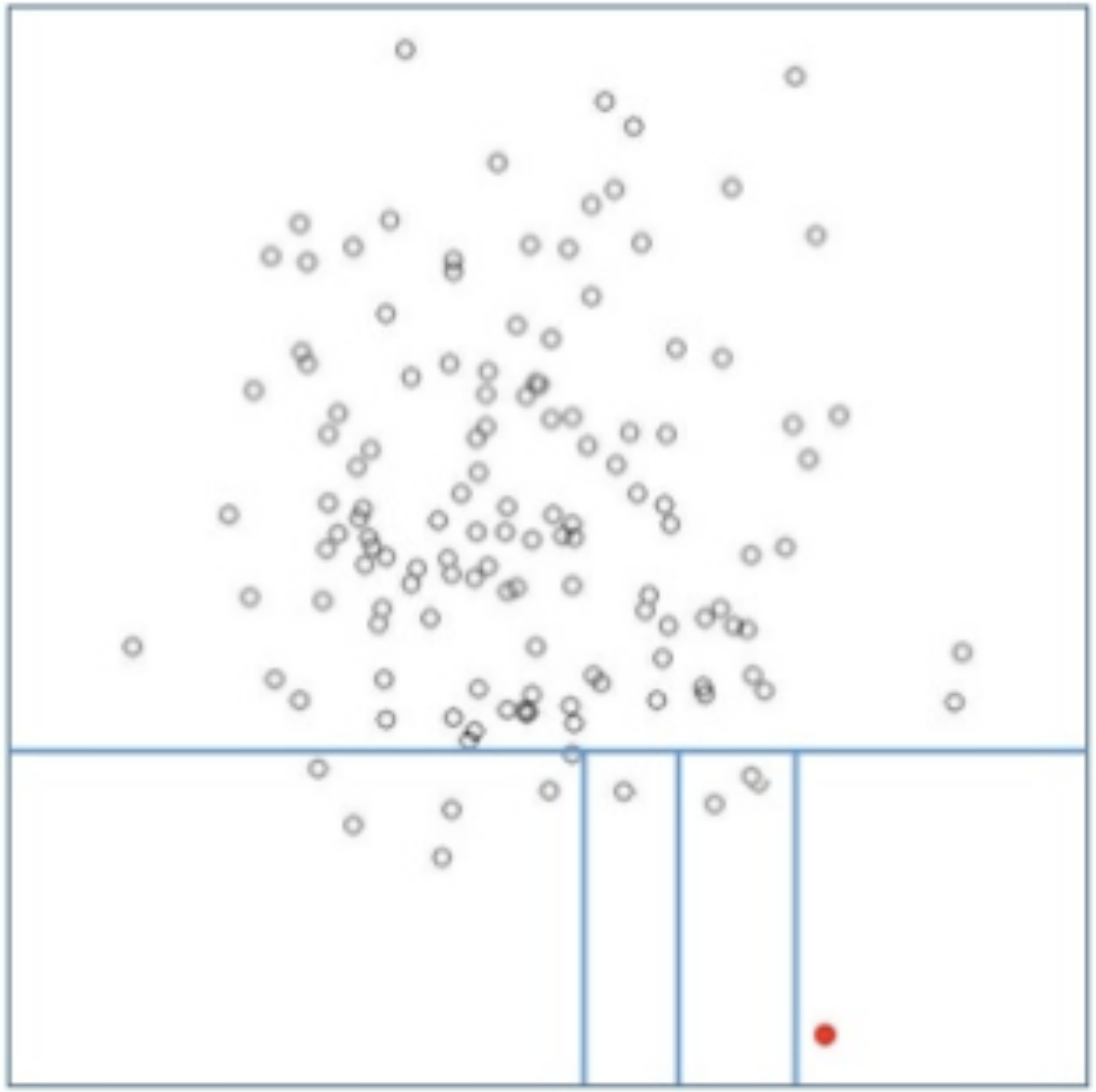
wikipedia.org/wiki/F-score

Anomaly Detection

Isolation Forests

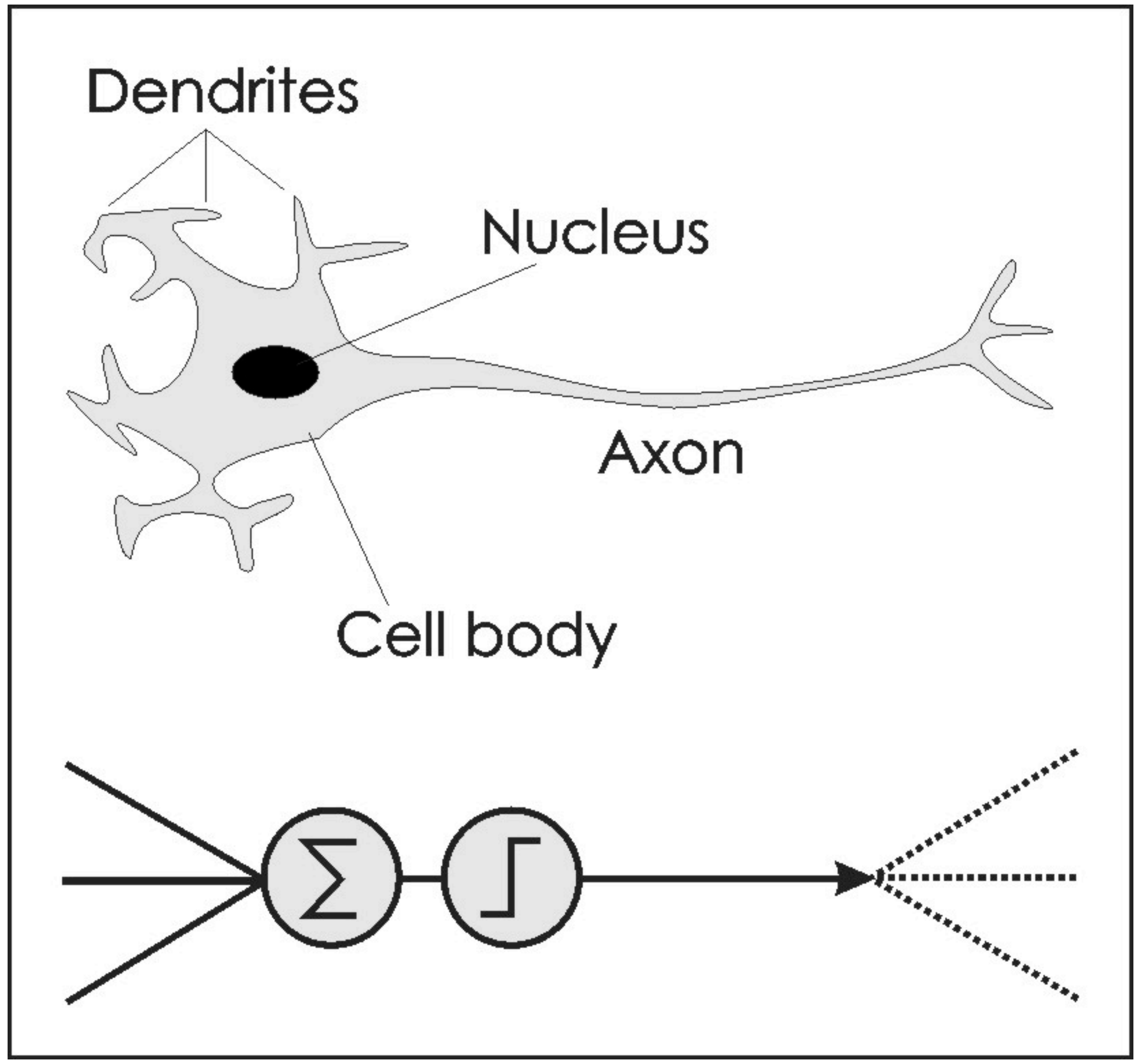


10 lines are needed to isolate this data point (not anomalous)



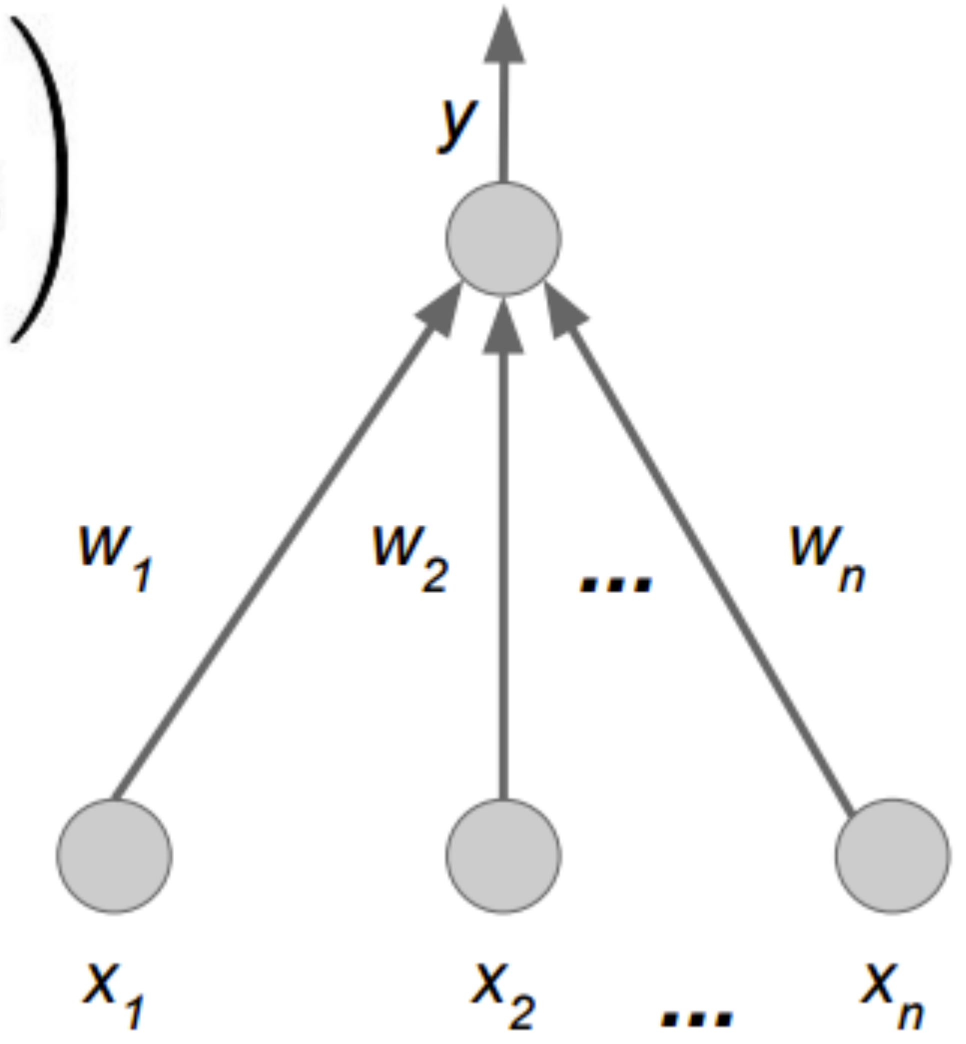
Only **4 lines** are needed to isolate this data point (highly anomalous)

Neural Networks

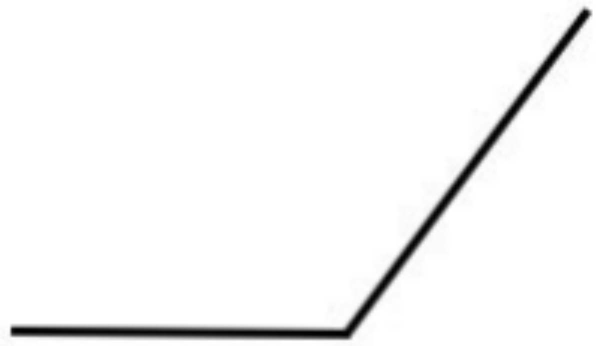


Neuron-inspired math (1958)

$$y = F\left(\sum_i w_i x_i + b\right)$$



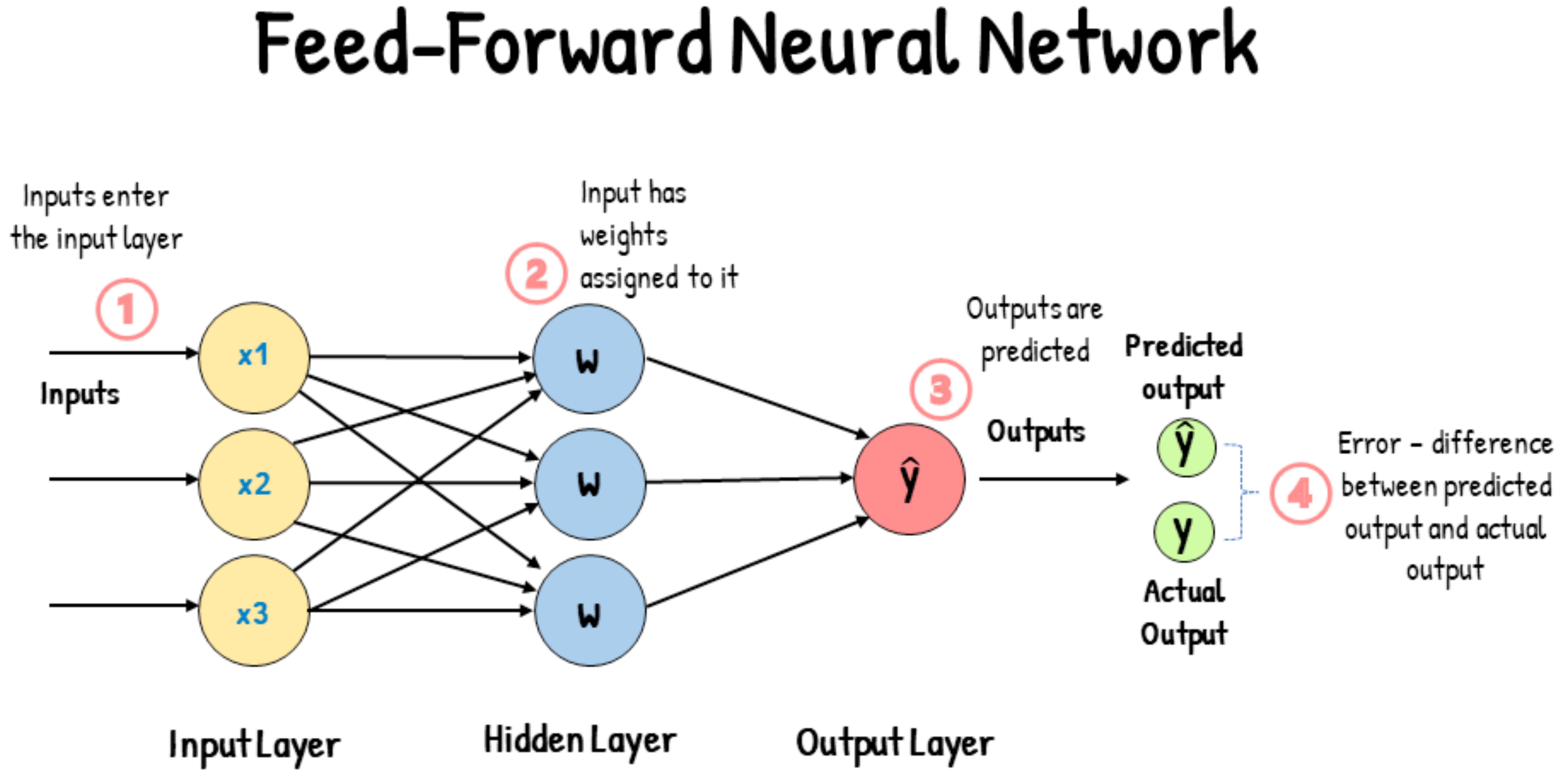
"1 layer" perceptron



$F(x) = \max(0, x)$

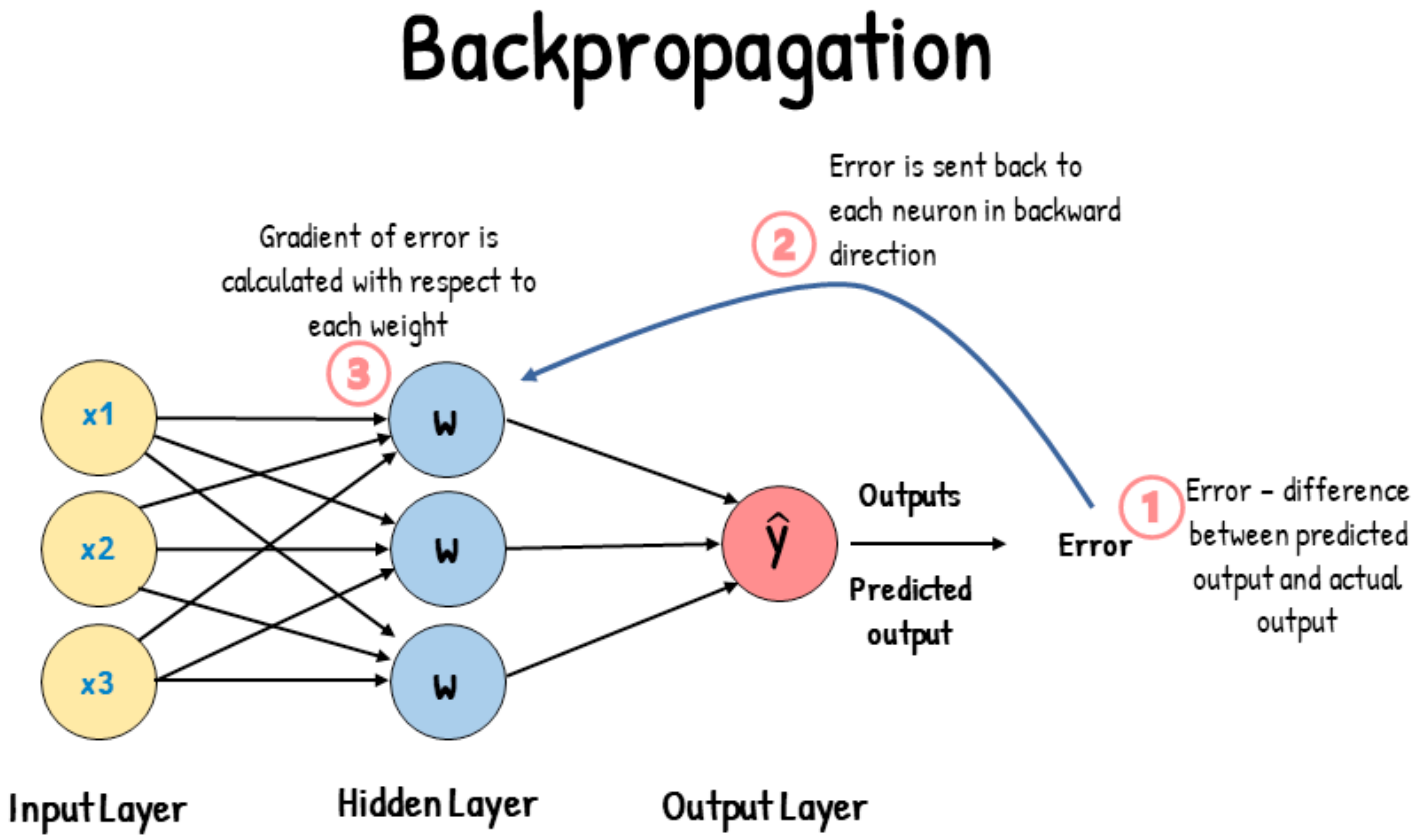
Activation function
(non-linearity)

Neural Networks

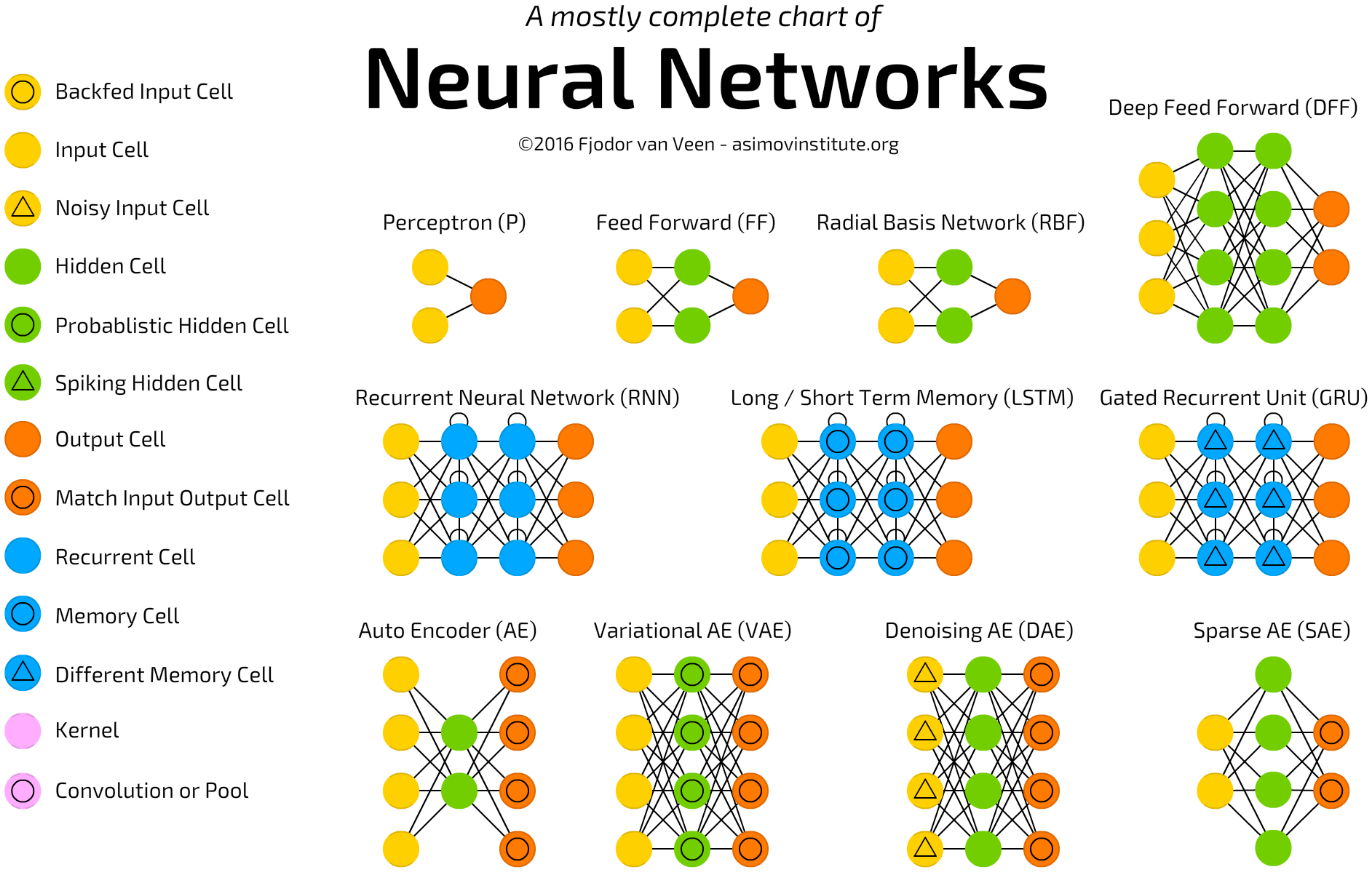


https://www.analyticsvidhya.com/blog/2023/01/gradient-descent-vs-backpropagation-whats-the-difference/

Neural Networks



https://www.analyticsvidhya.com/blog/2023/01/gradient-descent-vs-backpropagation-whats-the-difference/


A mostly complete chart of
Neural Networks
©2016 Fjodor van Veen - asimovinstitute.org
Backfed Input Cell
Input Cell
Noisy Input Cell
Hidden Cell
Probablistic Hidden Cell
Spiking Hidden Cell
Output Cell
Match Input Output Cell
Recurrent Cell
Memory Cell
Different Memory Cell
Kernel
Convolution or Pool
Perceptron (P)
Feed Forward (FF)
Radial Basis Network (RBF)
Deep Feed Forward (DFF)
Recurrent Neural Network (RNN)
Long / Short Term Memory (LSTM)
Gated Recurrent Unit (GRU)
Auto Encoder (AE)
Variational AE (VAE)
Denoising AE (DAE)
Sparse AE (SAE)

Basic Training & Hyperparameter Optimization

- Preprocess (and clean) raw data using domain-specific understanding. Coercion into data formats appropriate for the model

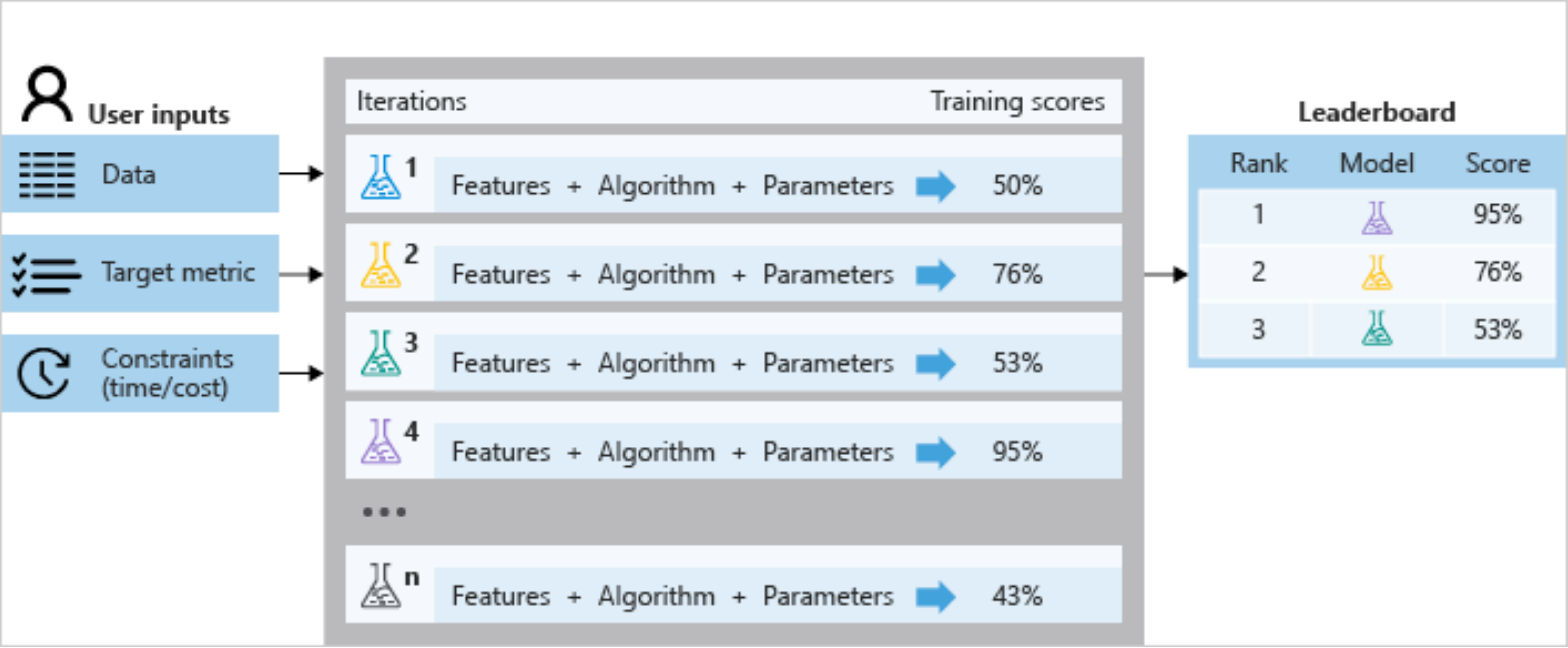
- Split $(\vec{x}, y)$ data into three distinct sets called **train**, **validation**, **test**

- Define a use-case specific **score** which we try to optimize — e.g. mean-square error MSE (regression) or binary cross entropy (classification)

- With a fix set of hyperparameters, learn model on train set. Determine quality (and stopping criteria) using valdiation set.

- Vary hyperparameters to obtain best valdiation score. Use best model and report its quality on test set

Basic Training & Hyperparameter Optimization



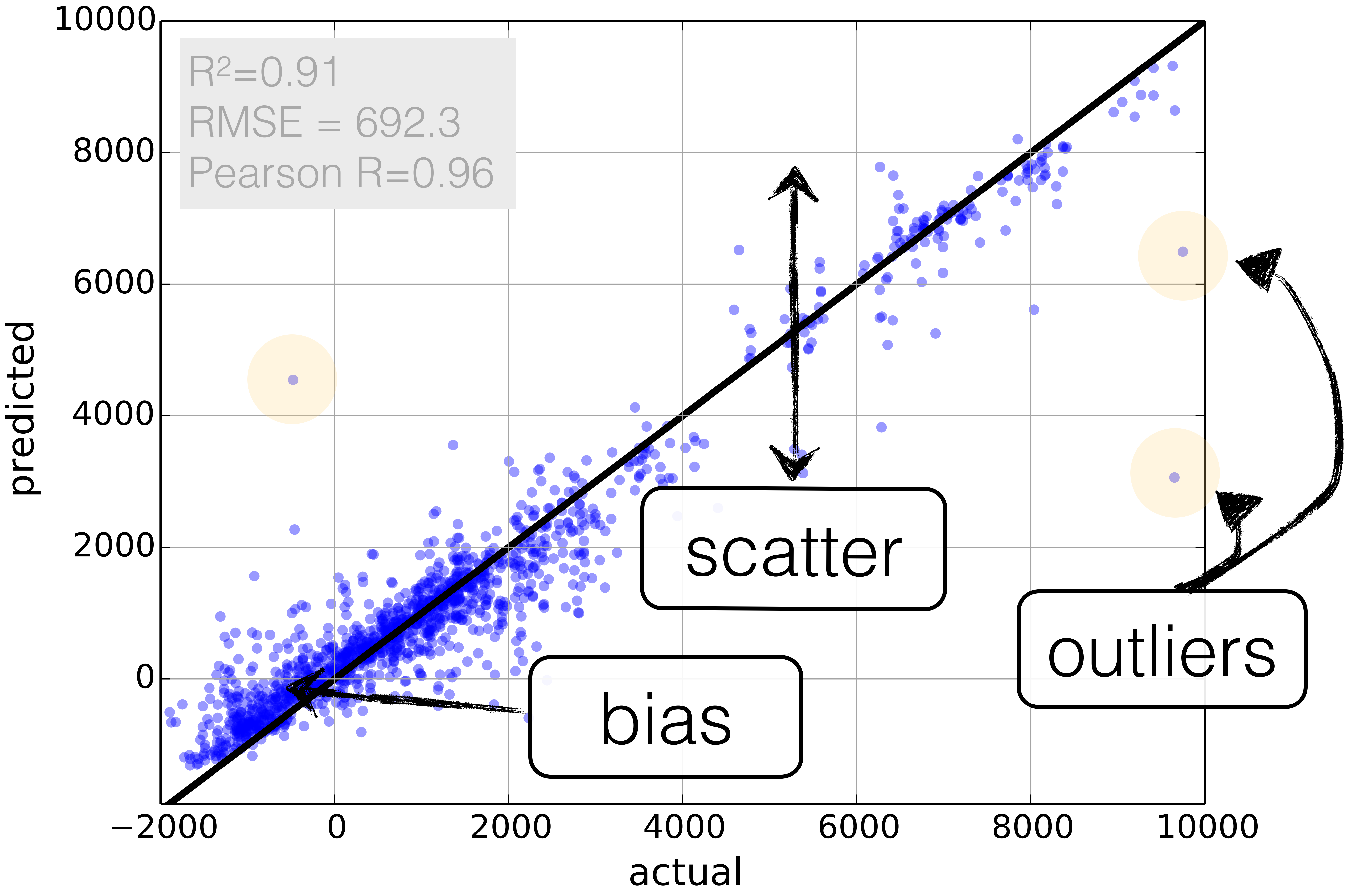
Optimization strategies: grid search, random, Bayes (*hyperopt*, *automl)*
Model training management: Weights & Biases, Tensorboard, …

Basic Training & Hyperparameter Optimization

Scoring metrics are domain-specific…be thoughtful about this

Scalar proxies:

- RMSE
- RMSLE
- [adjusted] R^2
- …



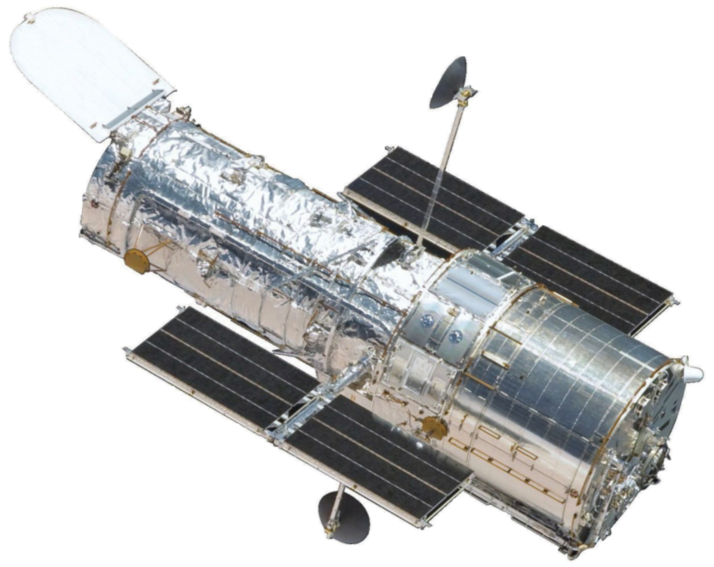
Example Use Cases

Why do ML? Too Many Transients Tax (Follow-Up) Resources

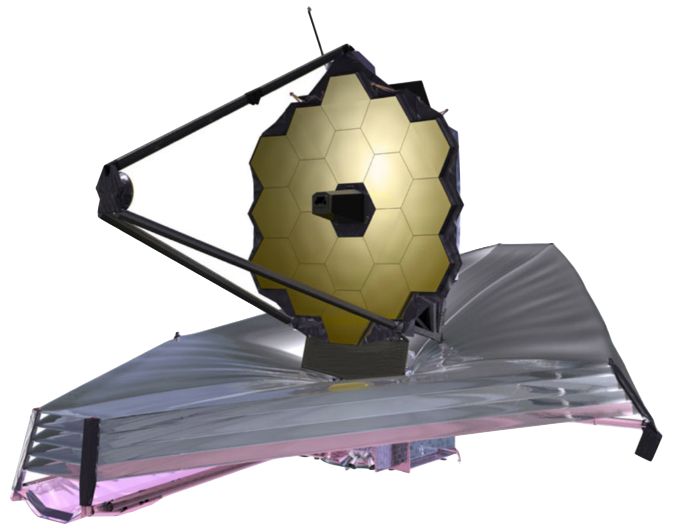
	Palomar Transient Factory (PTF) 2009-2016	**Zwicky Transient Factory (ZTF)** 2017-2024	**Large Synoptic Survey Telescope (LSST)** 2024-2034
Image data rate	1 GB/90s	3 GB/45 s	6 GB/5 s
Transient Alerts per night	4×10^4	3×10^5	2×10^6

“cheap” discovery

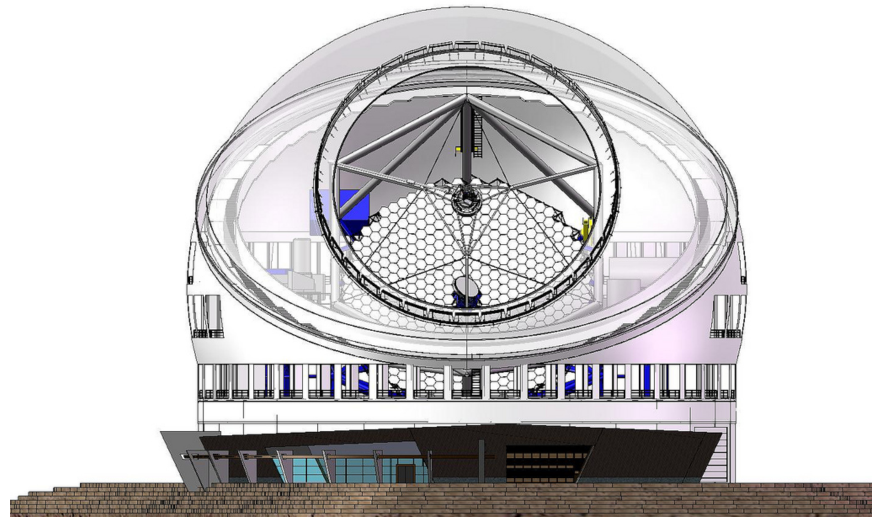
“expensive” followup



Hubble Space Telescope (HST)



James Webb Space Telescope (JWST)

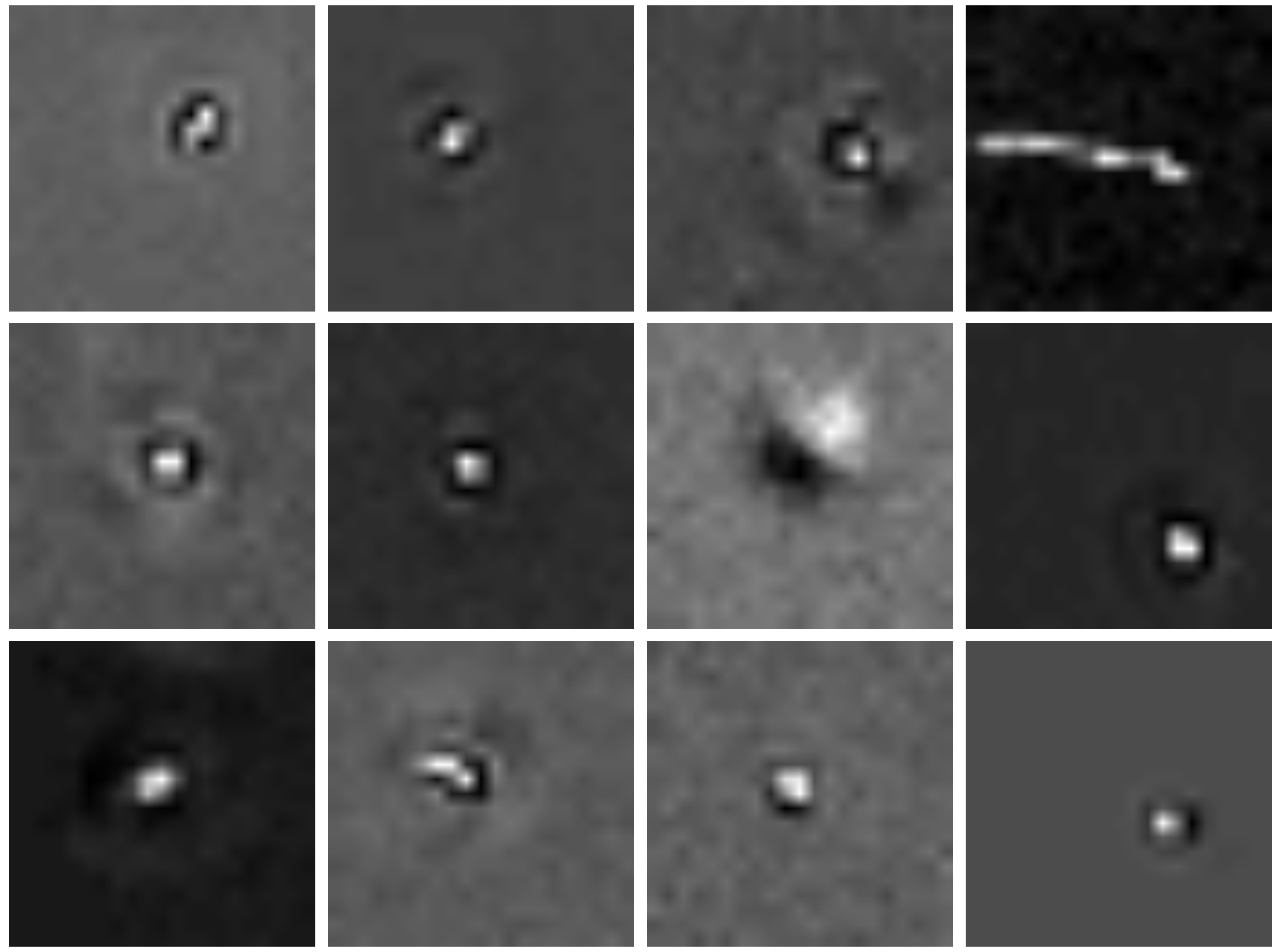


Thirty-Meter Telescope (TMT)

Harvard College Observatory c. 1890

Example Classification Task: Discovery for Astro Survey Images

"bogus"



"real"

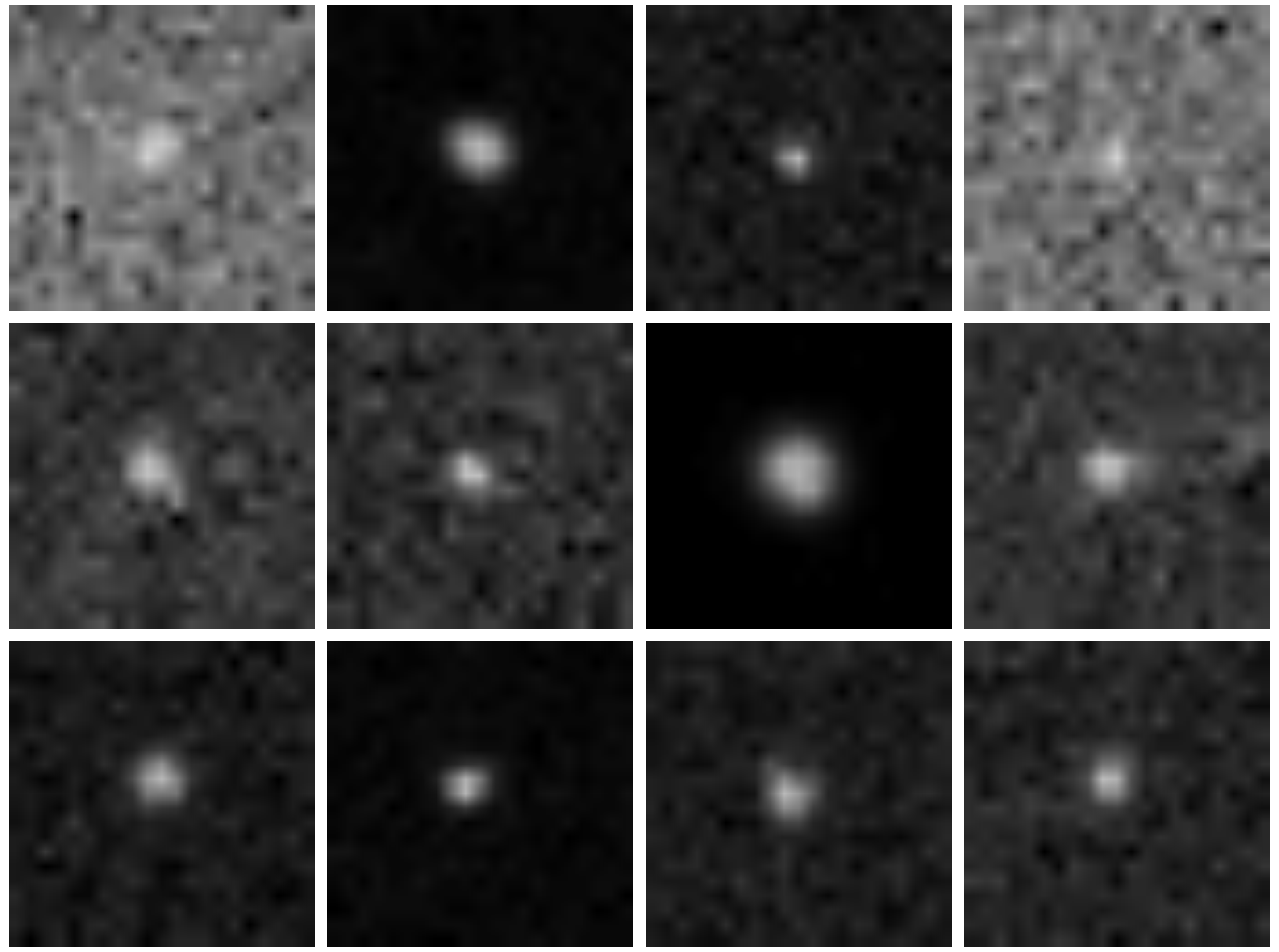


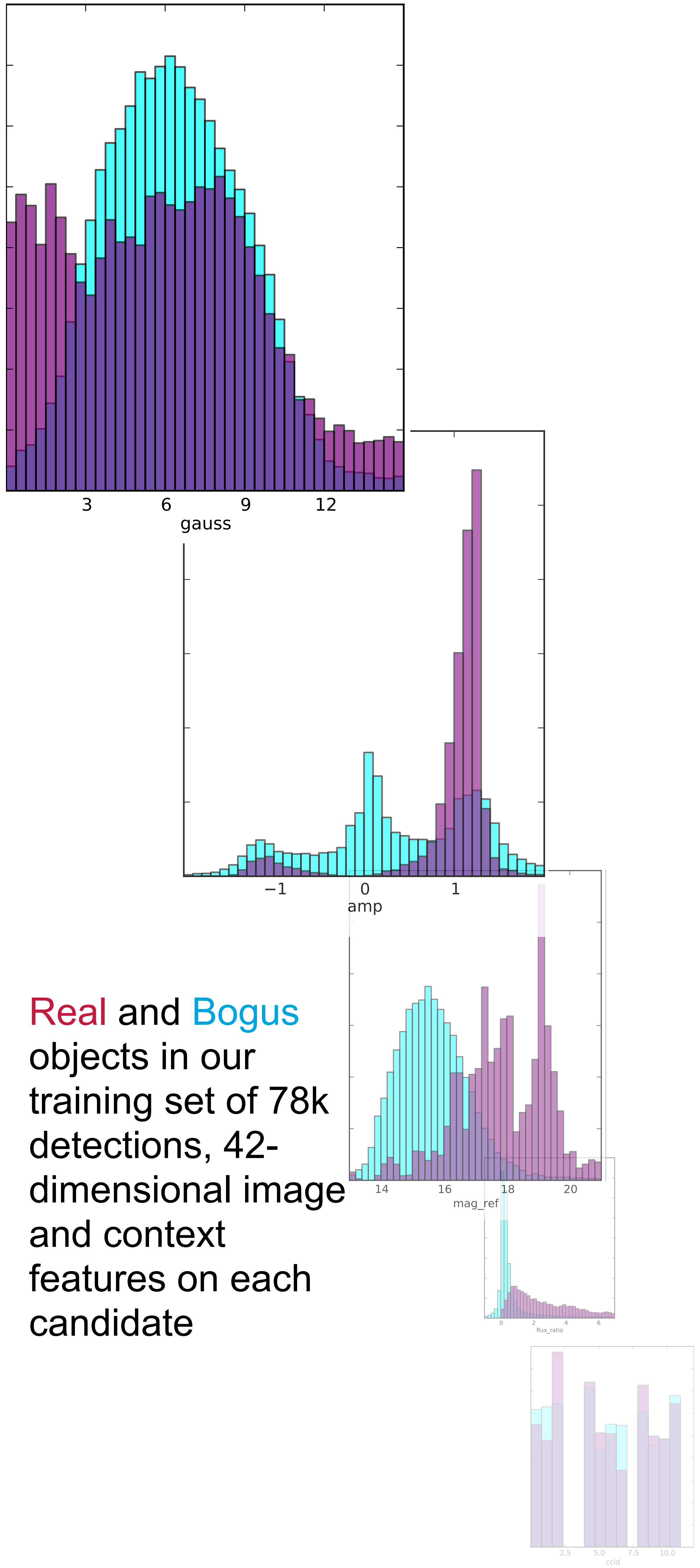
image "subtractions"

Built a real-time framework to discover variable/transient sources without people

- fast (compared to people)
- parallelizable
- transparent
- deterministic
- versionable

1000 to 1 needle in the haystack problem

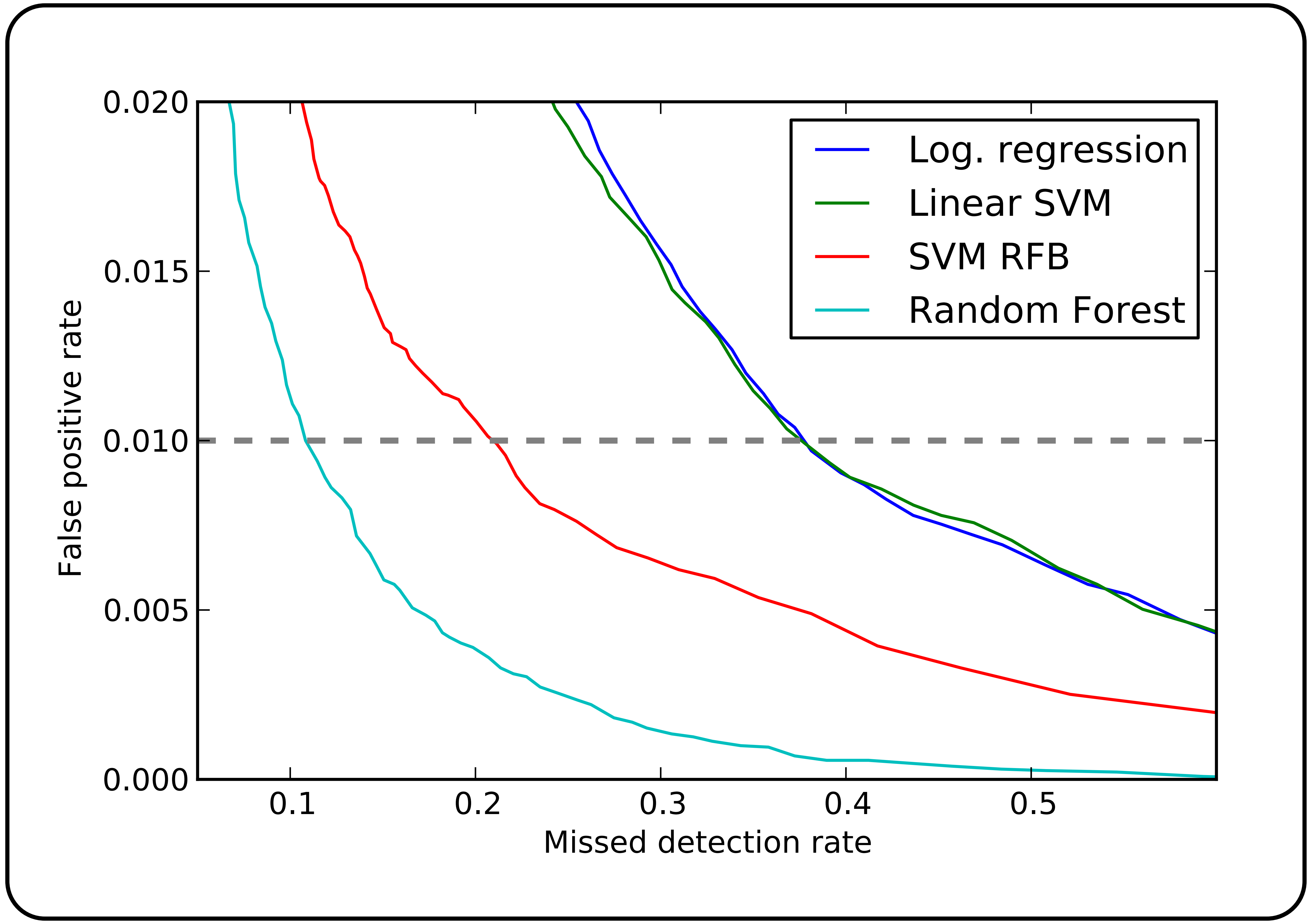
arXiv.org > astro-ph > arXiv:1209.3775



Real and Bogus objects in our training set of 78k detections, 42-dimensional image and context features on each candidate

Some classifiers work better than others

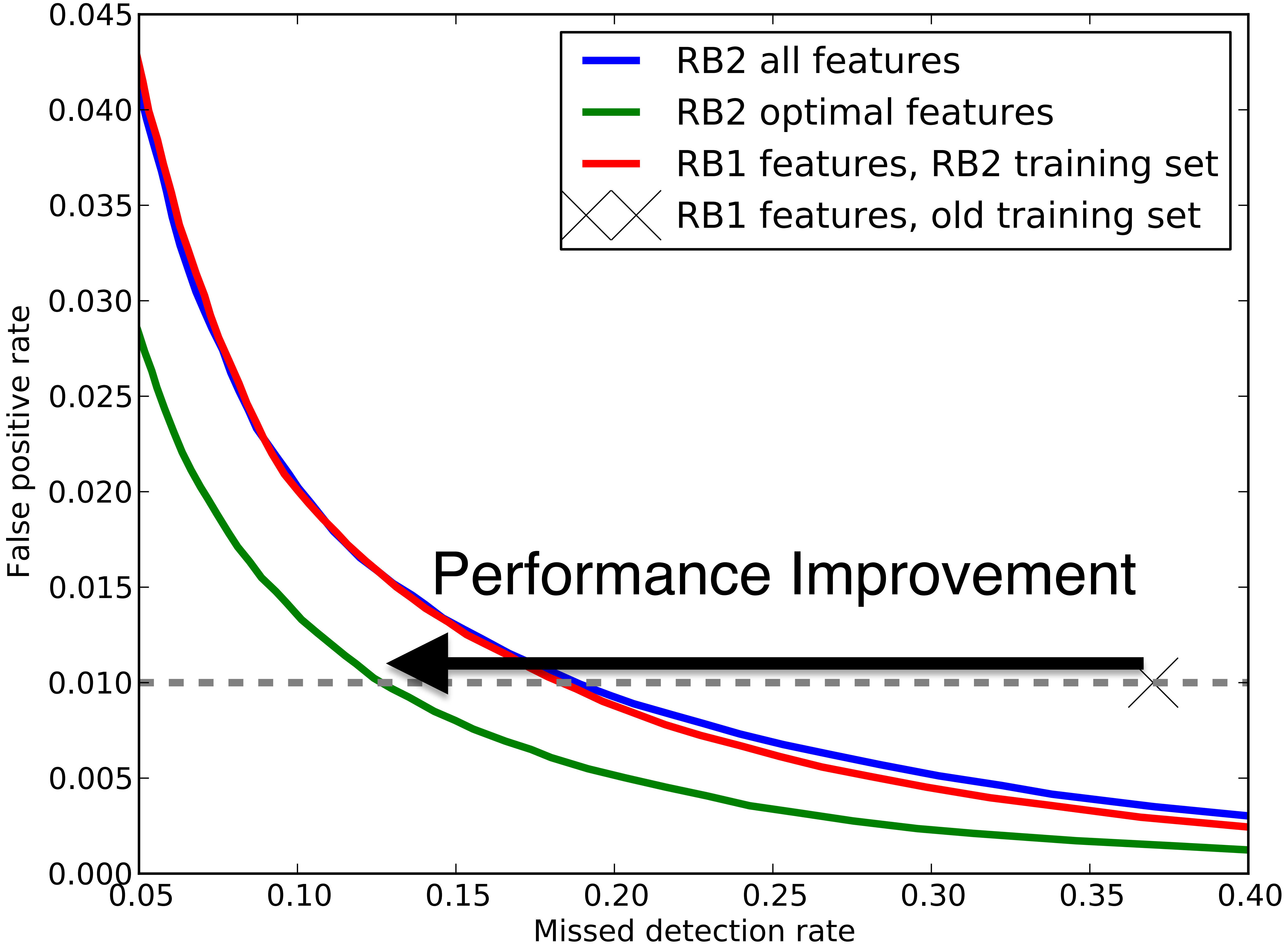
ROC Curve



Brink+2012

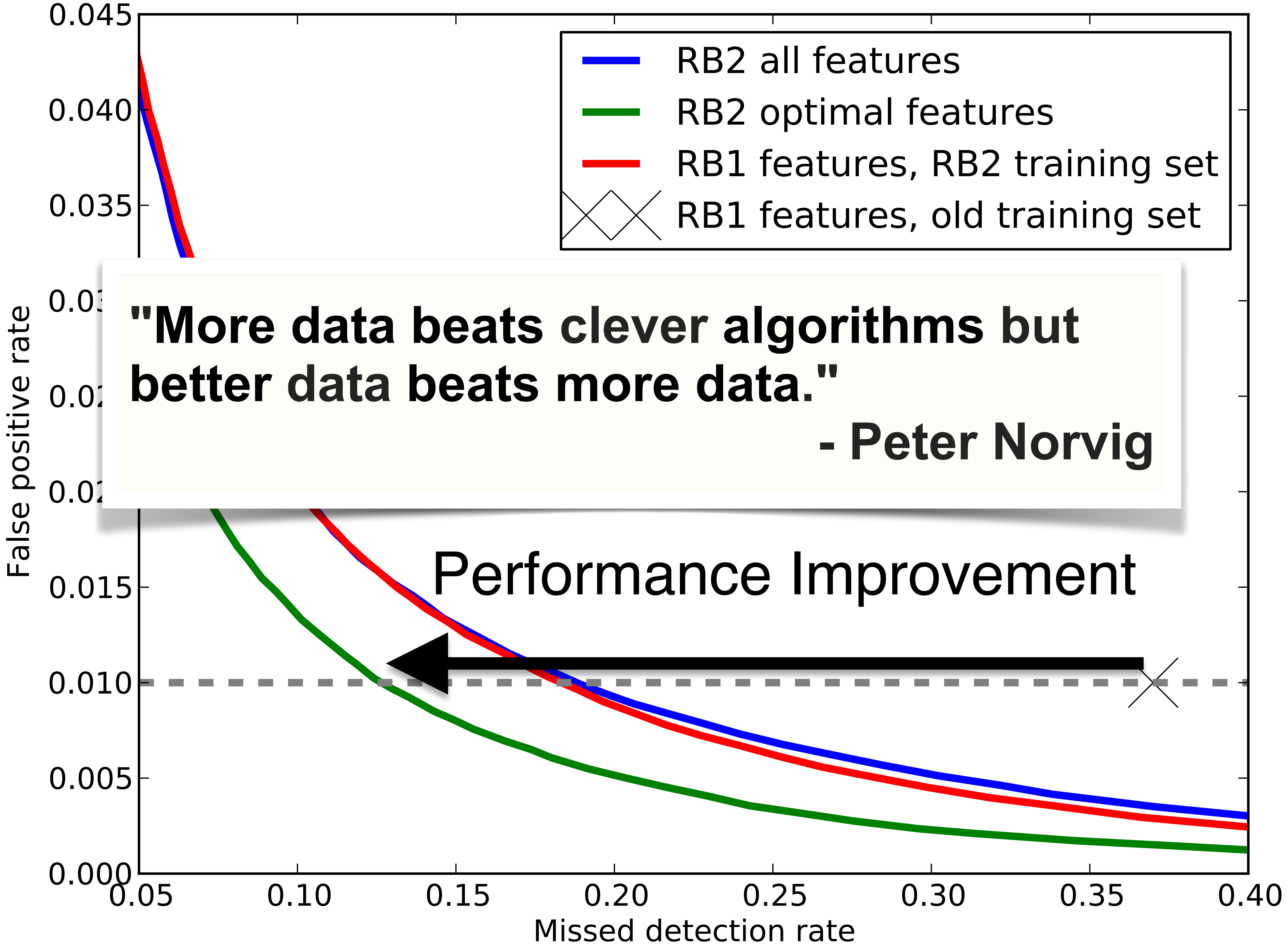
arXiv.org > astro-ph > arXiv:1209.3775

Better Models…

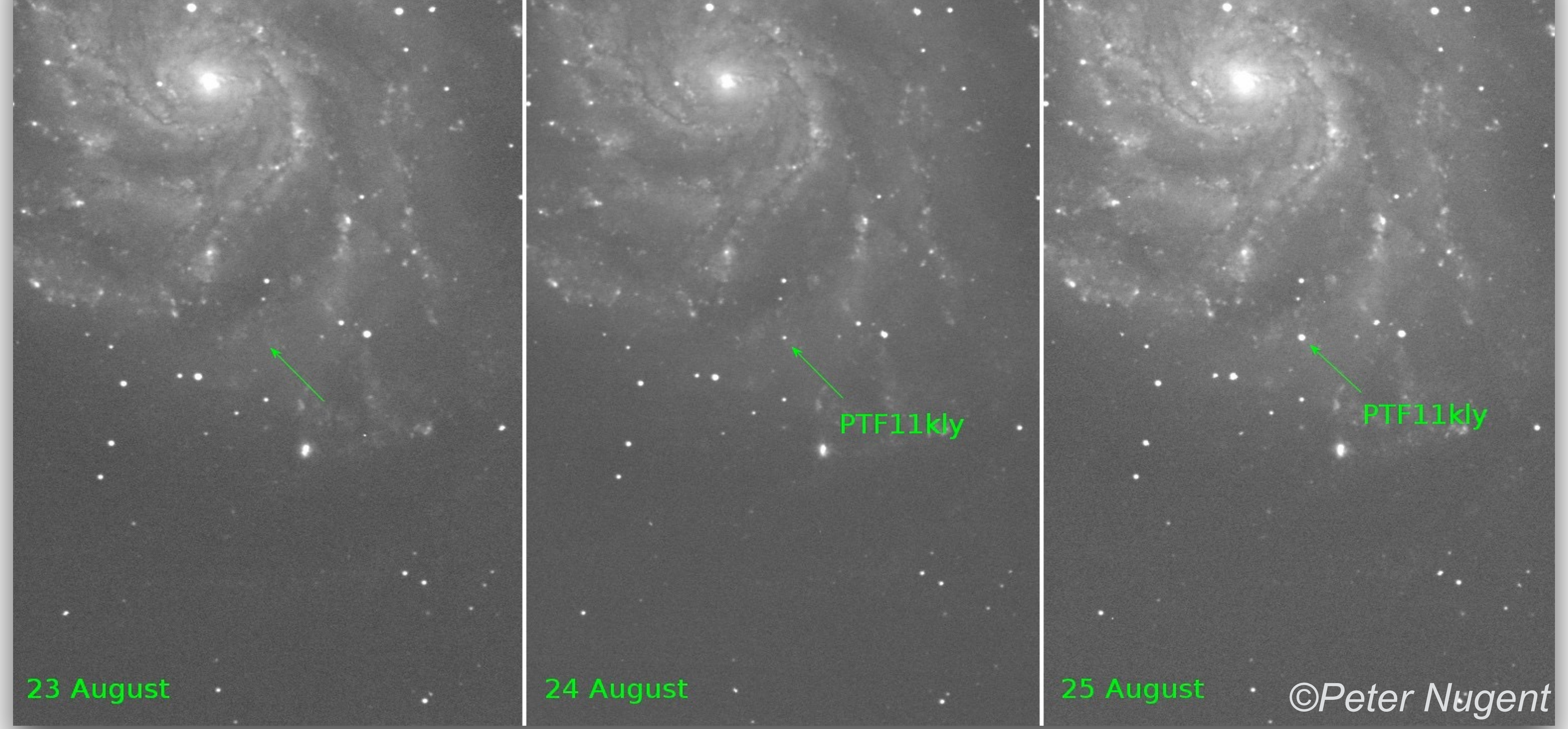


Brink+2012

Better Models…



Brink+2012



Supernova Discovery in the Pinwheel Galaxy (M101)

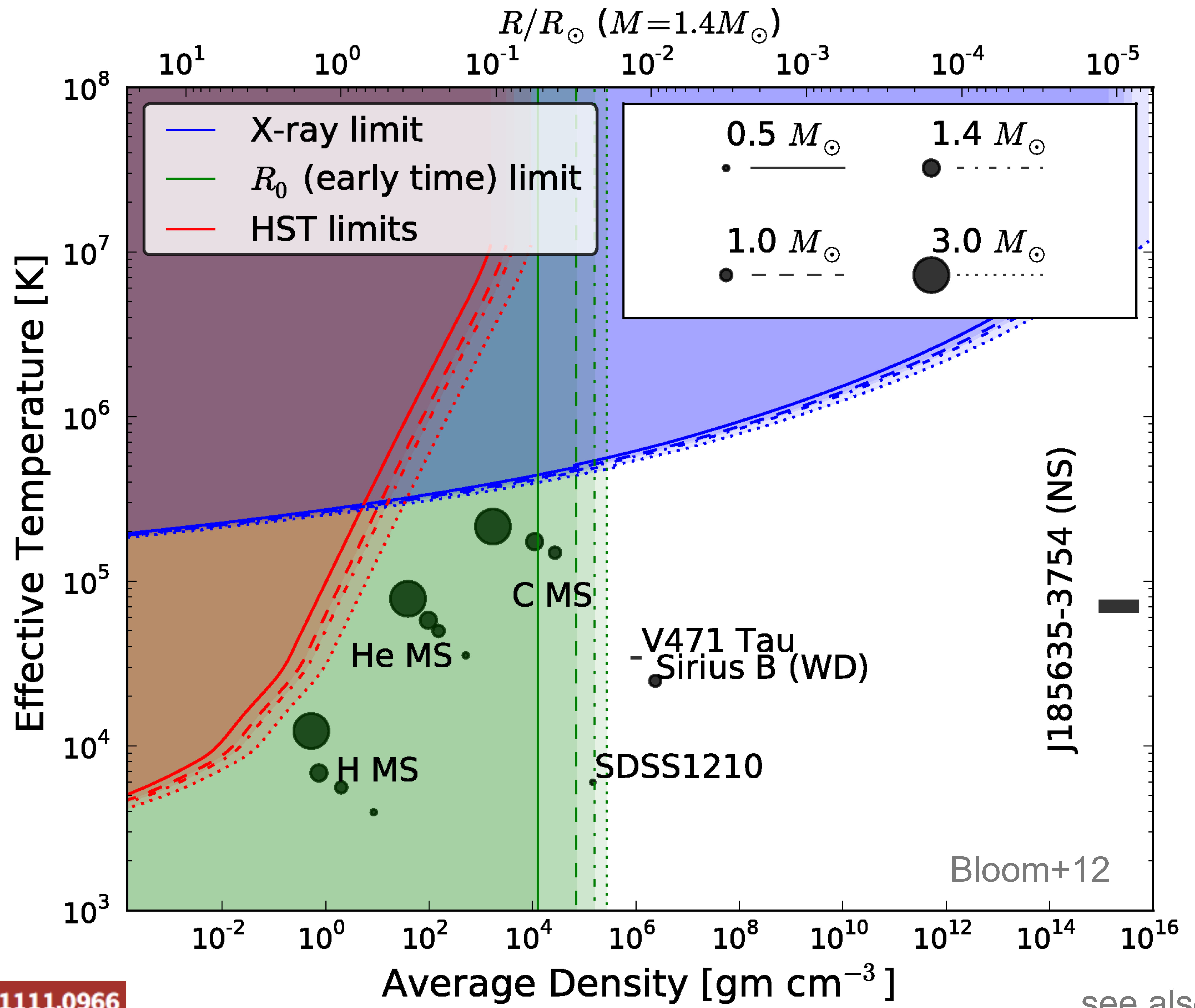
11 hr after explosion

nearest SN Ia in >3 decades

ML-assisted "real-bogus" discovery

arXiv.org > astro-ph > arXiv:1209.3775

Nugent, …, JSB+12, Nature, 1110.6201



arXiv.org > astro-ph > arXiv:1111.0966

see also Nugent+12, Nature

Discovery (& classification) on images is now a cottage industry

Convolutional Neural Networks for Transient Candidate Vetting in Large-Scale Surveys

Fabian Gieseke,[1,2]★ Steven Bloemen,[3,4] Cas van den Bogaard,[1] Tom Heskes,[1]

achine Learning Classification of SDSS Transient Survey Images

L. du Buisson[1,2]★, N. Sivanandam[2]†, Bruce A. Bassett[1,2,3]‡ and M. Smith[4,5]

A transient search using combined human and machine classifications

Darryl E. Wright, Chris J. Lintott, Stephen J. Smartt, Ken W. Smith, Lucy Fortson,

Machine learning for transient discovery in Pan-STARRS1 difference imaging

D. E. Wright, S. J. Smartt, K. W. Smith, P. Miller, R. Kotak, A. Rest, W. S. Burgett, K. C. Chambers, H. Flewelling, K. W. Hodapp ... Show more

Machine Learning Based Real Bogus System for HSC-SSP Moving Object Detecting Pipeline

Hsing-Wen Lin[1], Ying-Tung Chen[2], Jen-Hung Wang[2], Shiang-Yu Wang[2],

Machine-learning Selection of Optical Transient in Subaru/Hyper Suprime-Cam Survey

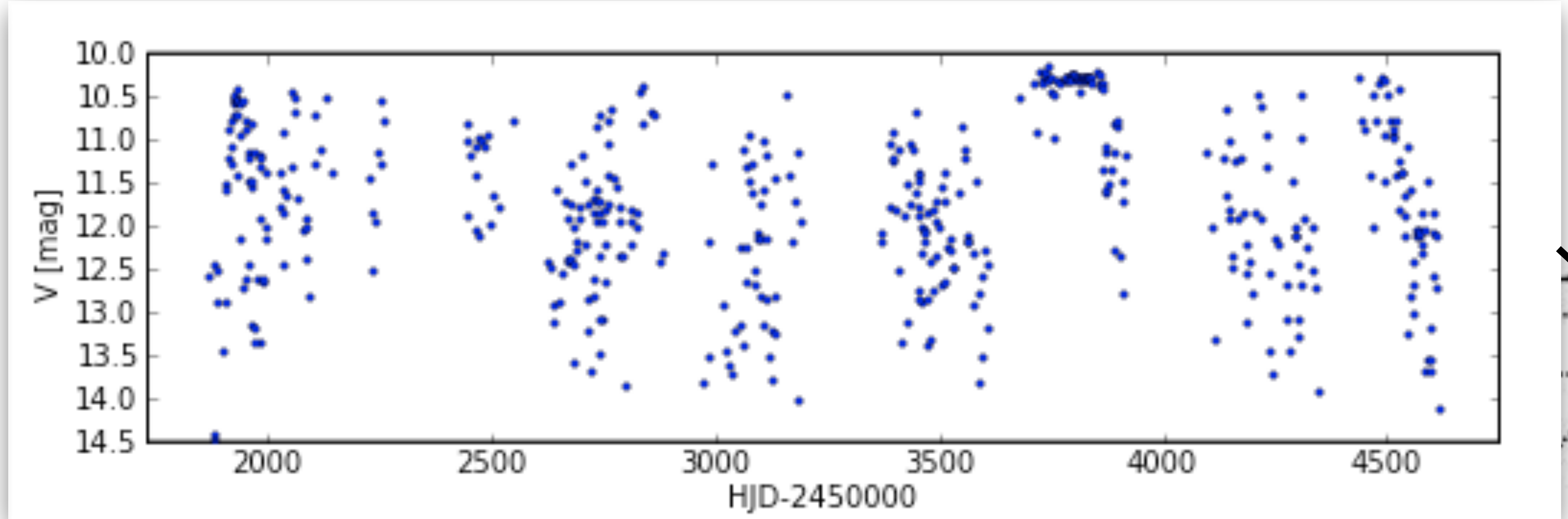
Mikio Morii[1], Shiro Ikeda[1], Nozomu Tominaga[2,3], Masaomi Tanaka[4,3],

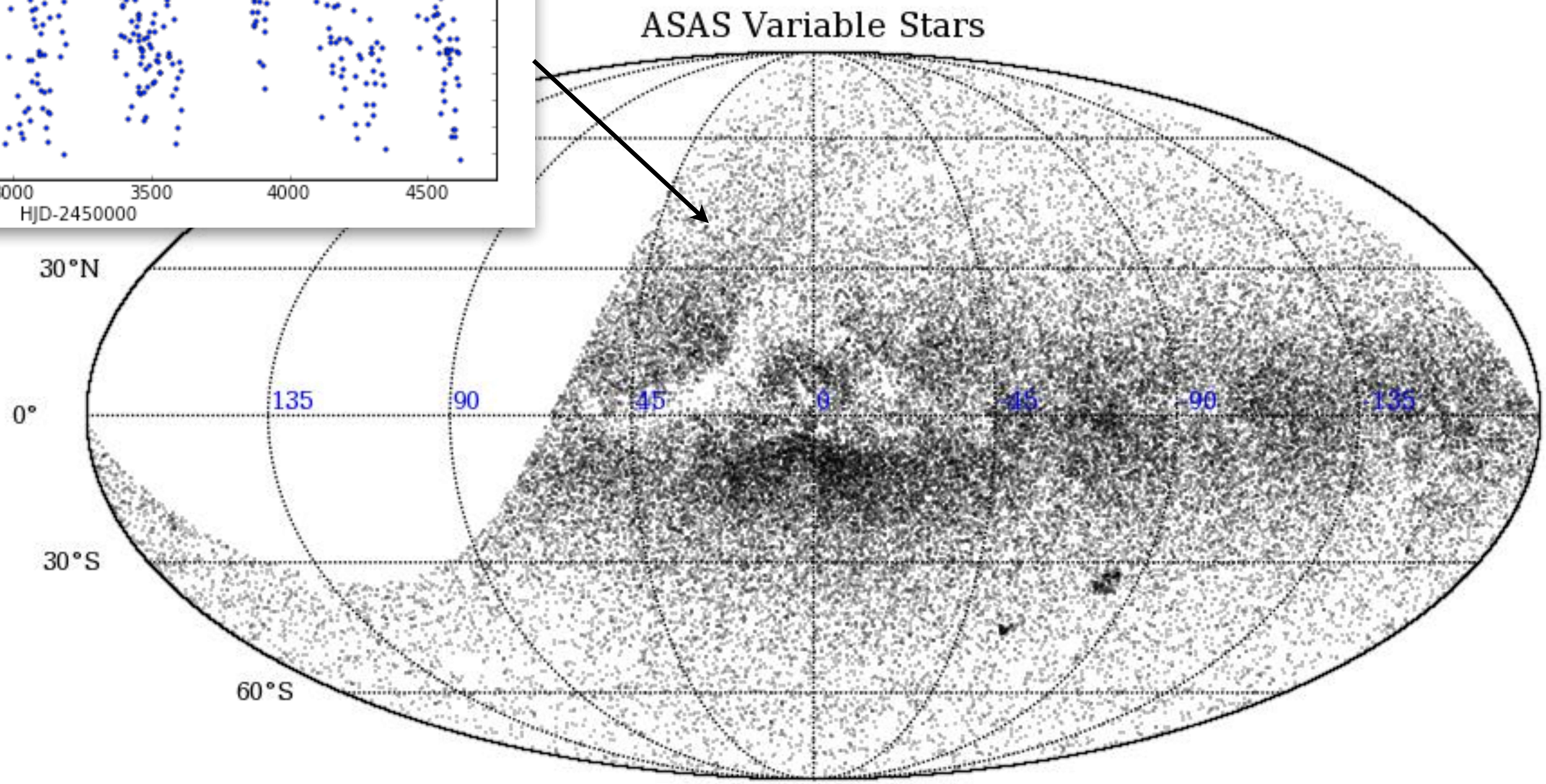
DEEP-HITS: ROTATION INVARIANT CONVOLUTIONAL NEURAL NETWORK FOR TRANSIENT DETECTION

Guillermo Cabrera-Vives[1,2,3,5], Ignacio Reyes[4,1,5], Francisco Förster[2,1], Pablo A. Estévez[4,1] and Juan-Carlos Maureira[2]
Email: gcabrera@dim.uchile.cl

Adapted from D. Goldstein

Variable Star Science





50k variables, 26 classes, 810 with known labels (timeseries, colors)

Richards+11, 12
Also, Amstrong+16 (10k K2 stars)

Self-Supervised (Autoencoder) Recurrent NN

nature astronomy

Letter

A recurrent neural network for classification of unevenly sampled variable stars

Brett Naul, Joshua S. Bloom, Fernando Pérez & Stéfan van der Walt

Nature Astronomy (2017)
doi:10.1038/s41550-017-0321-z
Download Citation
Computer science Stars

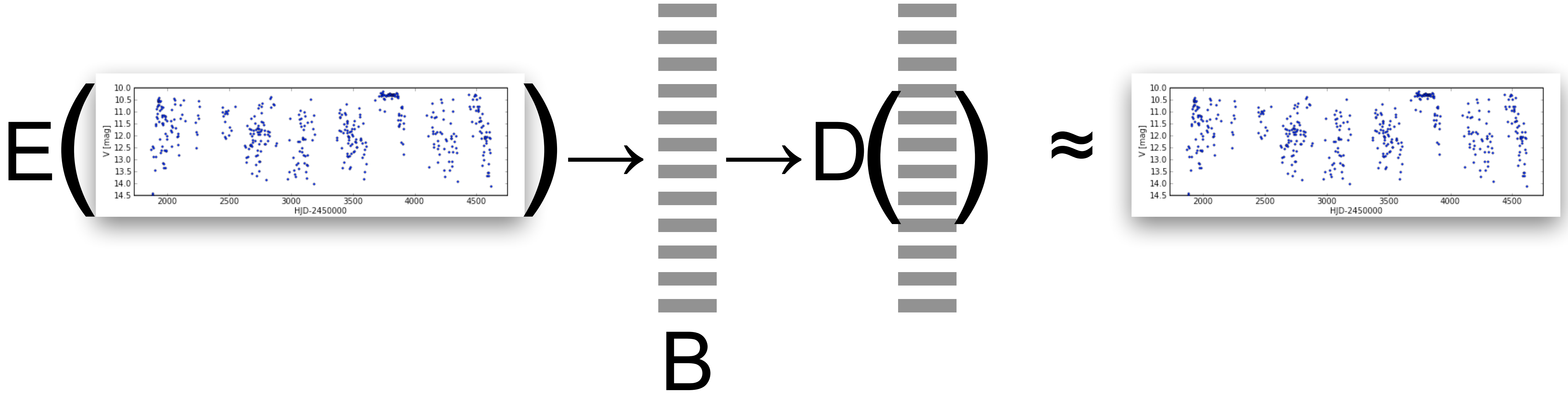
Received: 30 May 2017
Accepted: 24 October 2017
Published online: 27 November 2017


S. van der Walt


F. Peréz

DOI 10.5281/zenodo.1045560

1. AE learn to reproduce irregularly sampled light curves using an information bottleneck (B)



2. Use B as features and learn a traditional classifier (e.g., random forest)

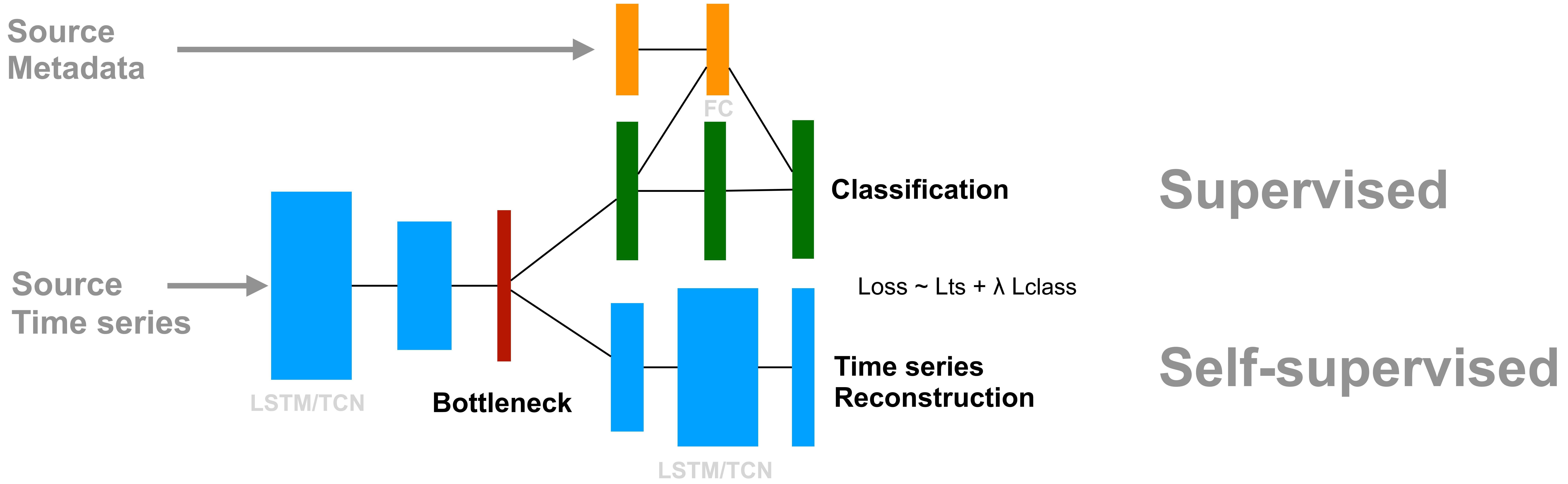
- **self-supervised feature learning** → leverage large corpus of unlabelled light curves

SOTA permutation invariant version: Zhang & Bloom (ICLR20, arxiv:2011.01243)

Self-Supervised (Autoencoder) Recurrent NN

Extensions/Active Research

- Co-training across multiple surveys & multiple bandpasses
- Semi-supervised topology + metadata ("Kitchen Sink")



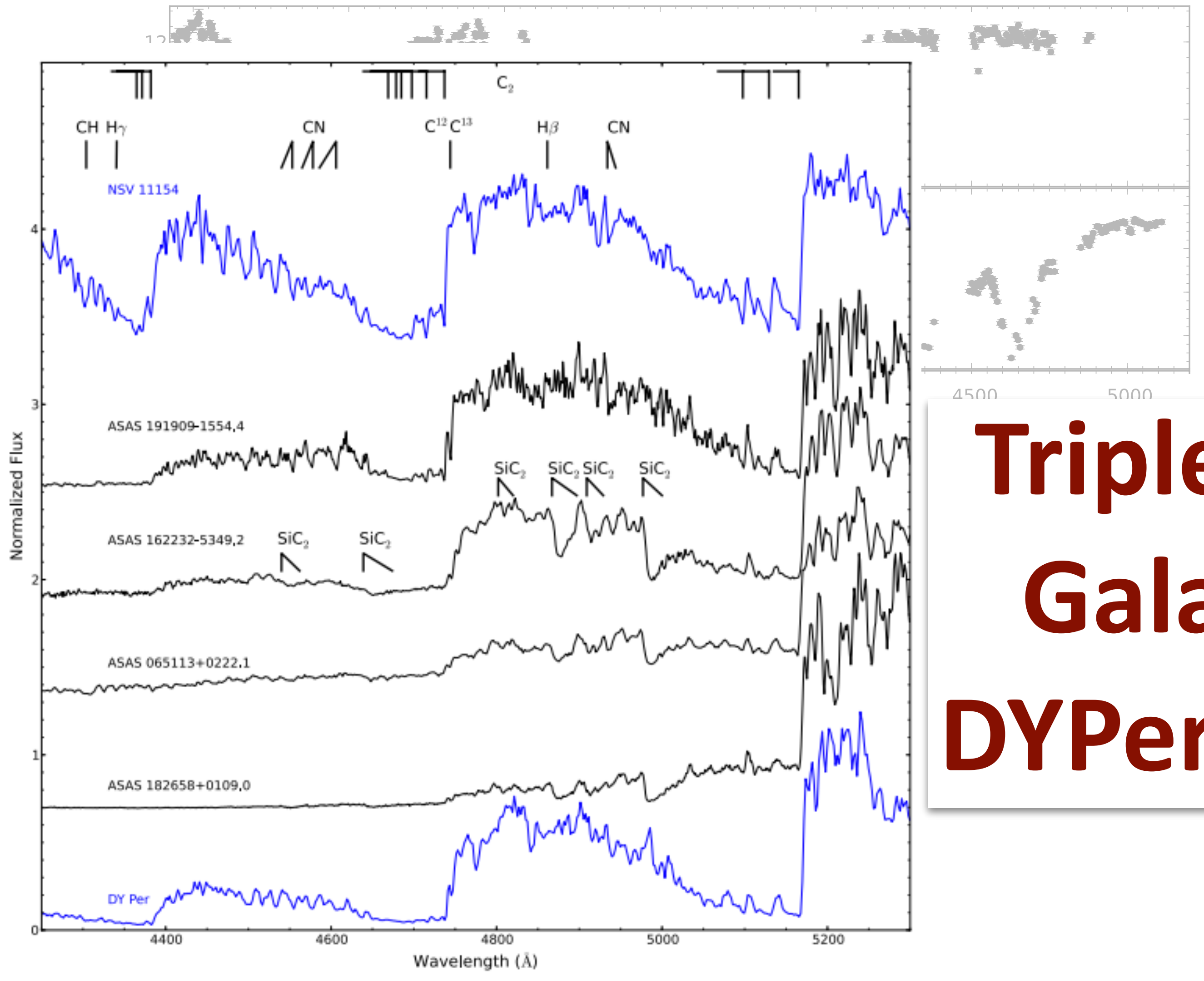
Jamal & JSB 2020, arxiv:2003.08618

Probabilistic Classification Of Variable Stars
→ Inform the use of precious followup resources

15 “RCB/DYP” candidates

106 “DEB” candidates

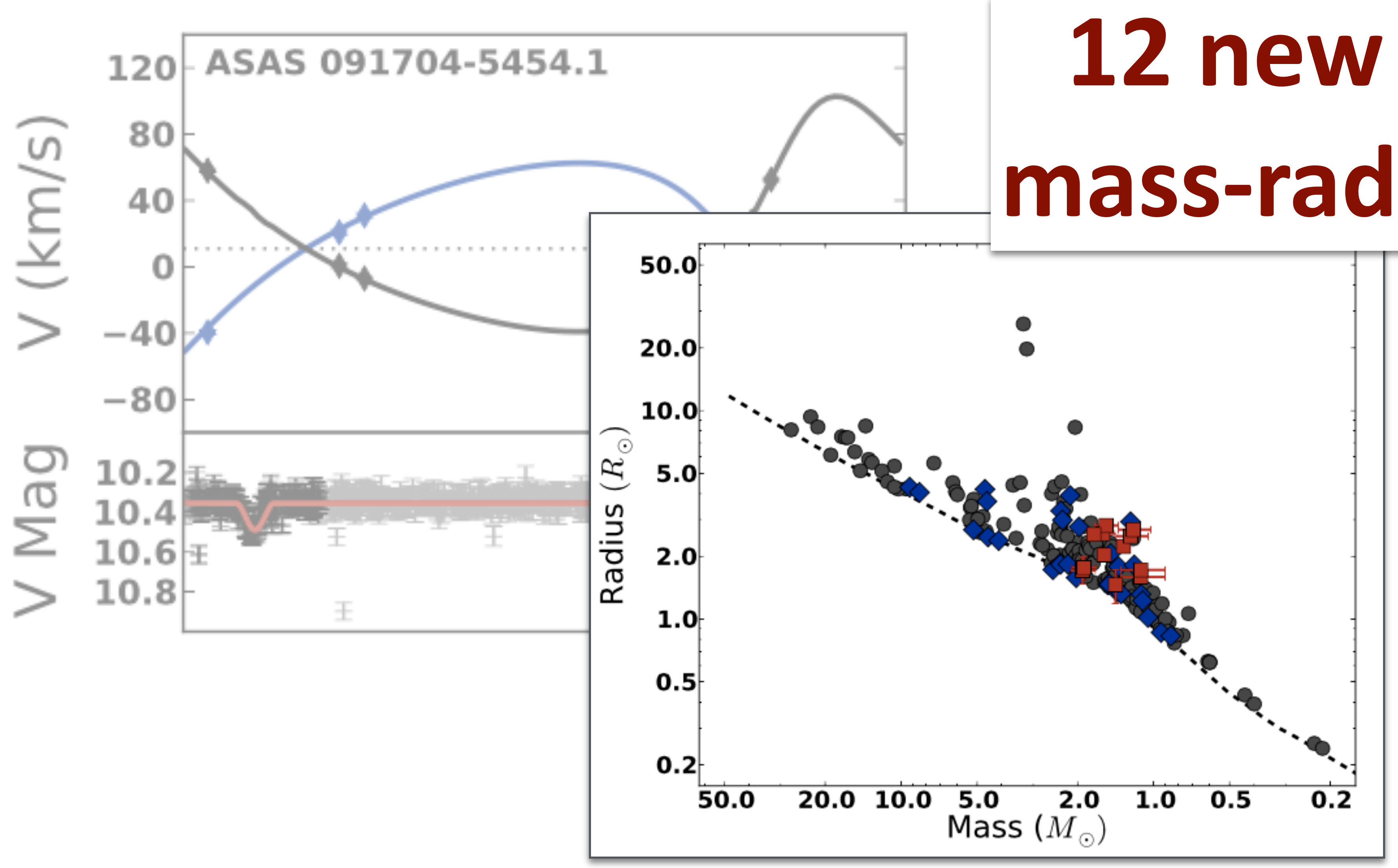
8 new discoveries



Triple # of Galactic DYPer Stars

DISCOVERY OF BRIGHT GALACTIC R CORONAE BOREALIS AND DY PERSEI VARIABLES: RARE GEMS MINED FROM ACVS

Miller, Richards, JSB,..ApJ 2012



12 new mass-radii

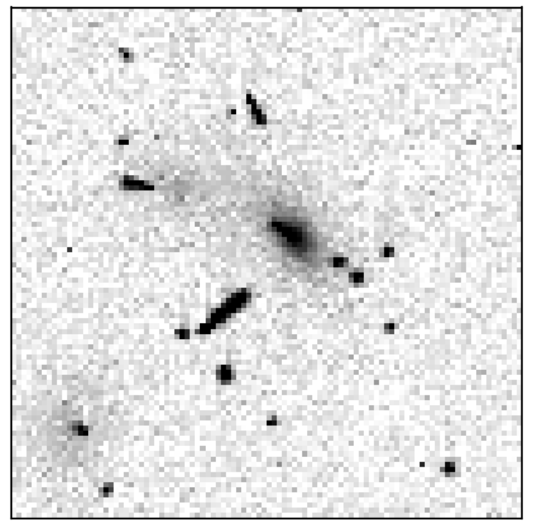
The Highly-Eccentric Detached Eclipsing Binaries in ACVS and MACC

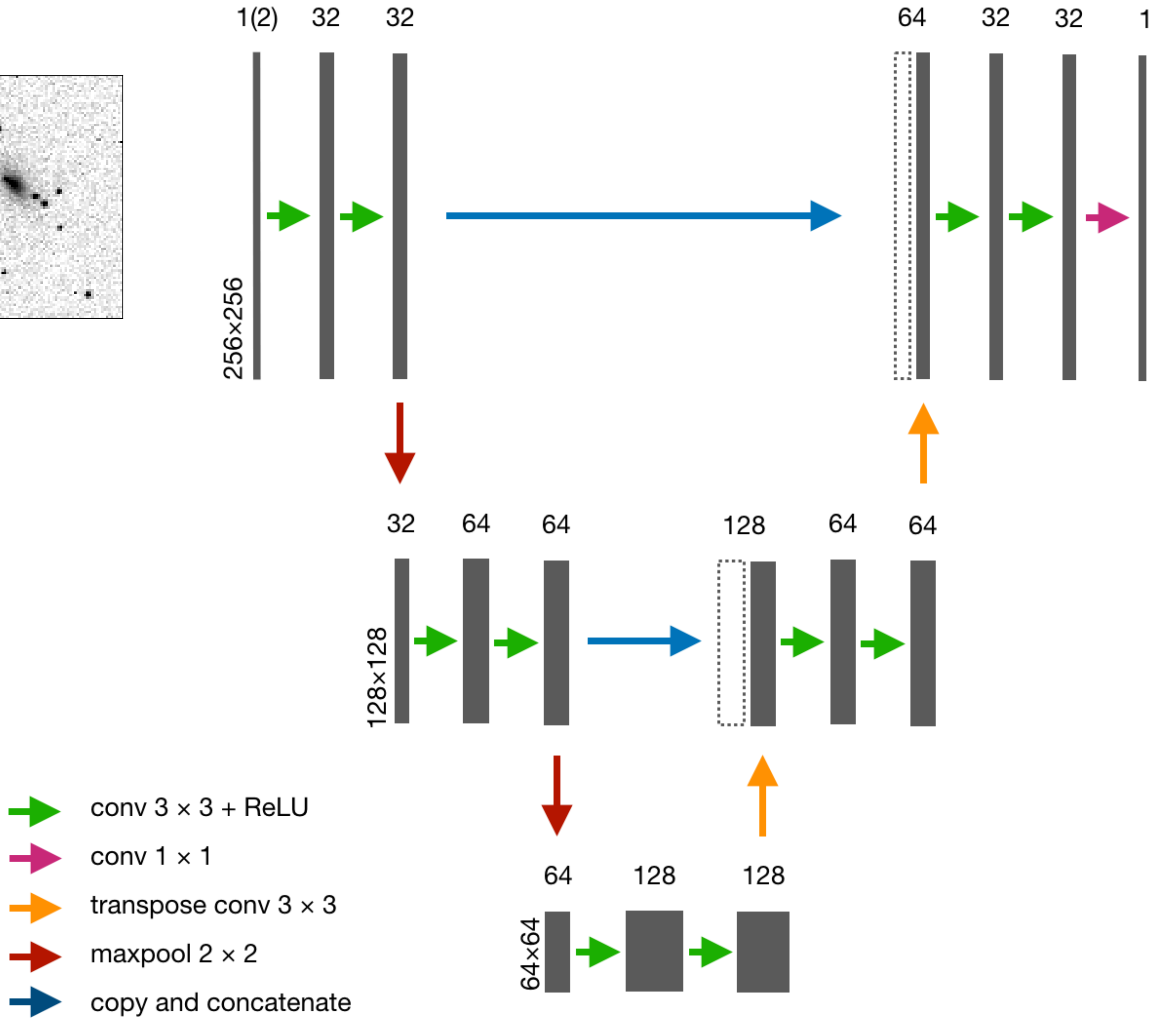
Shivvers,JSB,Richards MNRAS,2014

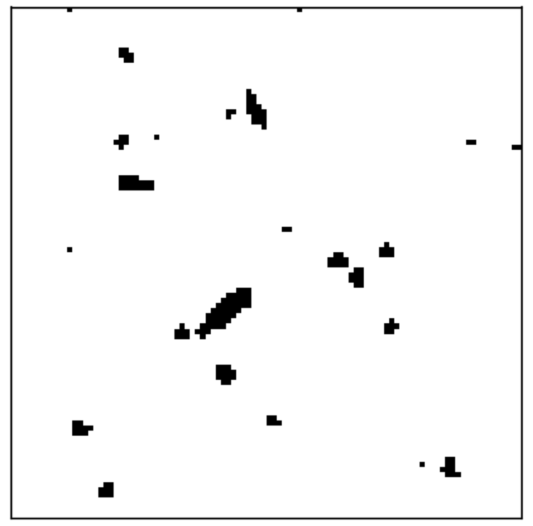
Local Distance Ladder: Spectroscopic Metallicity measurements for RRL, Cepheids, Mira…

Denoising Autoencoders for Imaging Pipelines

Fork me on GitHub







deepCR*: Network Architecture

K. Zhang

* deepCR-mask & deepCR-inpaint

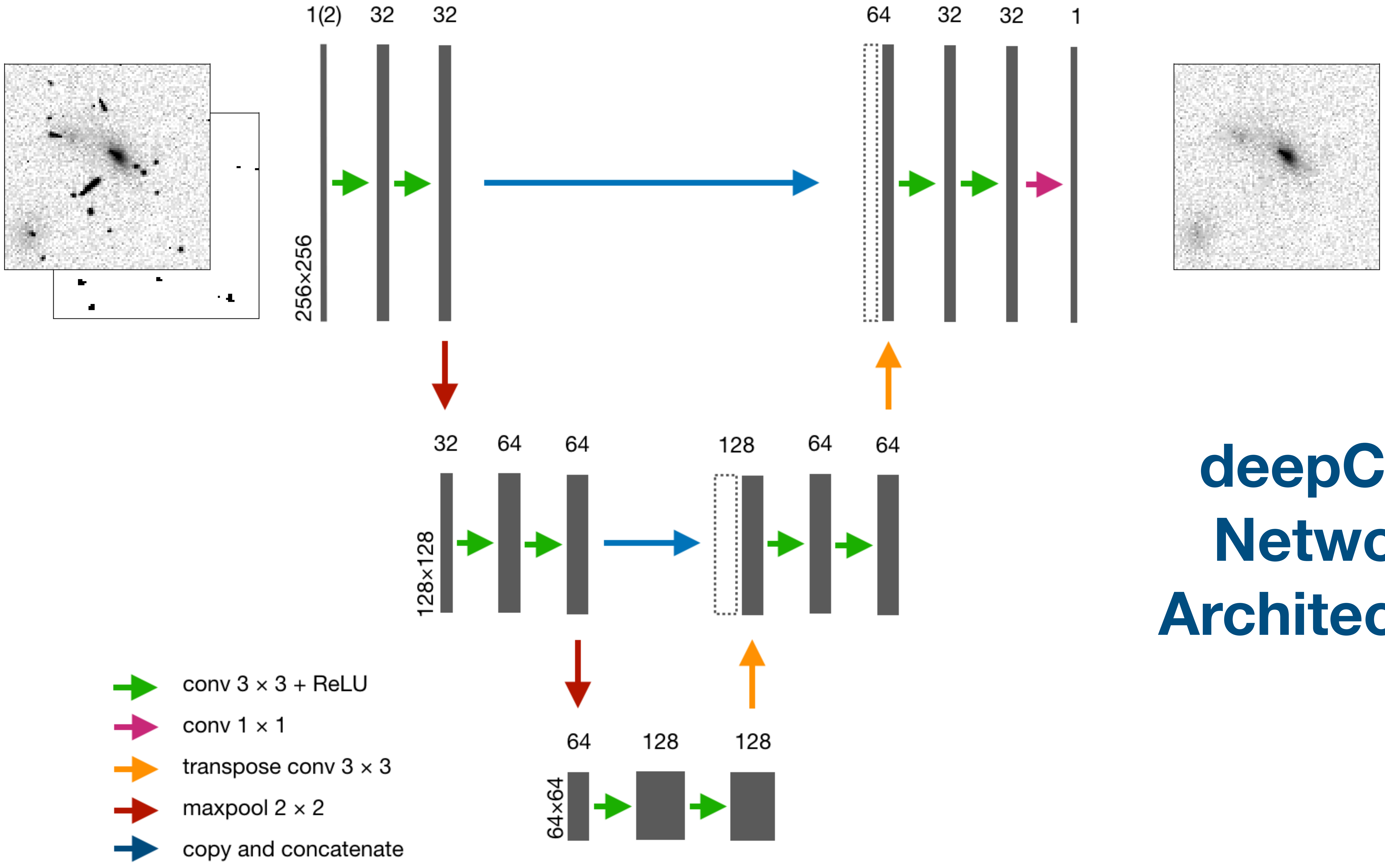
Zhang & Bloom, NeurIPS 1907.09500

JOSS 10.21105/joss.01651 astro-ph 1907.09500

UNet: Ronneberger et al. 2015

Denoising Autoencoders for Imaging Pipelines

Fork me on GitHub



deepCR*: Network Architecture

K. Zhang

* deepCR-mask & deepCR-inpaint

Zhang & Bloom, NeurIPS 1907.09500

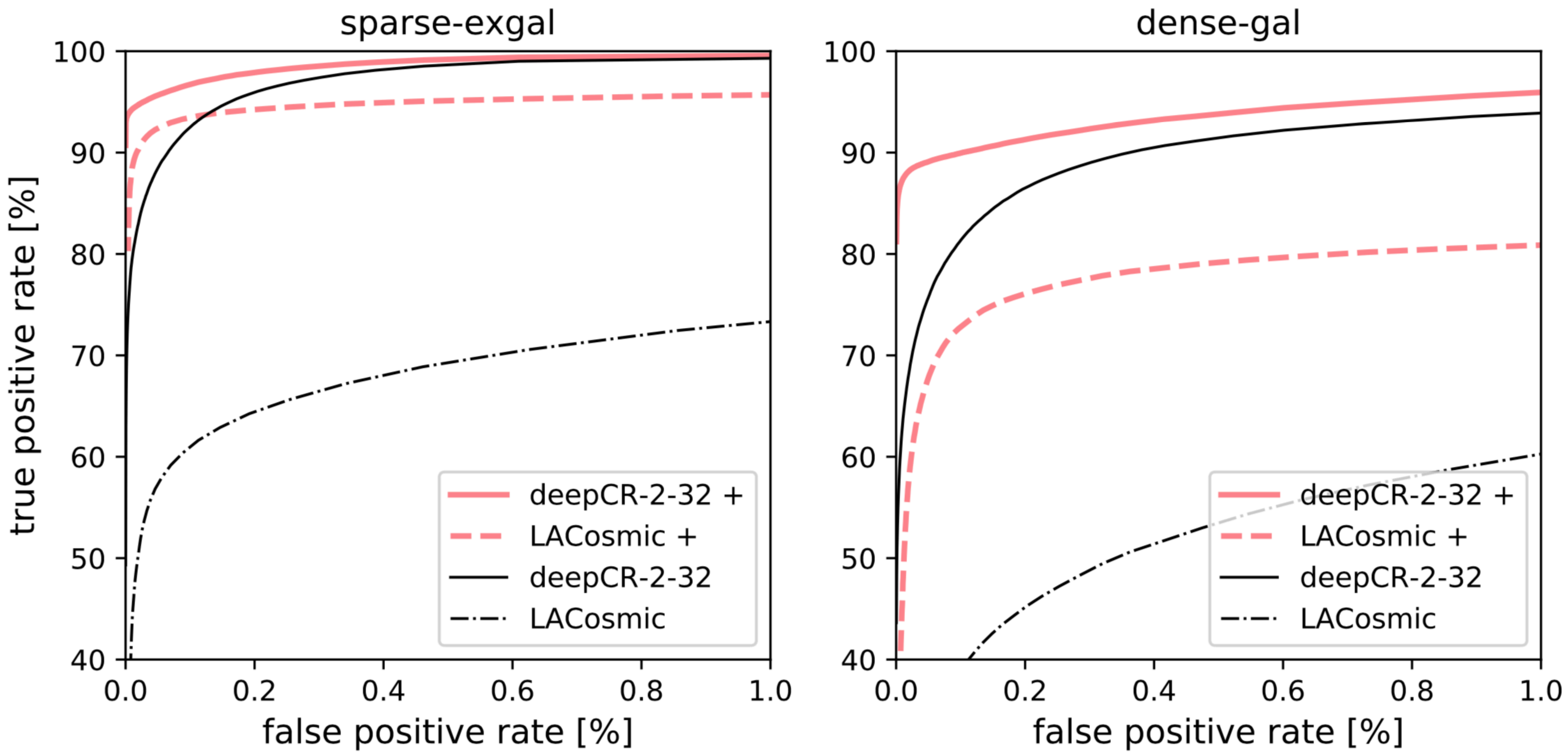
JOSS 10.21105/joss.01651 astro-ph 1907.09500

UNet: Ronneberger et al. 2015

Denoising Autoencoders for Imaging Pipelines

Denoising Autoencoders for Imaging Pipelines

deepCR-mask: better*, faster



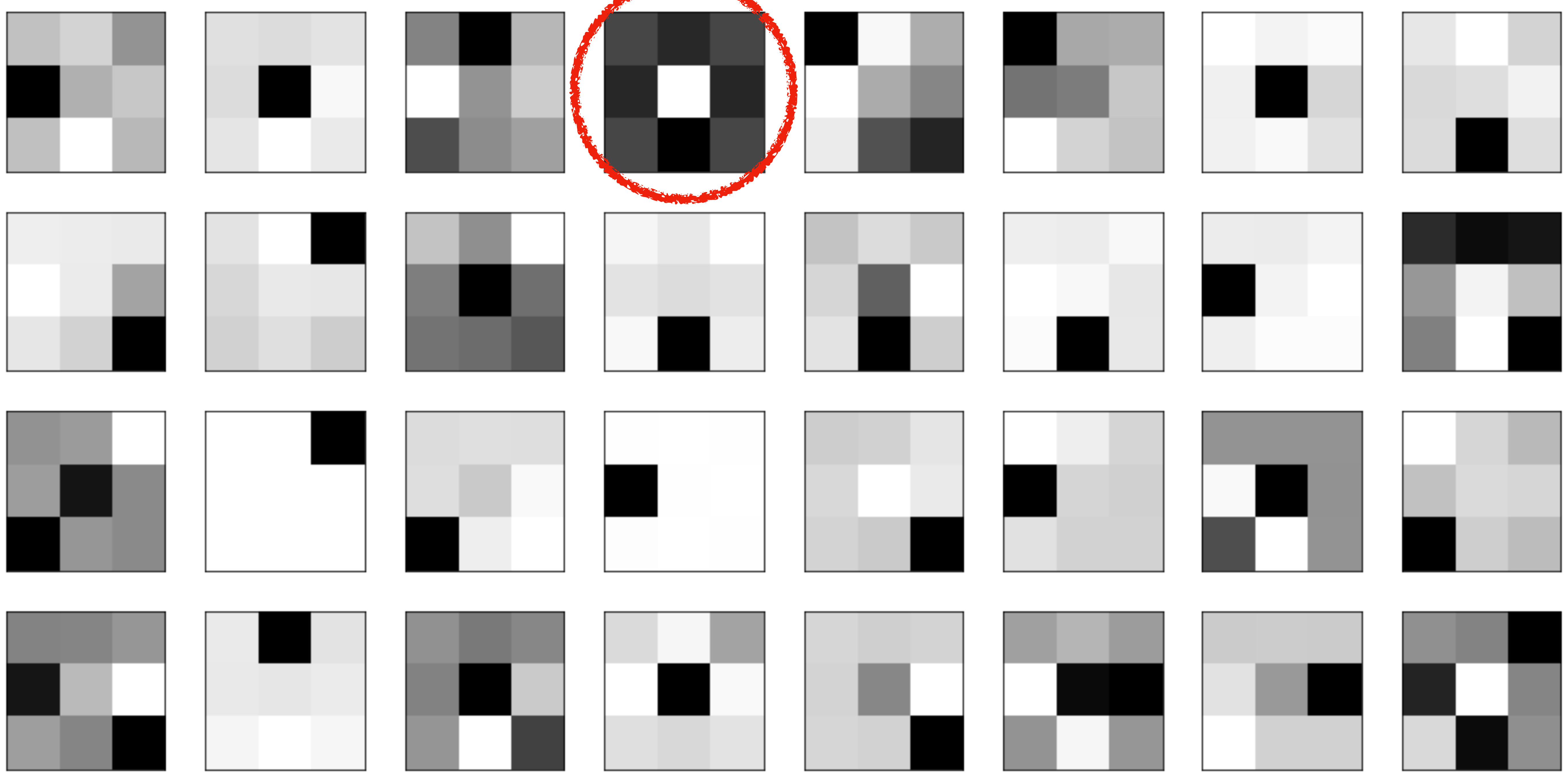
Model	sparse-exgal TPR (0.02%)	TPR (0.1%)	dense-gal TPR (0.02%)	TPR (0.1%)	Time (CPU)	Time (GPU)
deepCR-2-4	82.8% (97.9%)	92.3% (98.4%)	58.1% (88.5%)	74.7% (91.8%)	**1.4s**	**0.3s**
deepCR-2-32	**83.0%** (94.7%)	**92.4%** (96.7%)	**68.3%** (88.1%)	**81.3%** (89.9%)	7.7s	0.2s
LACosmic	50.4% (84.4%)	60.9% (89.7%)	24.3% (42.0%)	38.9% (64.5%)	13.2s	-

* at least on Hubble Space Telescope ACS/WFC

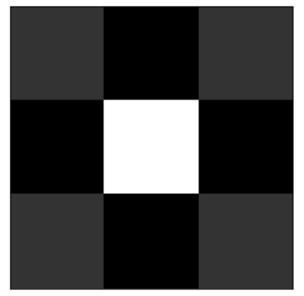
Denoising Autoencoders for Imaging Pipelines

Convolution Filters Learned in 1st layer

deepCR-mask



LACosmic



Current State of the Art Approach

Physics Informed ML

"although neural networks only work well for an exponentially tiny fraction of all possible inputs, the laws of physics are such that the data sets we care about for machine learning are also drawn from an exponentially tiny fraction of all imaginable data sets…"

"Why does deep and cheap learning work so well?"
Lin, Tegmark, Rolnick arXiv:1608.08225 (2017)

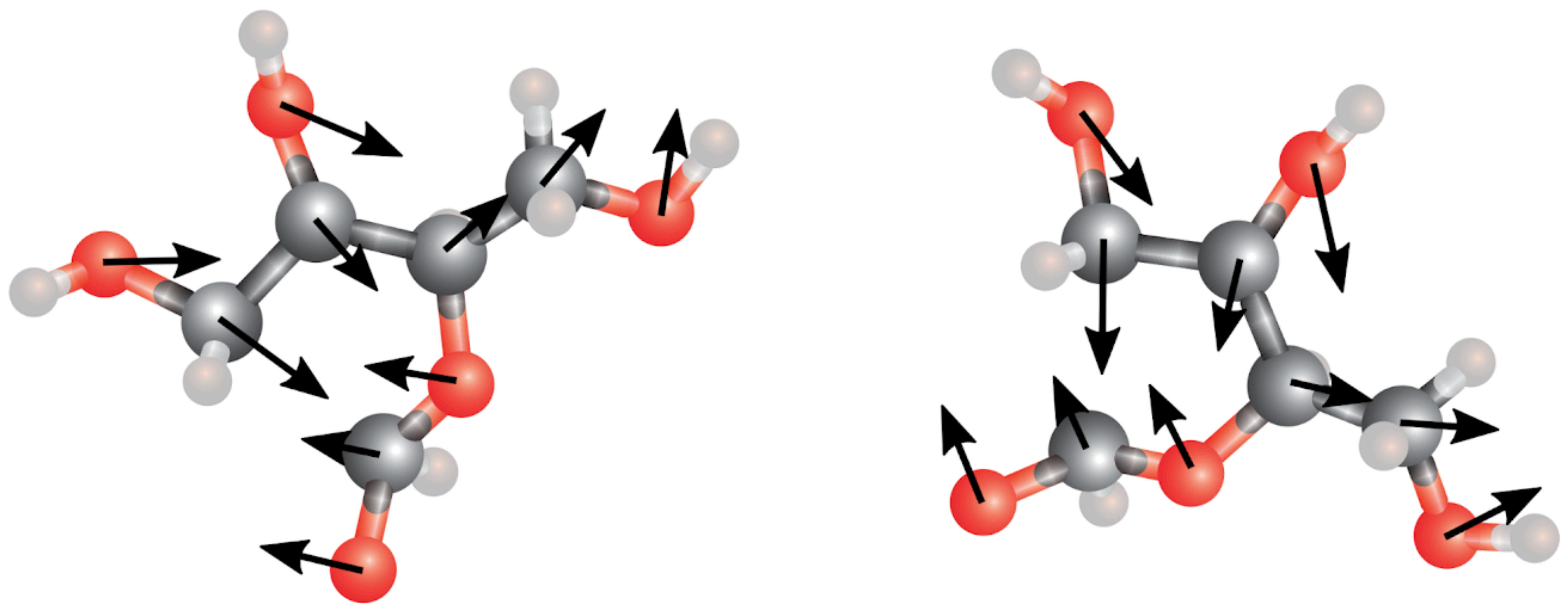
Impart/impose/imbue physical constraints into architecture

- Computer vision: e.g., Spatial Transformer Network, GVNN (s03 layer Euler,…)

  Jaderberg+1506.02025; Handa+1607.07405
- High-energy physics: “QCD-Aware Recursive NN for Jet Physics”

  Louppe+ 1702.00748
- Quantum Chemistry: “Ab-Initio Solution of the Many-Electron Schrödinger Equation with Deep Neural Networks”

  Pfau+ 1909.02487



Euclidean Neural Networks
rotation-, translation-, & permutation-equivariant convolutional neural networks for 3D point clouds for emulating *ab initio* calculations & generating atomic geometries

Tess Smidt

cf. "Machine learning and the physical sciences” Carleo+ 1903.10563

Impart/impose/imbue physical constraints into architecture

- Computer vision: e.g., Spatial Transformer Network, GVNN (s03 layer Euler,…)

  Jaderberg+1506.02025; Handa+1607.07405

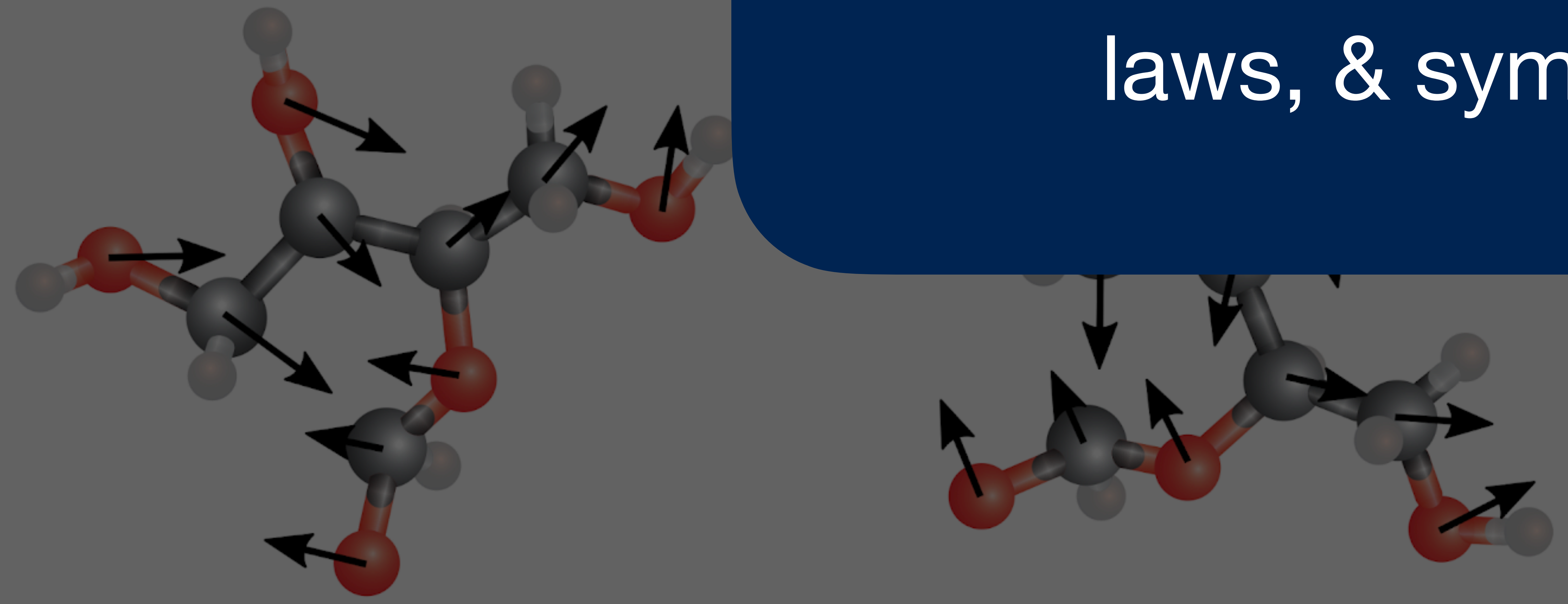
- High-energy physics: “QC

  Louppe+ 1702.00748

- Quantum Chemistry: “A
  Schrödinger Equation w

  Pfau+ 1909.02487

Challenge: Find data embeddings & network architectures that conform to known taxonomies, conservation laws, & symmetries



ral Networks
ation-, & permutation-
volutional neural
networks for 3D point clouds for emulating *ab initio* calculations & generating atomic geometries

Tess Smidt

cf. "Machine learning and the physical sciences” Carleo+ 1903.10563

Wrapping in a Ring: Polar Coordinate Convolution

Cyclic-Permutation Invariant Networks for Modeling Periodic Time Series

Keming Zhang *
Department of Astronomy
University of California at Berkeley
Berkeley, CA 94720, USA
kemingz@berkeley.edu

Joshua S. Bloom
Department of Astronomy
University of California at Berkeley
Berkeley, CA 94720, USA
joshbloom@berkeley.edu

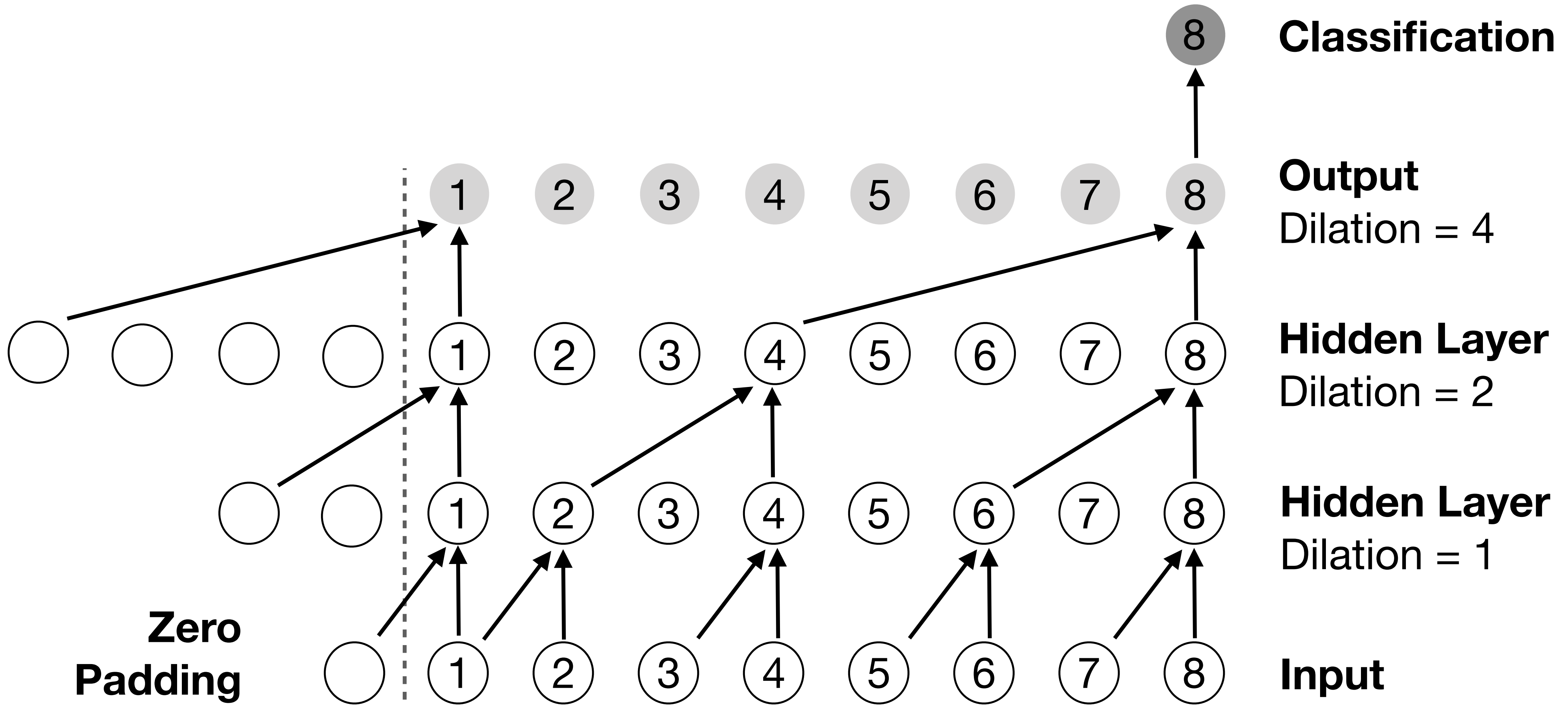
ICLR2020 Workshop

Abstract

Recurrent neural networks (RNNs) are sub-optimal for modeling periodic time series data which is common in the physical sciences, because their acyclic topology forbids explicit modeling of periodicity. In this paper, we present novel cyclic-permutation invariant networks, where the symmetry of periodicity is explicitly embedded in the network architecture by performing convolutions in polar coordinates, instead of Cartesian coordinates. We describe two specific implementations here, one named invariant Temporal Inception Networks (iTINs), which is based on 1-D dilated convolutions, and the other the invariant ResNet (iResNet). Applied to the classification of periodic variable star light curves, a physically relevant exemplar, the iResNets achieve state-of-the-art accuracy. The methodology we introduce is applicable to a wide range of science domains where periodic data abounds due to physical symmetries, and is highly scalable on modern GPU devices.

1 Introduction

Strictly periodic data is common in the physical sciences where periodicity occurs both spatially and temporally. Neural networks (NNs) for which invariances arising from periodicity are explicitly considered either in the input features or the loss function, have been previously applied to particle



Cartesian Coordinate Temporal Convolutional Network (TCN)

TCN: Bai et al. 2018

Wrapping in a Ring: Polar Coordinate Convolution

CYCLIC-PERMUTATION INVARIANT NETWORKS FOR MODELING PERIODIC TIME SERIES

Keming Zhang *
Department of Astronomy
University of California at Berkeley
Berkeley, CA 94720, USA
kemingz@berkeley.edu

Joshua S. Bloom
Department of Astronomy
University of California at Berkeley
Berkeley, CA 94720, USA
joshbloom@berkeley.edu

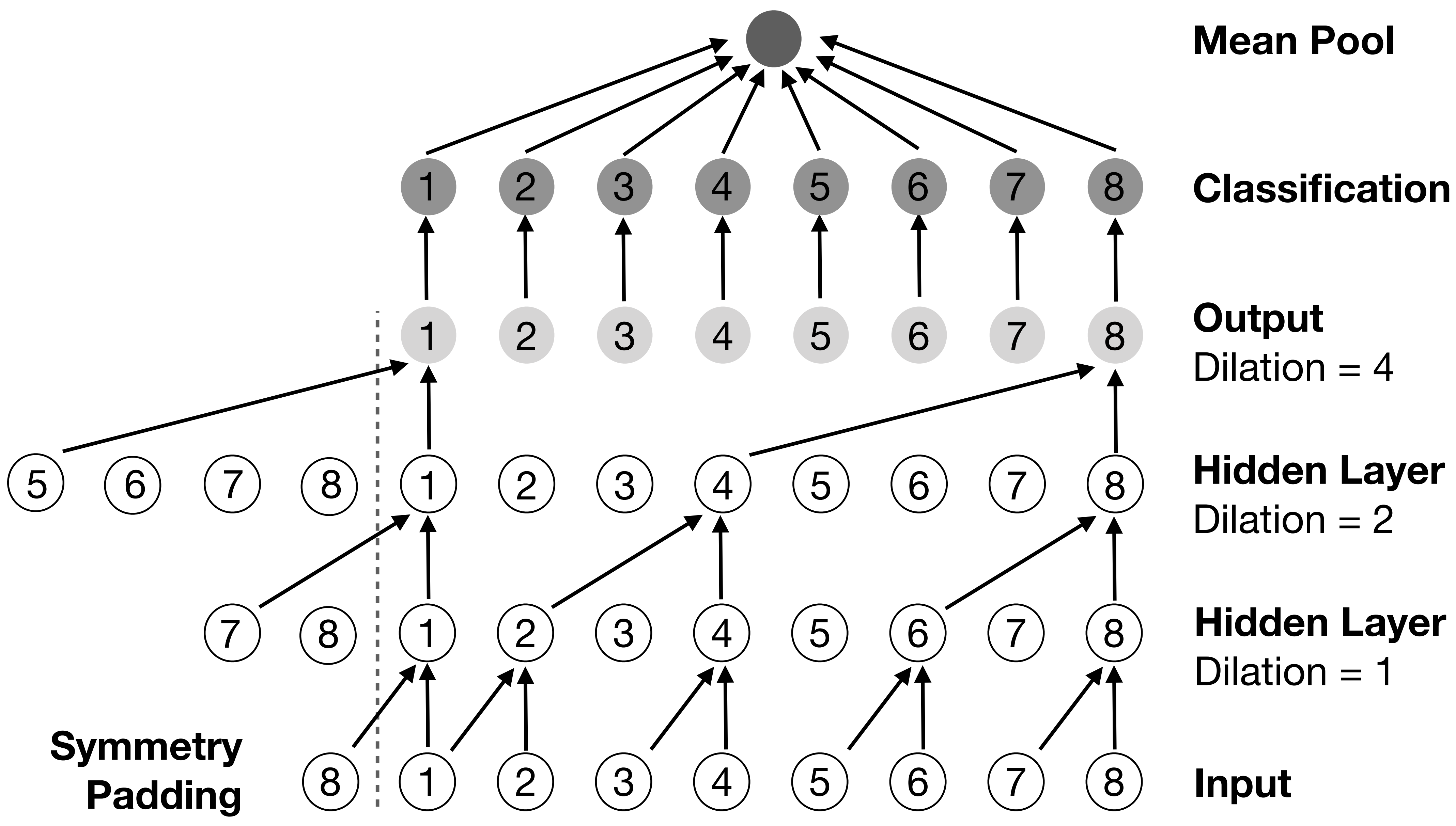
ICLR2020 Workshop

ABSTRACT

Recurrent neural networks (RNNs) are sub-optimal for modeling periodic time series data which is common in the physical sciences, because their acyclic topology forbids explicit modeling of periodicity. In this paper, we present novel cyclic-permutation invariant networks, where the symmetry of periodicity is explicitly embedded in the network architecture by performing convolutions in polar coordinates, instead of Cartesian coordinates. We describe two specific implementations here, one named invariant Temporal Inception Networks (iTINs), which is based on 1-D dilated convolutions, and the other the invariant ResNet (iResNet). Applied to the classification of periodic variable star light curves, a physically relevant exemplar, the iResNets achieve state-of-the-art accuracy. The methodology we introduce is applicable to a wide range of science domains where periodic data abounds due to physical symmetries, and is highly scalable on modern GPU devices.

1 INTRODUCTION

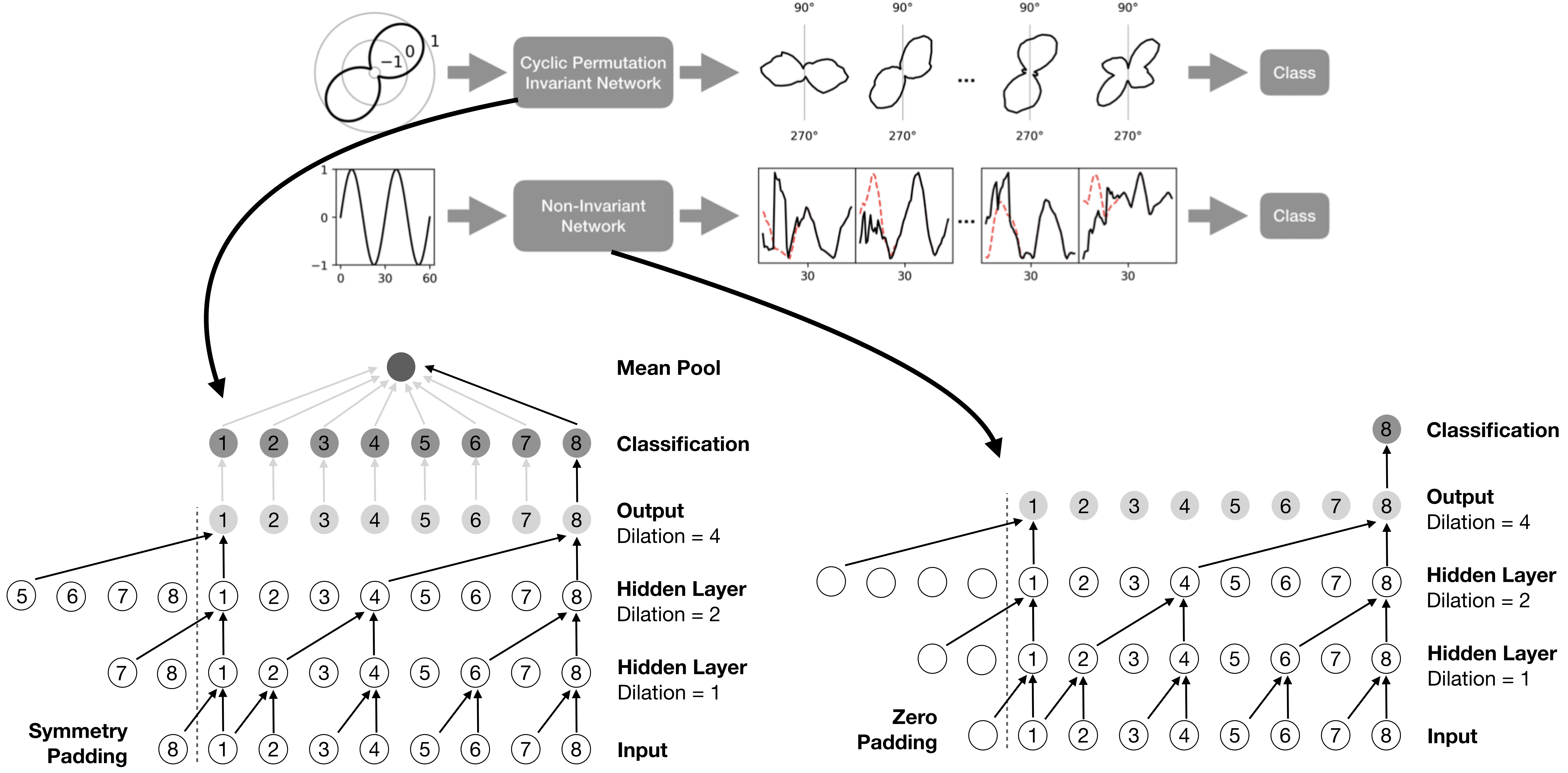
Strictly periodic data is common in the physical sciences where periodicity occurs both spatially and temporally. Neural networks (NNs) for which invariances arising from periodicity are explicitly considered either in the input features or the loss function, have been previously applied to particle



Polar Coordinate / Invariant Temporal Convolutional Network (iTCN)

TCN: Bai et al. 2018

Wrapping in a Ring: Polar Coordinate Convolution



Polar Coordinate / invariant Temporal Convolutional Network (iTCN)

Cartesian Coordinate Temporal Convolutional Network (TCN)

Results: Periodic Variable Star Classification

	MACHO	OGLE-III	ASAS-SN
Size	130,484	187,571	418,207
Number of Class	8	8	15
Sequence Length	128	128	128

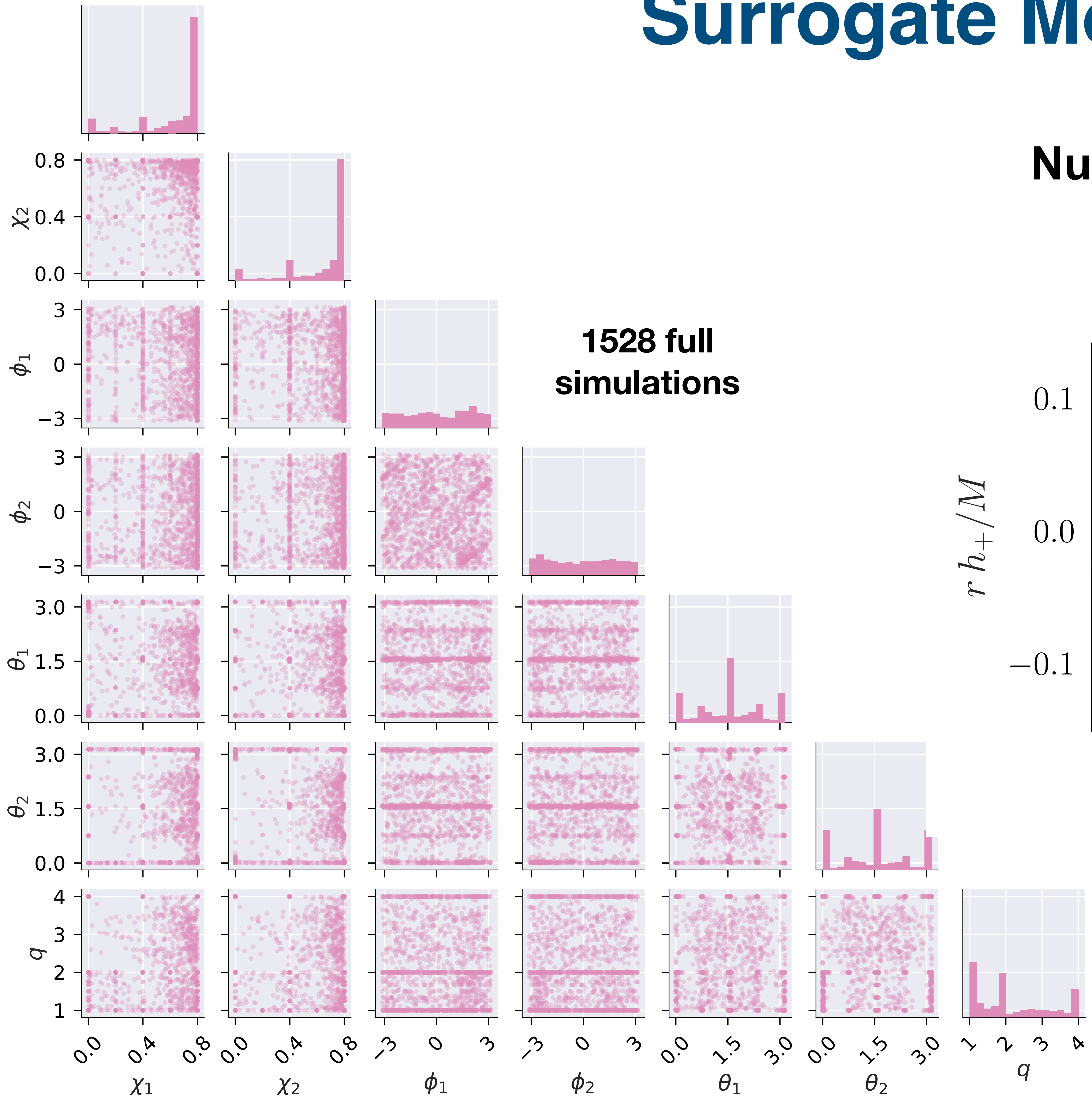
Model	MACHO	OGLE-III	ASAS-SN
iResNet	**92.8 %**	**97.8 %**	**93.7 %**
ResNet	92.7 %	97.6 %	93.3 %
—[1]	$(-0.07\%^{+0.01\%}_{-0.03\%})$	$(-0.16\%^{+0.01\%}_{-0.04\%})$	$(-0.44\%^{+0.14\%}_{-0.04\%})$
iTIN	**92.8 %**	97.7 %	**93.7 %**
TIN	92.4 %	97.4 %	93.3 %
—[1]	$(-0.41\%^{+0.04\%}_{-0.01\%})$	$(-0.30\%^{+0.16\%}_{-0.05\%})$	$(-0.37\%^{+0.08\%}_{-0.05\%})$
iTCN	92.7 %	97.7 %	93.6 %
TCN*	92.1 %	97.3 %	93.0 %
—[1]	$(-0.54\%^{+0.02\%}_{-0.13\%})$	$(-0.32\%^{+0.05\%}_{-0.02\%})$	$(-0.67\%^{+0.03\%}_{-0.04\%})$
GRU*	92.5 %	97.5 %	93.3 %
—[2]	$(-0.33\%^{+0.06\%}_{-0.12\%})$	$(-0.29\%^{+0.04\%}_{-0.02\%})$	$(-0.36\%^{+0.09\%}_{-0.11\%})$
LSTM*	92.4 %	97.2 %	93.2 %
—[2]	$(-0.43\%^{+0.17\%}_{-0.24\%})$	$(-0.65\%^{+0.14\%}_{-0.11\%})$	$(-0.52\%^{+0.06\%}_{-0.15\%})$

[1] Compared to the invariant version of the same network

[2] Compared to the best performing network

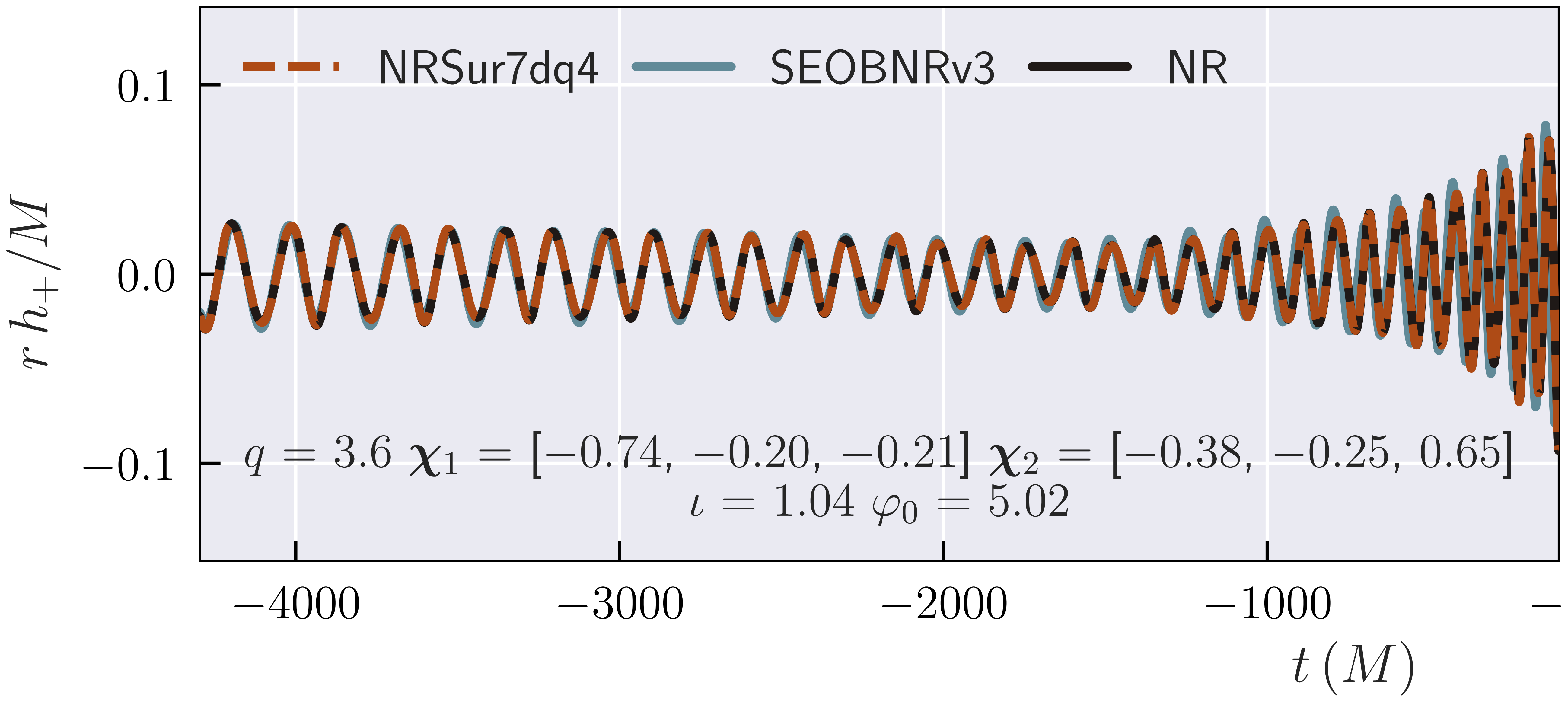
*Has been previously applied to variable star classification

Surrogate Modeling



1528 full simulations

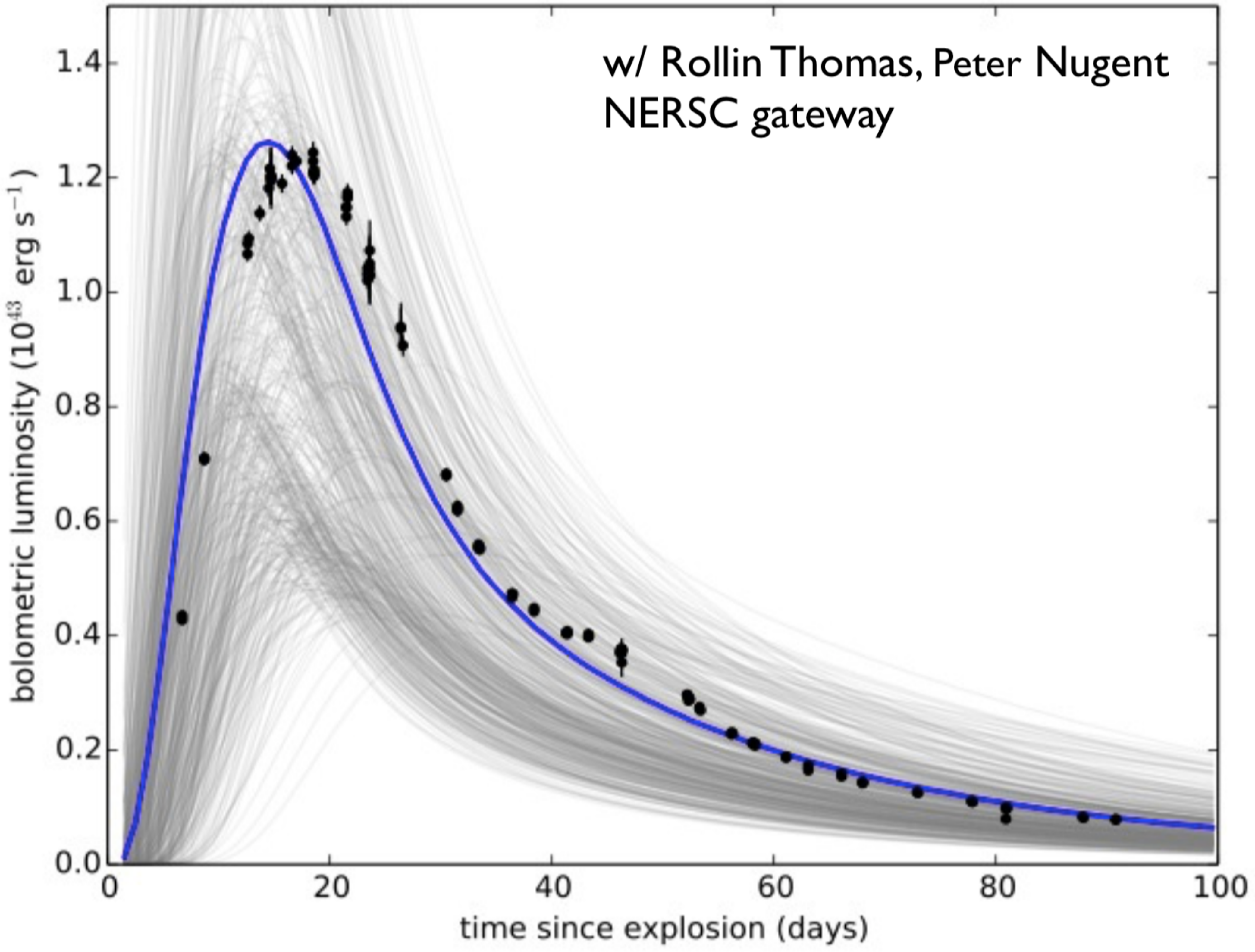
Numerical Relativity calculations of black hole merger waveforms



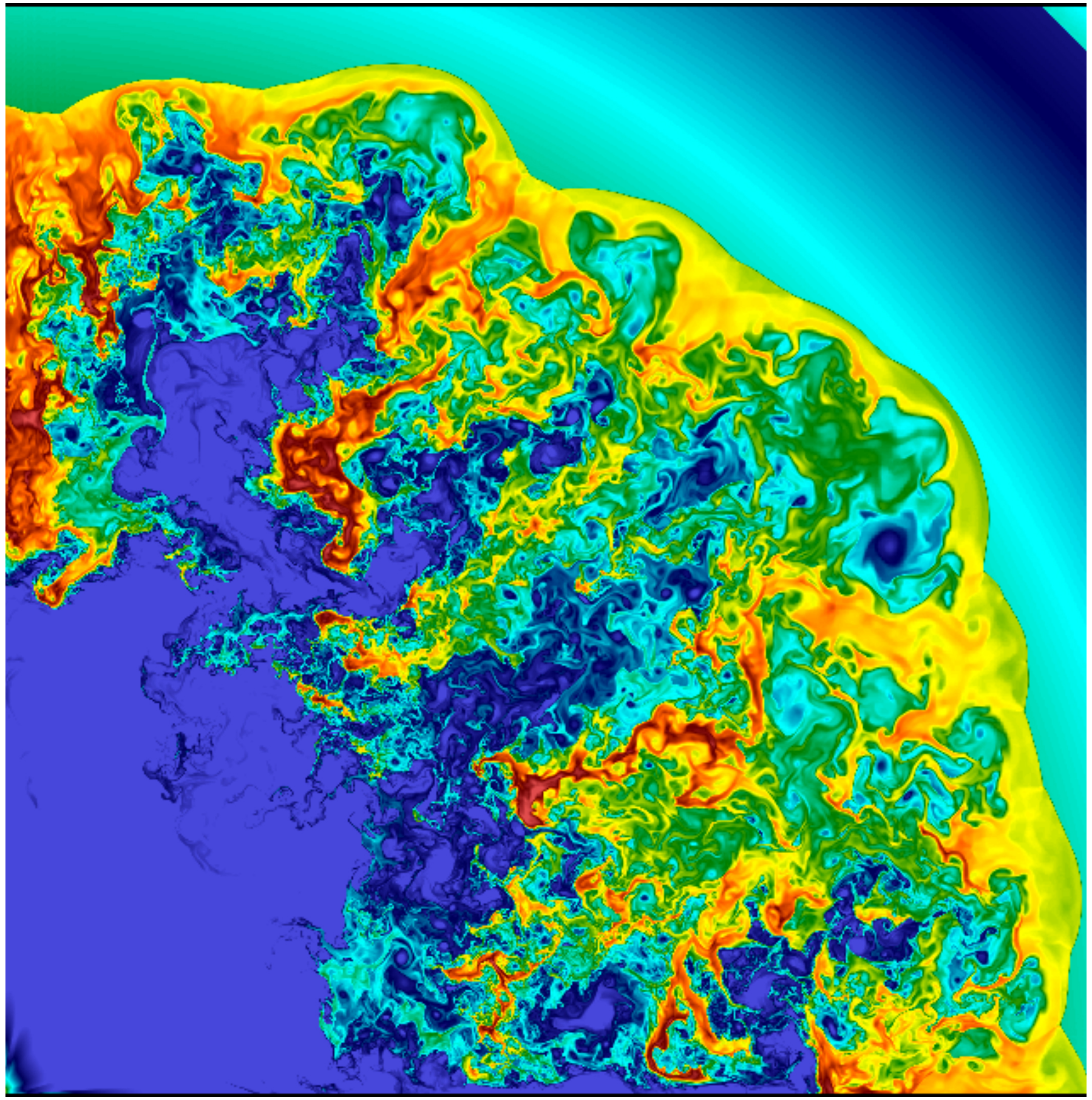
"Surrogate models for precessing binary black hole simulations with unequal masse"

Varma+ 1905.09300

Surrogate Modeling



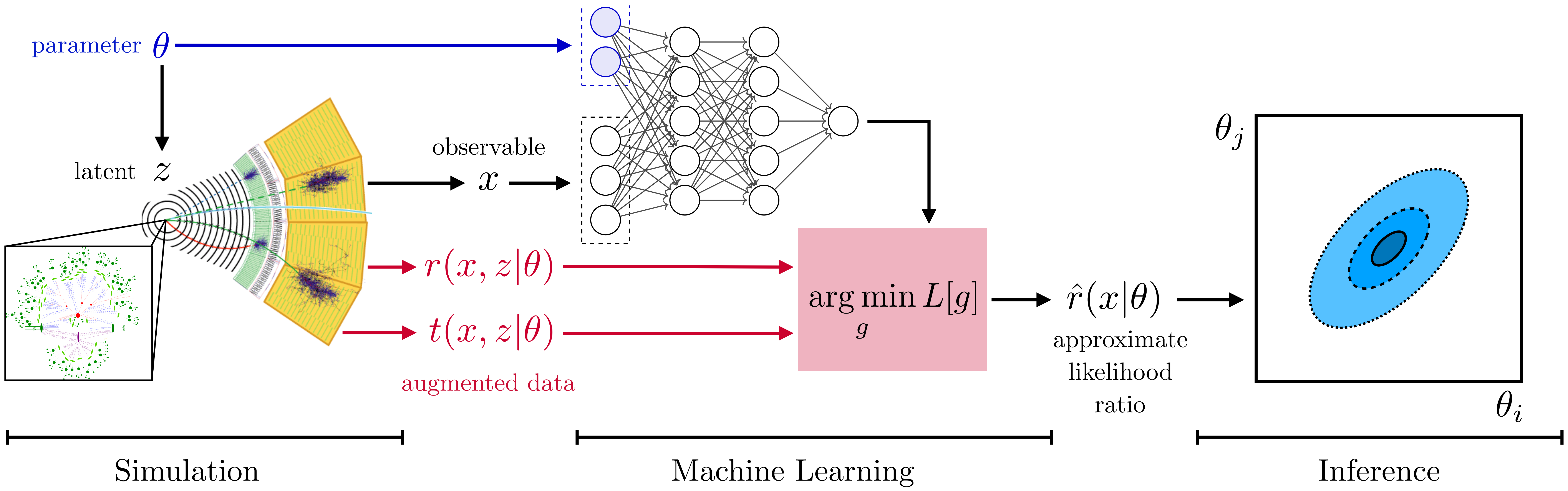
Supernova (Thomas/Nugent); Exoplanets (Ford+11)



Chen+ 2016 ApJ 836

Likelihood-Free Inference (LFI) / Simulation-based Inference (SBI)

Turn inference into density estimation task using simulated data

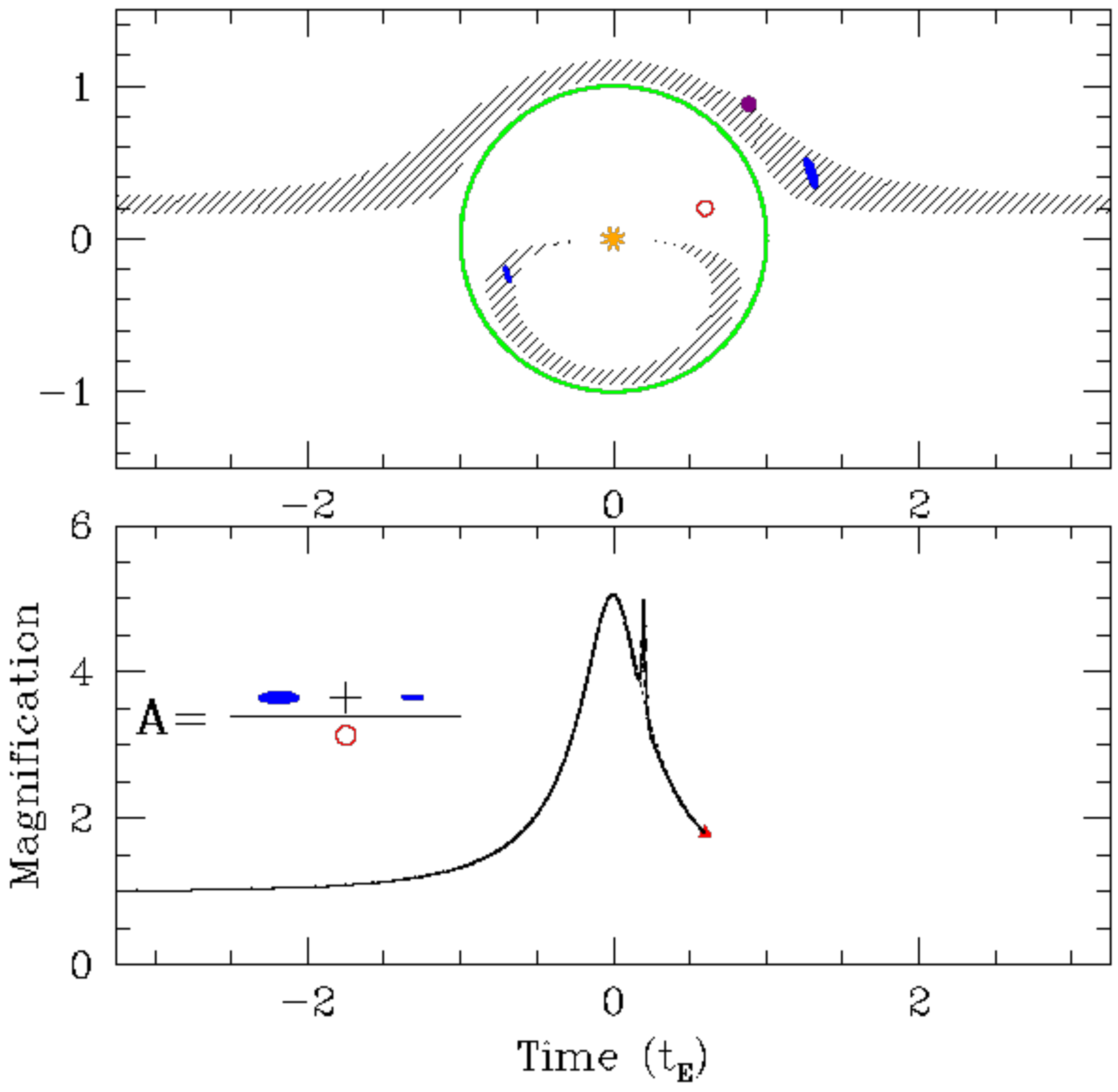


Brehmer+ 1805.00013
See also Cranmer, Brehmer & Louppe 2020

Astro example: "Fast likelihood-free cosmology with neural density estimators and active learning"

Alsing+ 1903.00007

Microlensing for Exoplanet Discovery & Characterization



Goal: measure masses, separations, orbits.

😅 Grid search+MCMC is slow (millions of forward model computations) & require experts in the loop

Animation: B. S. Gaudi

Microlensing for Exoplanet Discovery & Characterization

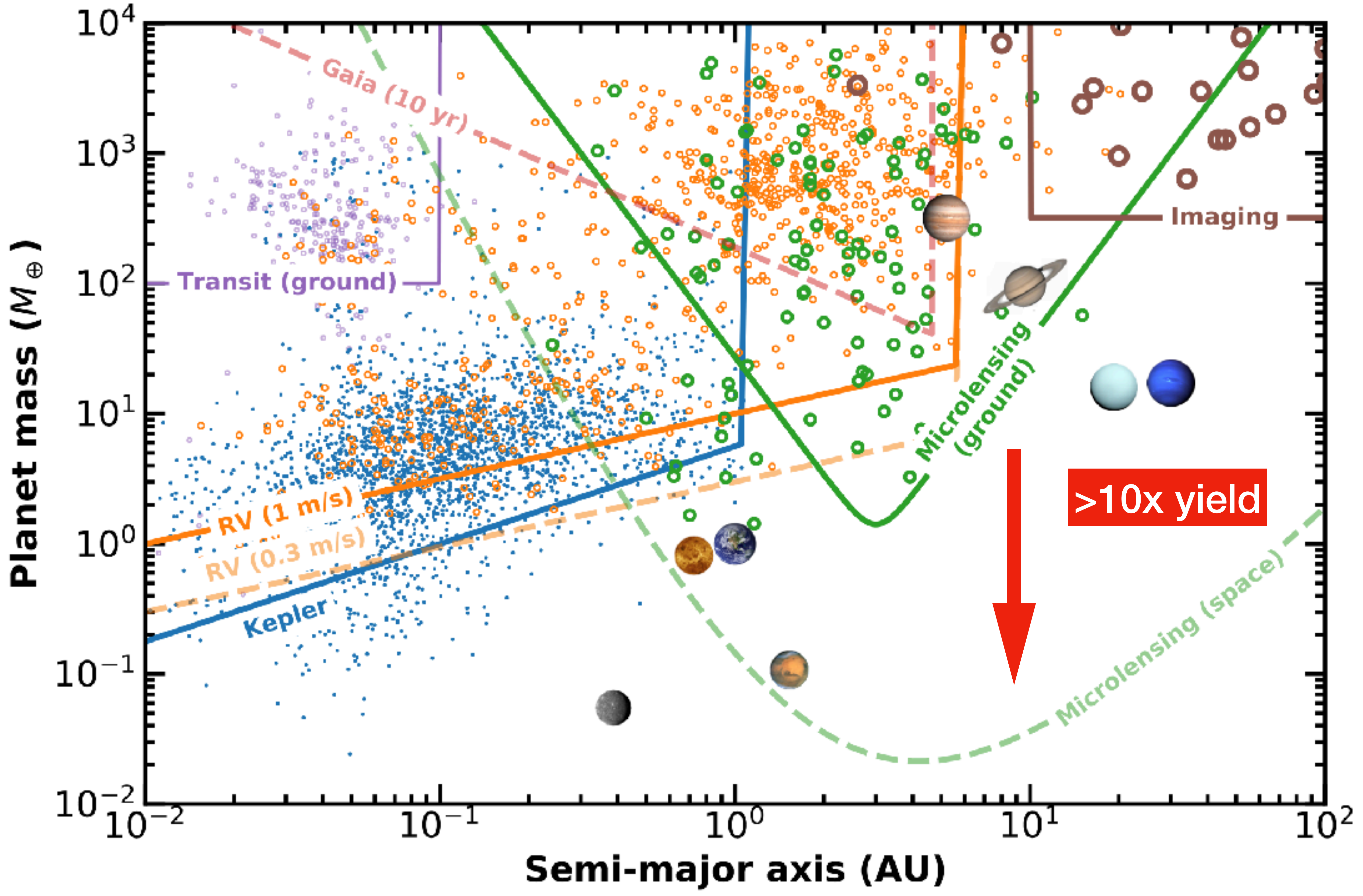


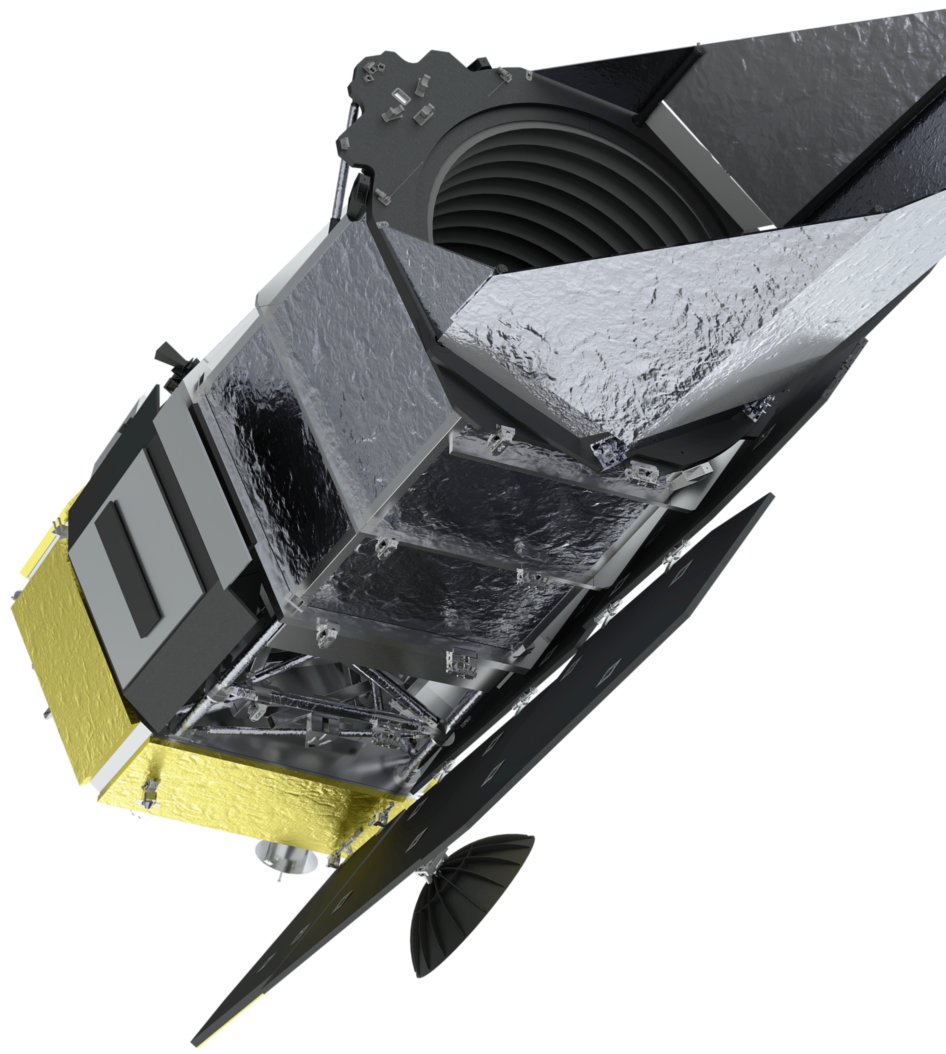
Figure from Zhu & Dong 2021

Goal: measure masses, separations, orbits.

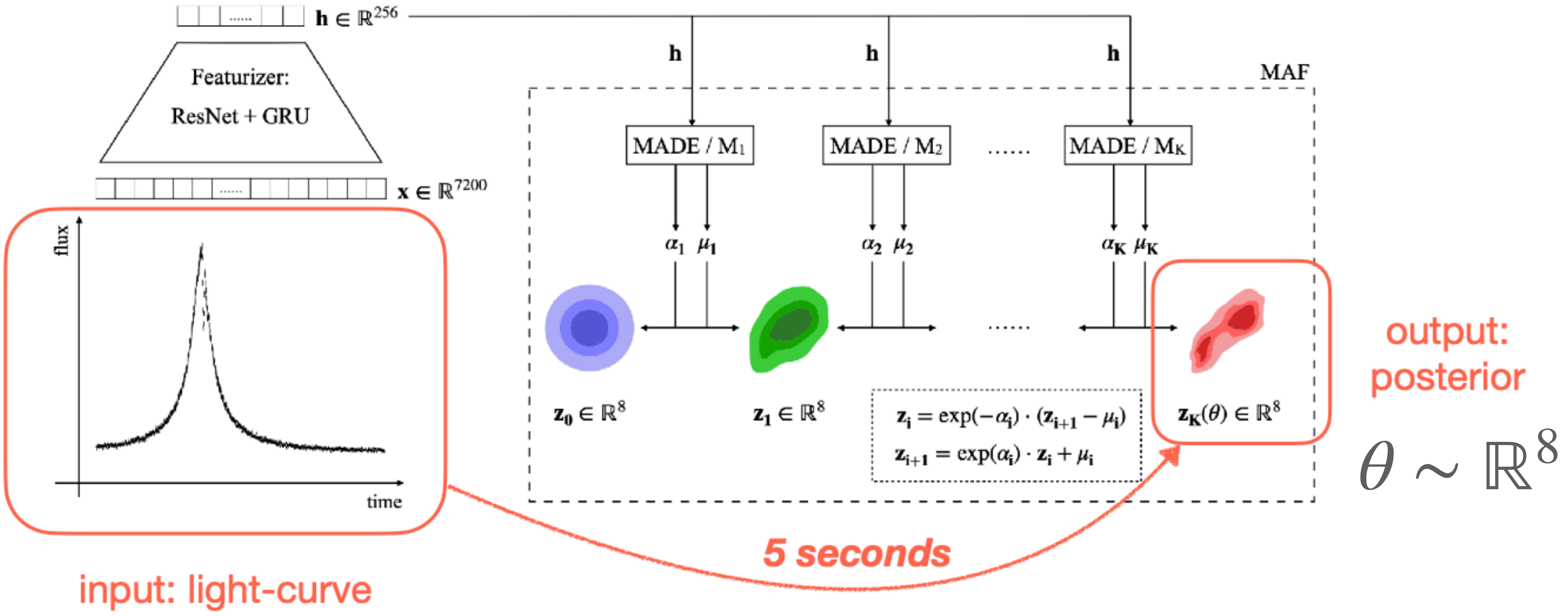
😅 Grid search+MCMC is slow (millions of forward model computations) & require experts in the loop

😱 Expecting *thousands* of events with Roman.

Calls for automated & more efficient inference approaches



Fast Inference with Neural Density Estimator

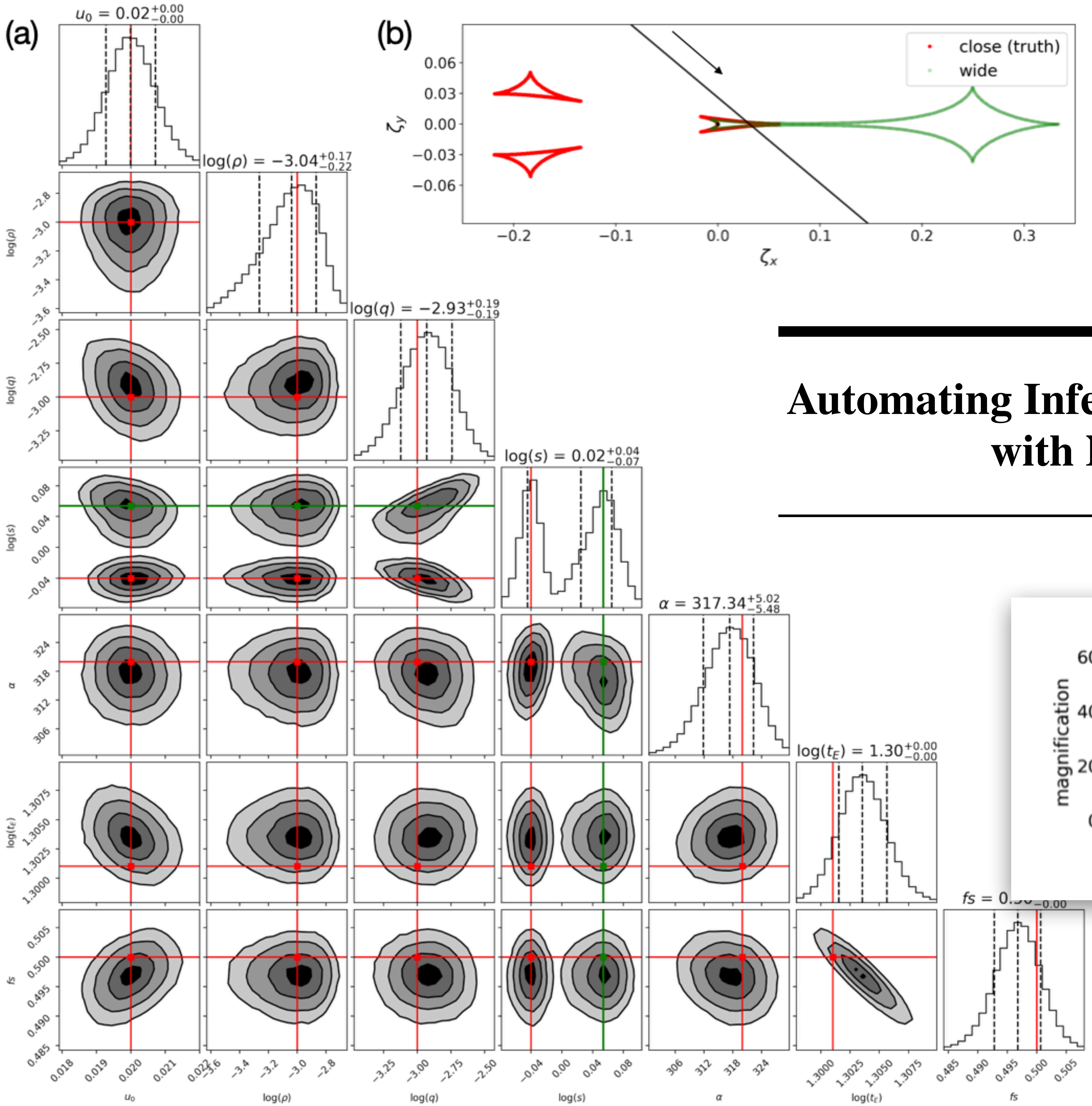


→Amortized inference, 10^5 faster

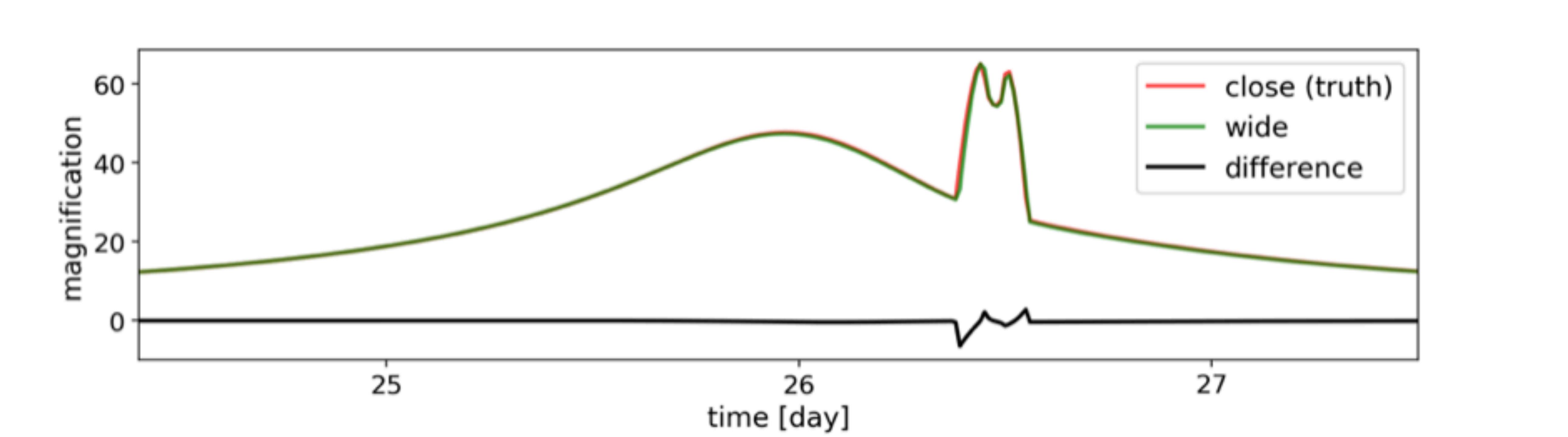
Zhang et al., *AJ* 161 262 (2021)

Zhang, JSB, … NeurIPS MP4PS (2010.04156)

Recovery of Known Caustic Degeneracies



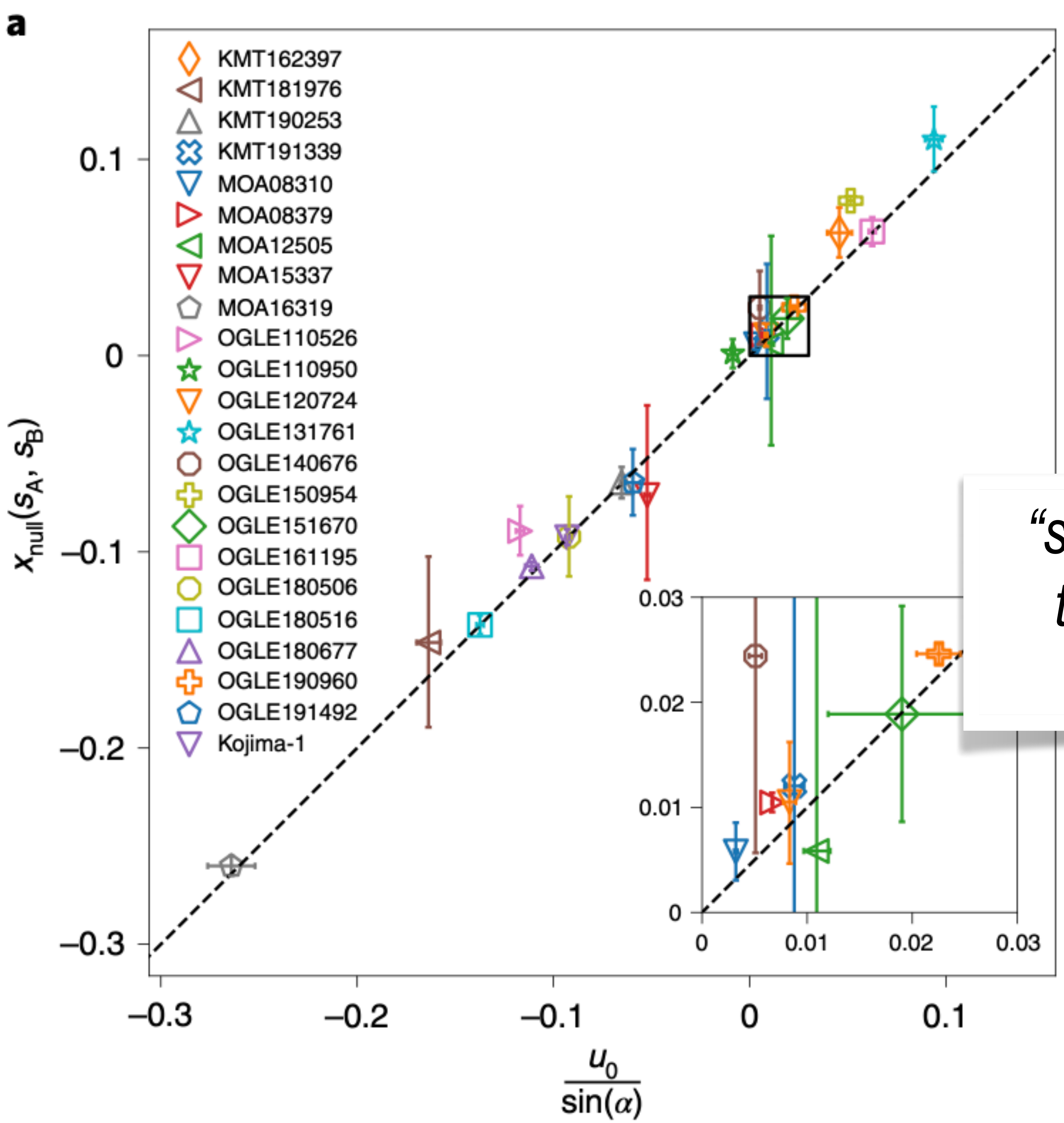
Automating Inference of Binary Microlensing Events with Neural Density Estimation



Zhang et al., *AJ* 161 262 (2021)

Zhang, JSB, … NeurIPS MP4PS (2010.04156)

Discovery of Magnification Degeneracies



Continuous set of "offset" degenerate light curves with inner-outer/close-wide as limiting cases

"suggests the existence of a deeper symmetry in the equations governing two-body lenses than previously recognized. "

Reanalysis of 23 previous 2-mode solutions shows one source location predicts the other

$$s_A = \frac{1}{2}\left(2x_0 - (s_B - 1/s_B) + \sqrt{[2x_0 - (s_B - 1/s_B)]^2 + 4}\right)$$

Zhang, Gaudi Bloom, Nat. Ast. 2022

Advancing astronomy by guiding human intuition with AI…

Letter | Published: 23 May 2022

A ubiquitous unifying degeneracy in two-body microlensing systems

Keming Zhang, B. Scott Gaudi & Joshua S. Bloom

Nature Astronomy **6**, 78

THE ASTROPHYSICAL JOURNAL LETTERS, 936:L22 (8pp), 2022 September 10 https://doi.org/10.3847/2041-8213/ac8c2b

OPEN ACCESS

CrossMark

A Mathematical Treatment of the Offset Microlensing Degeneracy

Keming Zhang (张可名)[1] and B. Scott Gaudi[2]
[1] Department of Astronomy, University of California, Berkeley, CA 94720-3411, USA; kemingz@berkeley.edu
[2] Department of Astronomy, The Ohio State University, Columbus, OH 43210, USA

Received 2022 May 11; revised 2022 August 19; accepted 2022 August 22; published 2022 September 9

…while AI is unlikely to replace scientists in the foreseeable future, [this work] demonstrates that it can be harnessed to help us understand deeper mathematical patterns in the underlying theory.

Mroz, Nat. Ast. News and Views (2022)

See also Davies et al. Nature, 2021
https://joshbloom.org/post/just_the_beginning/

Lesson 1: Don’t do ML unless you have to

Overcome **Resource** Constraints

Computation

- Accelerate physics-based simulation
- Simulation-based inference

Hardware

- Data transport bottlenecks
- Survey & instrument design
- Optimize observing plans

People

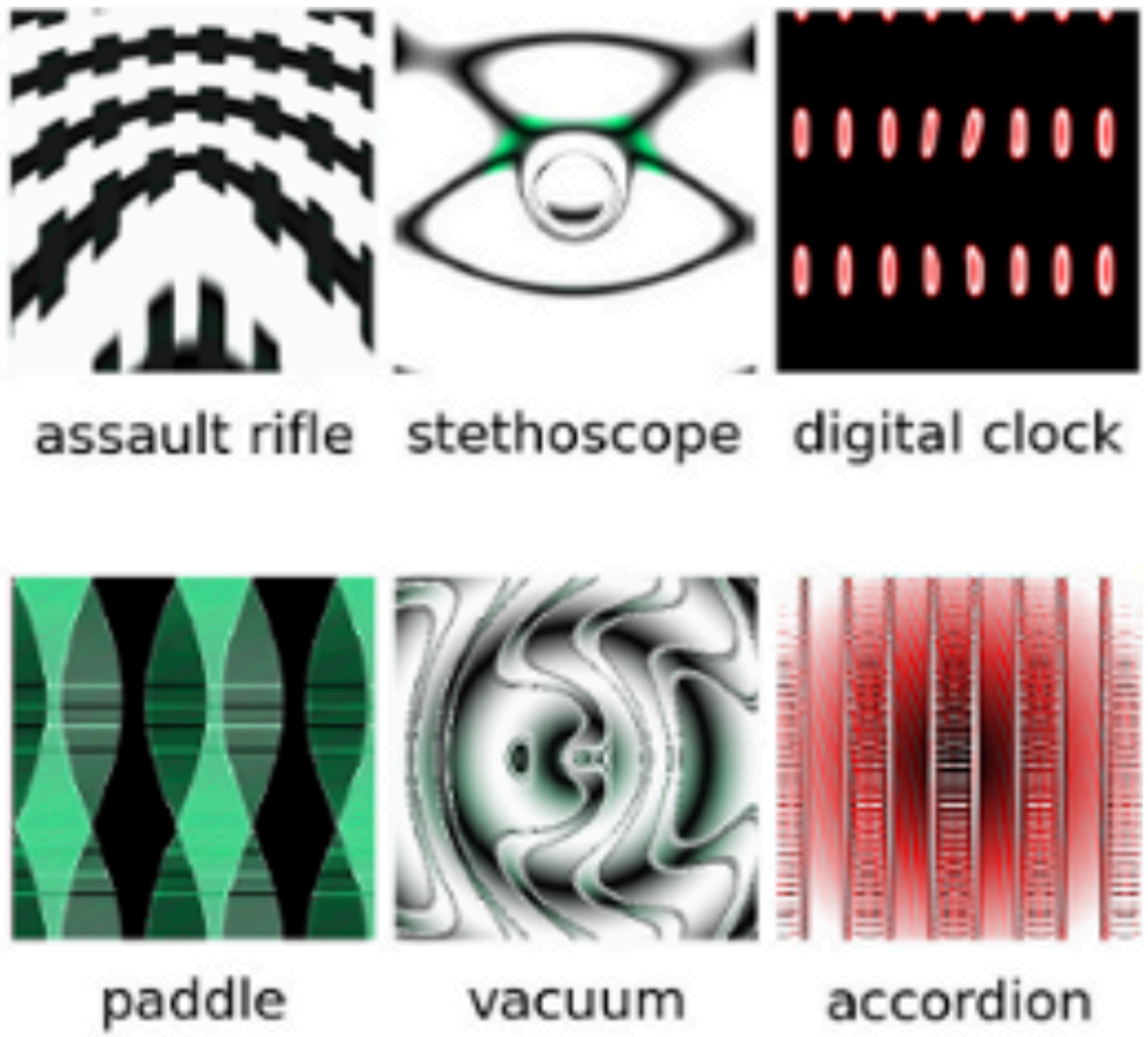
- Scaling decision support
- Automated Hypothesis generation
- Guided exploration & discovery

All Models Have Flaws

"It's common to forget the flaws of the model that you are most familiar…while the flaws of new models get exaggerated."

- John Langford (2007, Microsoft research)

http://hunch.net/?p=224

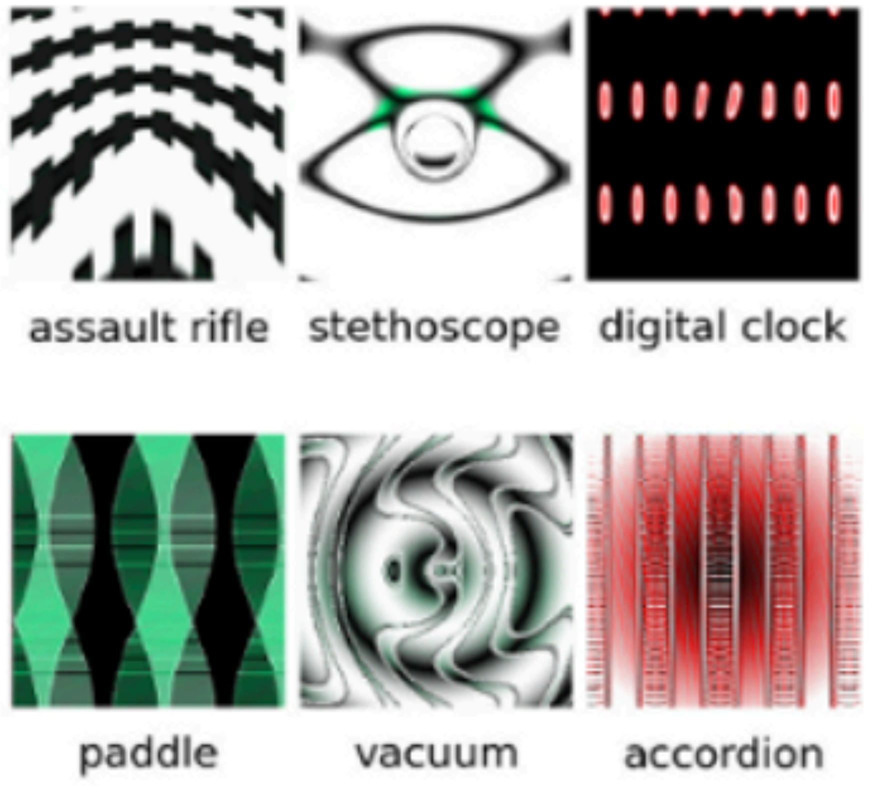


All Models Have Flaws

"It's common to forget the flaws of the model that you are most familiar…while the flaws of new models get exaggerated."

- *John Langford (2007, Microsoft research)*

http://hunch.net/?p=224



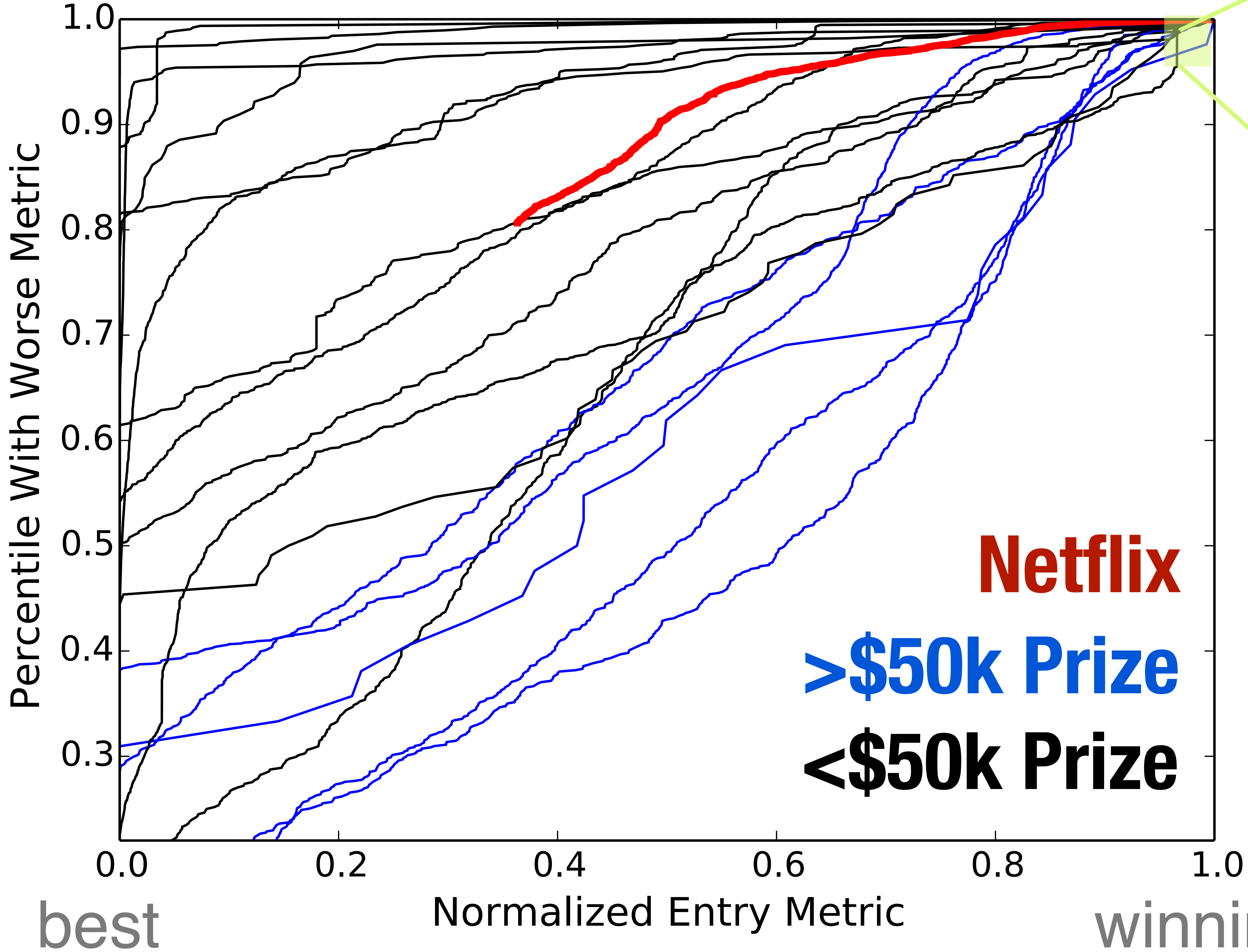
Nguyen et al, CVPR 2015

Magritte, ICML, 1929

task	Test on: / Train on:	SUN09	LabelMe	PASCAL	ImageNet	Caltech101	MSRC
"car" classification	SUN09	**28.2**	29.5	16.3	14.6	16.9	21.9
	LabelMe	14.7	**34.0**	16.7	22.9	43.6	24.5
	PASCAL	10.1	25.5	**35.2**	43.9	44.2	39.4
	ImageNet	11.4	29.6	36.0	**57.4**	52.3	42.7
	Caltech101	7.5	31.1	19.5	33.1	**96.9**	42.1
	MSRC	9.3	27.0	24.9	32.6	40.3	**68.4**
	Mean others	10.6	28.5	22.7	29.4	39.4	34.1

Torralba/Efros11 via L. Bottou (ICML 2015)

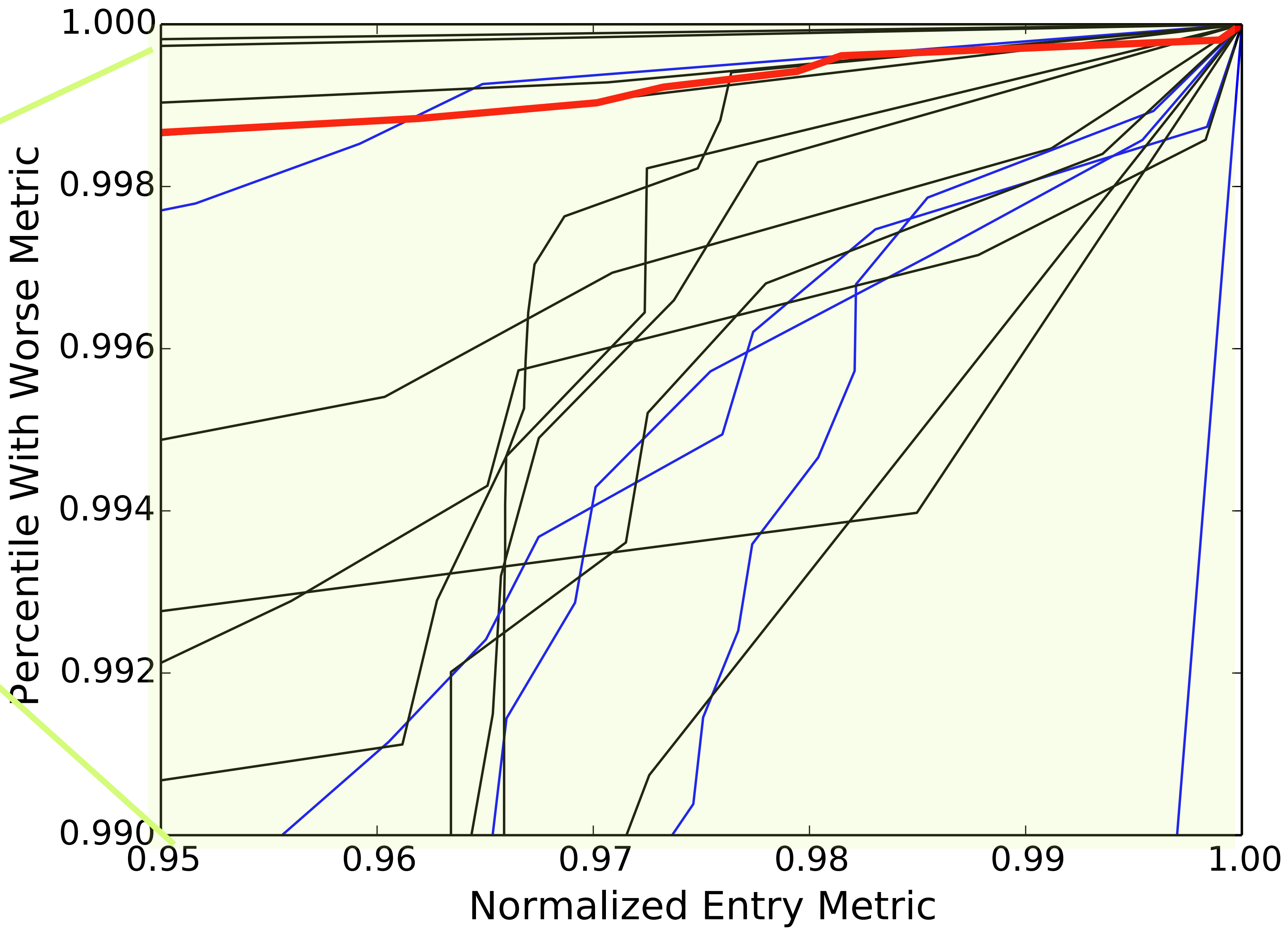
Optimization Metric



best benchmark

winning metric

Leaderboard data from Kaggle & Netflix



many teams get within ~few % of optimum

so which is easier to put into production?

On the NETFLIX Prize

"We evaluated some of the new methods offline but the **additional accuracy gains** that we measured **did not seem to justify the engineering effort** needed to bring them into a **production environment.**"

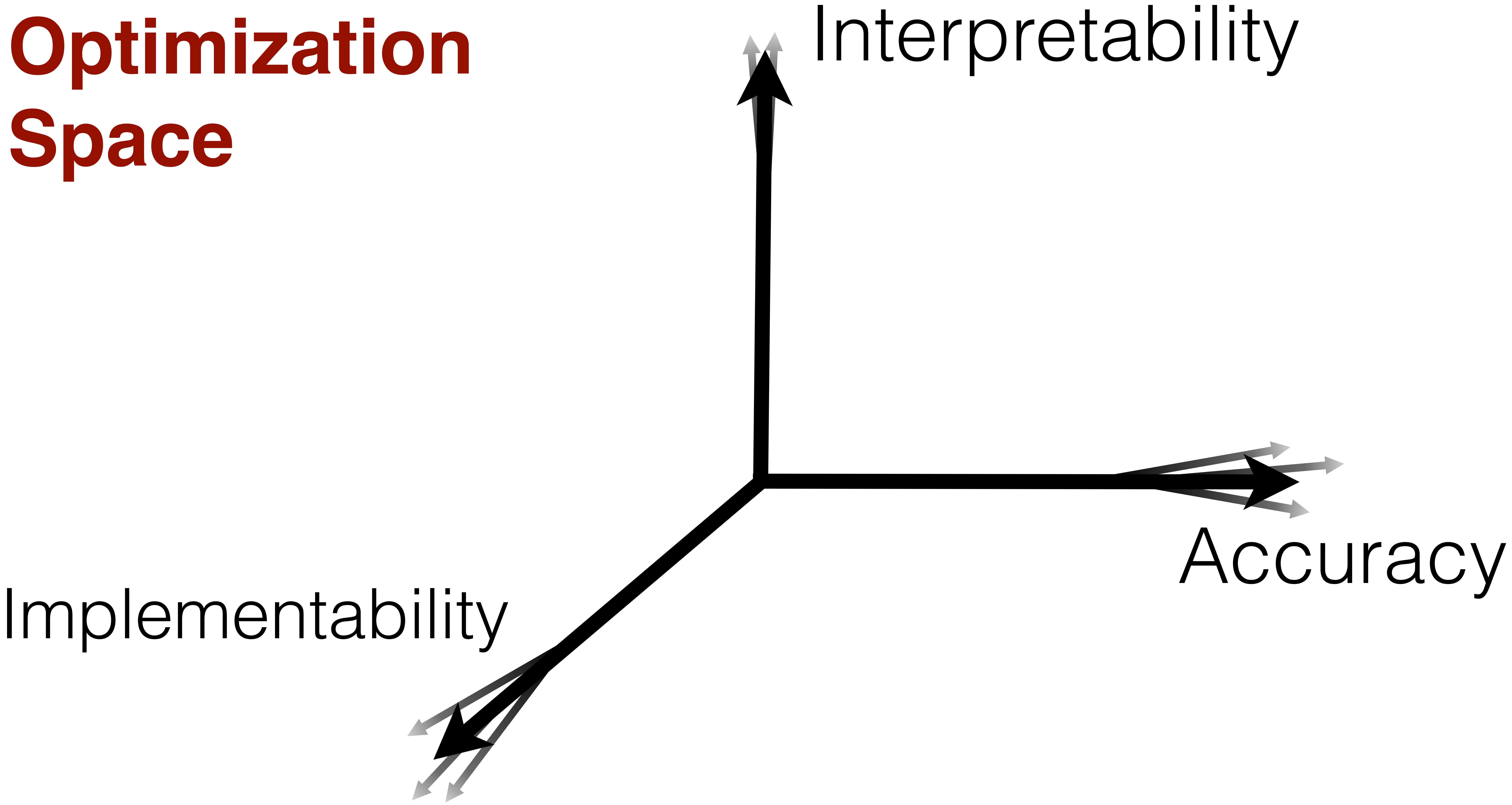
Xavier Amatriain and Justin Basilico (April 2012)

Lesson 2: Choose the right tool for the problem

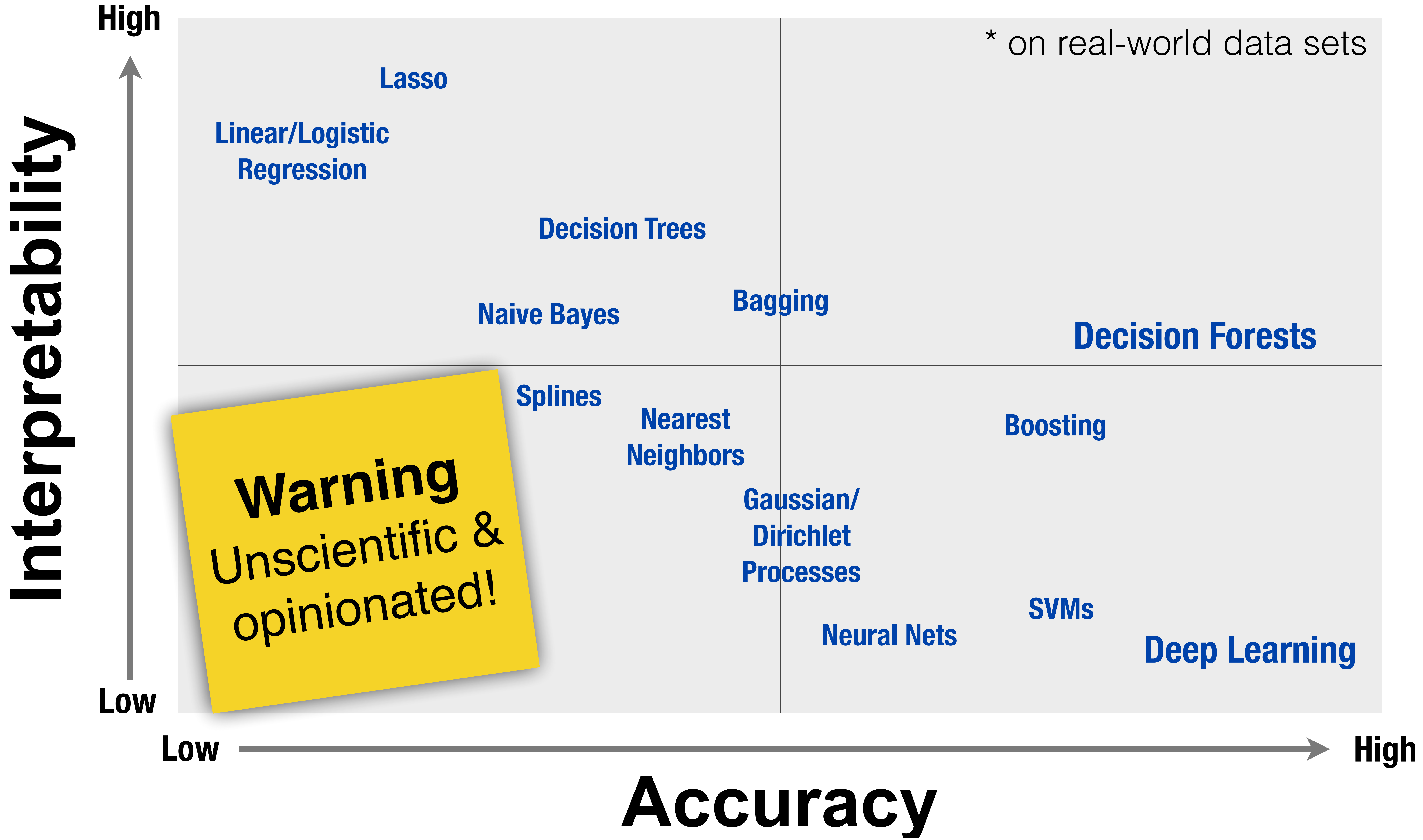
The **simplest** is usually the best

Results from simple approaches are, at worst, good **benchmarks** for you to beat with more complex solutions

Optimization Space



Machine Learning Algorithmic Trade-Off



All ML in production is a Systems Challenge

What are we optimizing for?

Component	What
Algorithm/Model	Learning rate, convexity, error bounds, scaling, …
+ Software/Hardware	Accuracy, Memory usage, Disk usage, CPU needs, time to learn, time to predict
+ Project Staff	time to implement, people/resource costs, reliability, maintainability, experimentability
+ Consumers	direct value, useability, explainability, actionability, security, privacy
+ Society	indirect value, ethics

- multi-axis optimizations in a given component
- highly coupled optimization considerations between components
- myopic view can be costly further up the stack

Lesson 3: Writing papers is easy, but ML in Production is *Hard*

Only real test of the model is if its falsifiable on data that does not yet exist

Since *all* models are fallible & people are always on the receiving end, we need to invest in how model are hot-swapped, predictions are consumed & acted upon

The mail you want, not the spam you don't

Posted: Thursday, July 09, 2015

Posted by Sri Harsha Somanchi, Product Manager

The Gmail team is always working hard to make sure that every message you care about arrives in your inbox, and all the spam you don't want remains out of sight. In fact, less than 0.1% of email in the average Gmail inbox is spam, and the amount of wanted mail landing in the spam folder is even lower, at under 0.05%.

Official Gmail Blog

News, tips and tricks from Google's Gmail team and friends.

Linus Torvalds

Shared publicly - Jul 17, 2015

Dear Google Mail Team,

I've said very nice things about your spam filter in the past, but I'm afraid I am going to have to take it all back. I'm currently going through the spam for the last week, and have gone through about a third of it.

Something you did recently has been an unmitigated disaster. Of the roughly 1000 spam threads I've gone through so far, right now 228 threads were **incorrectly** marked as spam.

That's not the 0.1% false positive rate you tried to make such a big deal about last week. No. That's over 20% of my spambox being real emails with patches and pull requests. Almost a quarter!

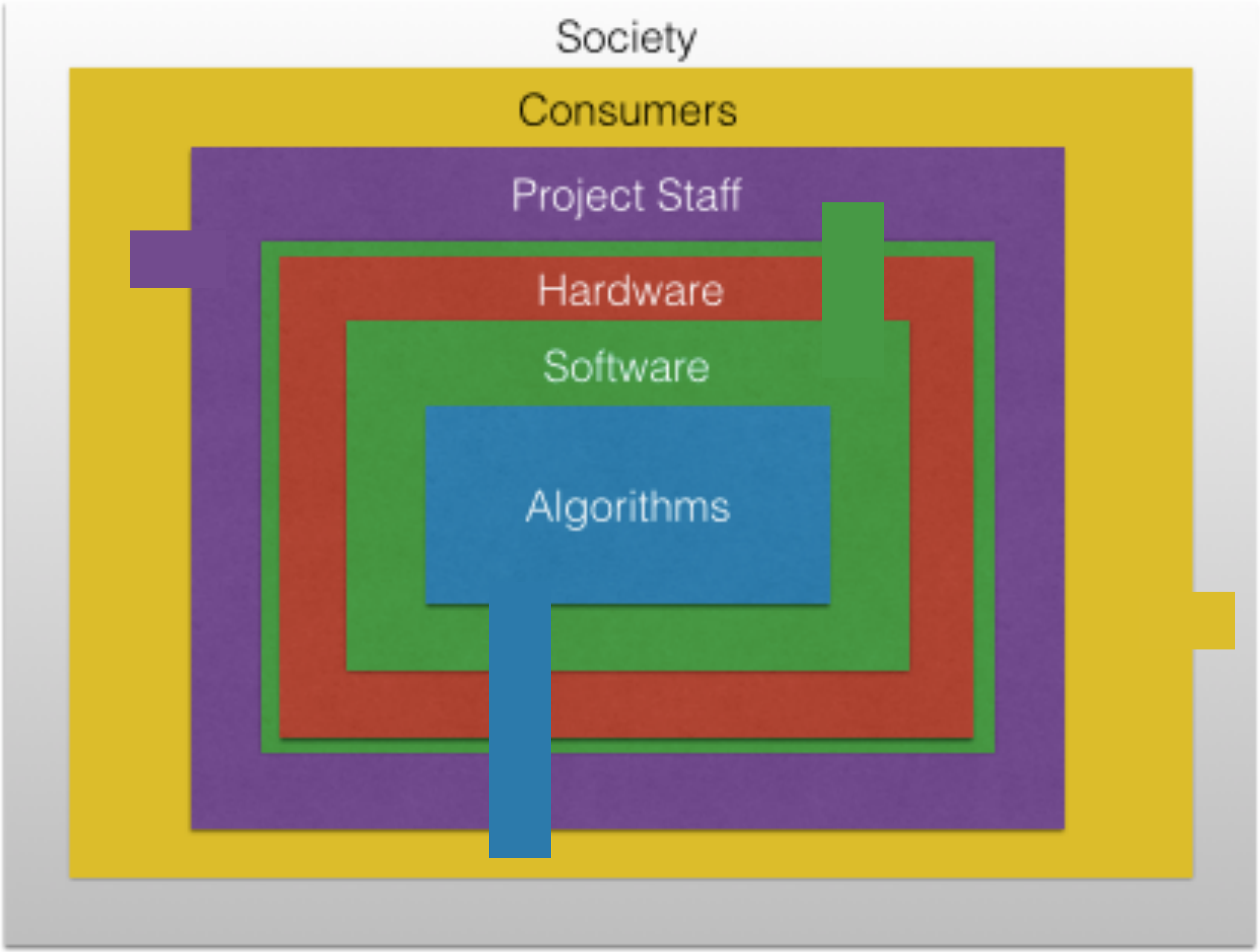
I don't know how to even describe the level of brokenness in those kinds of spam

"Yes Netflix, because my 6 year old will enjoy the animated fun of Sons of Anarchy"

https://www.reddit.com/r/funny/comments/3e7gy4/yes_netflix_because_my_6_year_old_will_enjoy_the/

“Weak Contracts” ie. Abstractions within components bleed through to other components

cf. Sculley …



Example (via Bottou)

1. A smart programmer makes an inventive use of a trained object recognizer.
2. The object recognizer receives data that does not resemble the testing data and outputs nonsense.
3. The code of the smart programmer does not work.

> *"It may be surprising to the academic community to know that only a fraction of the code … is actually doing 'machine learning'. A mature system might end up being (at most)* ***5% machine learning code*** *and (at least)* ***95% glue code****."*

- Complex models erode abstraction boundaries
- Data dependencies cost more than code dependencies: weak contracts
- System-level Spaghetti
- Changing External World

see also, Bottou (Facebook) ICML

http://research.google.com/pubs/pub43146.html

Machine Learning: The High-Interest Credit Card of Technical Debt

D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov,
Todd Phillips, Dietmar Ebner, Vinay Chaudhary, Michael Young
{dsculley,gholt,dgg,edavydov}@google.com
{toddphillips,ebner,vchaudhary,mwyoung}@google.com
Google, Inc

Abstract

Machine learning offers a fantastically powerful toolkit for building complex systems quickly. This paper argues that it is dangerous to think of these quick wins as coming for free. Using the framework of *technical debt*, we note that it is remarkably easy to incur massive ongoing maintenance costs at the system level when applying machine learning. The goal of this paper is highlight several machine learning specific risk factors and design patterns to be avoided or refactored where possible. These include boundary erosion, entanglement, hidden feedback loops, undeclared consumers, data dependencies, changes in the external world, and a variety of system-level anti-patterns.

1 Machine Learning and Complex Systems

Real world software engineers are often faced with the challenge of moving quickly to ship new products or services, which can lead to a dilemma between speed of execution and quality of en-

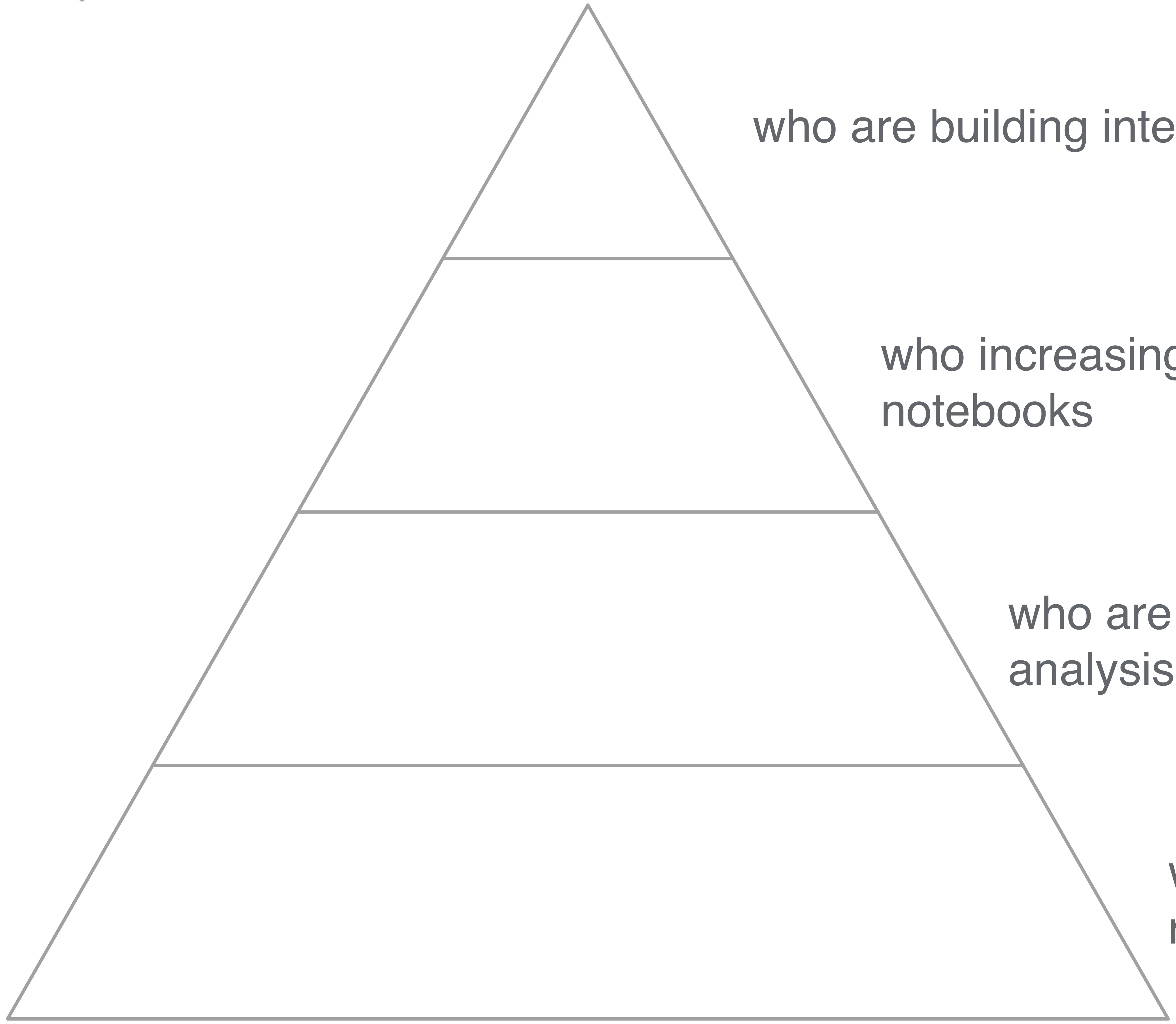
Lession 4:
Get the people and roles right

O'REILLY

What are machine learning engineers?

A new role focused on creating data products and making data science work in production.

By Ben LoricaMike Loukides June 6, 2017



who are building intelligent products from data

who increasingly develop prototypes using notebooks

who are responsible for statistical analysis and modeling

who carry out ad-hoc analysis and reporting

Tools, talent, and org structure should align with this reality

Doing ML is a Team Sport

Π	vs.	Γ
deep domain skill/knowledge/training deep methodological knowledge/skill		deep domain or methodological skill/knowledge/training strong methodological or domain knowledge/skill

Summary

- **Wide range of ML approaches**
  Decision Forests are the go-to for tabular data, neural approaches for most other types of data.
  BUT always try simple approaches first -> benchmark

- Clear exemples of scientific acceleration with ML, but **do not do ML unless you HAVE to**…

- **ML in production is HARD**. Easy to convince yourself of efficacy of ML solutions with off-the-shelf data…only real testing data is that that has not been created yet

- Work in domain and techically diverse teams. It’s more fun.

Thanks and enjoy the school!

Joshua Bloom
Astronomy

ML4FP School Lecture
Aug 12, 2024
Lawrence Berkeley National Laboratory