3rd ASTERICS-OBELICS International School

8-12 April 2019, Annecy, France.

H2020-Astronomy ESFRI and Research Infrastructure Cluster
(Grant Agreement number: 653477).

Machine Learning Tutorial
IV - Beyond textbook ML

3rd ASTERICS-OBELICS International School
8-12 April 2019, Annecy - France

Emille E. O. Ishida
Laboratoire de Physique de Clermont - Université Clermont-Auvergne
Clermont Ferrand, France

Summary

I. Quick recap of the week
with a few interesting additions

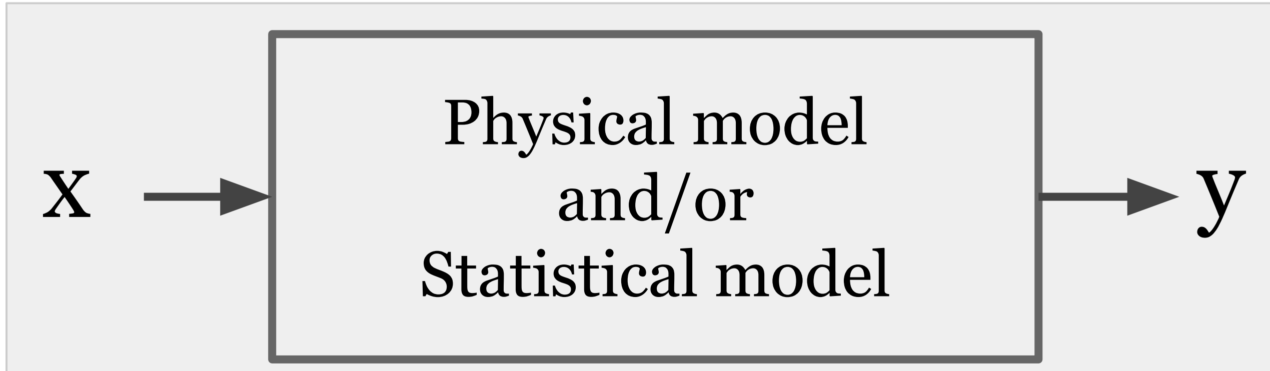
II. Representativeness matters

III. Adaptive Learning Techniques

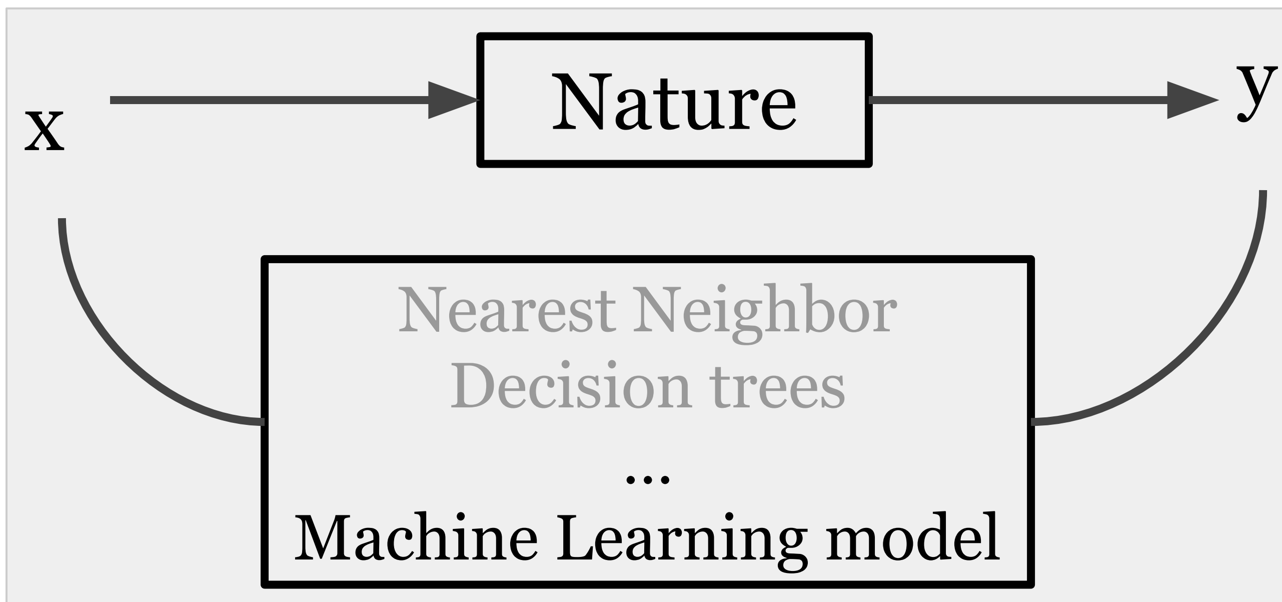
IV. The human factor

Hypothesis:

Traditiona
l
data
modeling:



Algorithmic
modeling:



Breiman, L., *Statistical Modeling: The Two Cultures*, Stat. Sci, Volume 16 (2001)

Supervised ML **model**

data **training**, target

χ set of all samples, x

Y set of possible labels, y

h_{train} learner: $y_{est;i} = h_{train}(x_i)$

L Loss function

Representativeness matters!

Data generation model:

$x_i \sim P_\chi$

$f \rightarrow$ true labeling function, $y_i = f(x_i)$

$L_{data,f}(h) \equiv P_{x\sim data}(h_{train}(x) \neq f(x))$

Shai and Shai, *Understanding ML: From Theory to Algorithms*, 2014, CUP

Supervised ML **model**

data — **training**, target

Machine Learning algorithm

X — set of all examples, x

Y — set of possible labels, y

$\boxed{h_{train} \quad \text{learner: } y_{est;i} = h_{train}(x_i)}$

L — Loss function

Representativeness matters!

Data generation model:

$x_i \sim P_X$

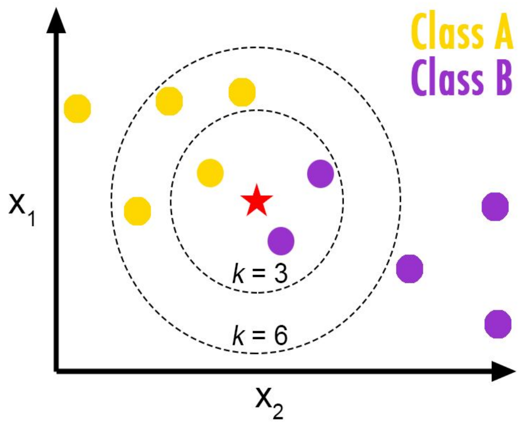
f — true labeling function, $y_i = f(x_i)$

$L_{data,f}(h) \equiv P_{x\sim data}(h_{train}(x) \neq f(x))$

Shai and Shai, *Understanding ML: From Theory to Algorithms*, 2014, CUP

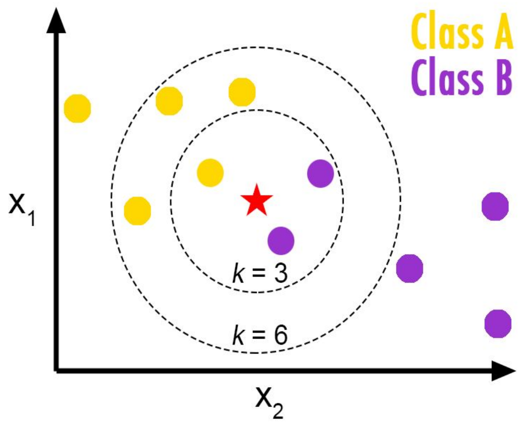
Machine Learning algorithms

K Nearest Neighbor

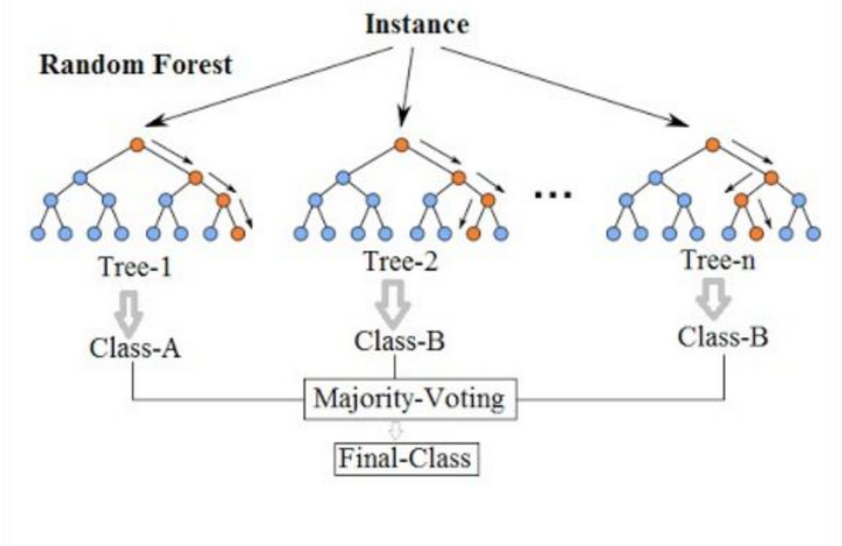


Machine Learning algorithms

K Nearest Neighbor

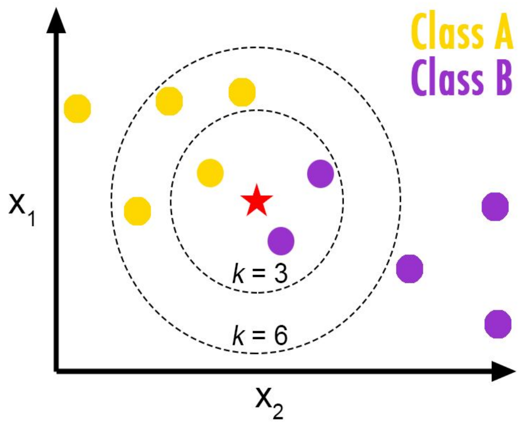


Decision trees and random forests

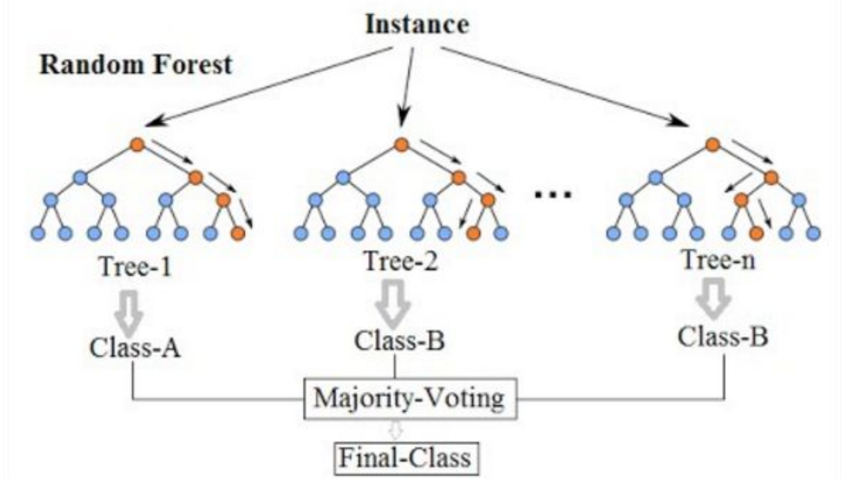


Machine Learning algorithms

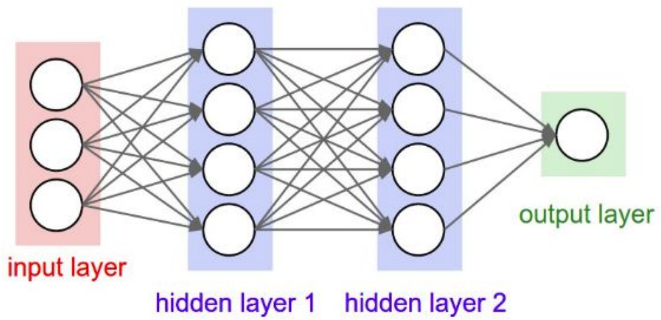
K Nearest Neighbor



Decision trees and random forests

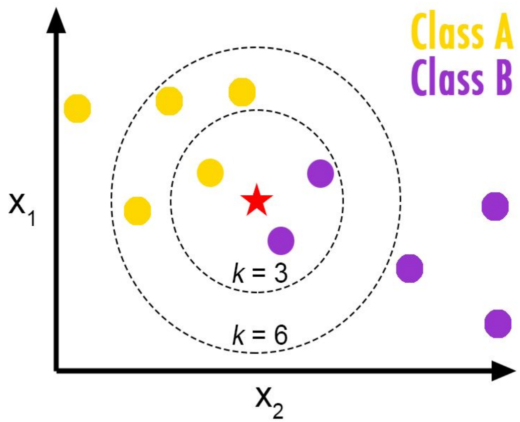


Neural networks and deep learning

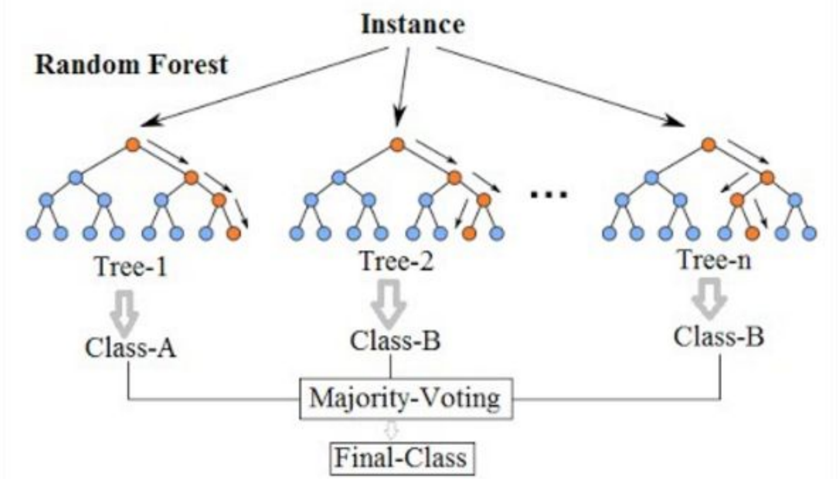


Machine Learning algorithms

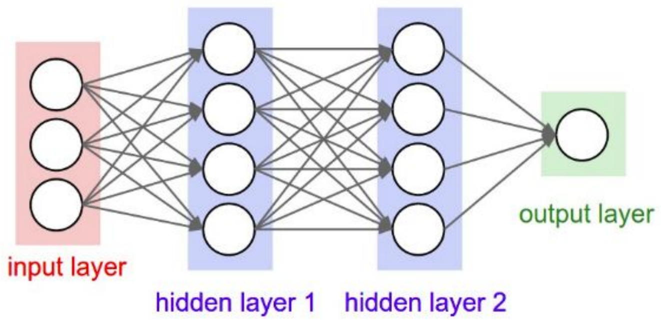
K Nearest Neighbor



Decision trees and random forests



Neural networks and deep learning



A comment on

Feature extraction

"Raw" data

Homogeneous matrix

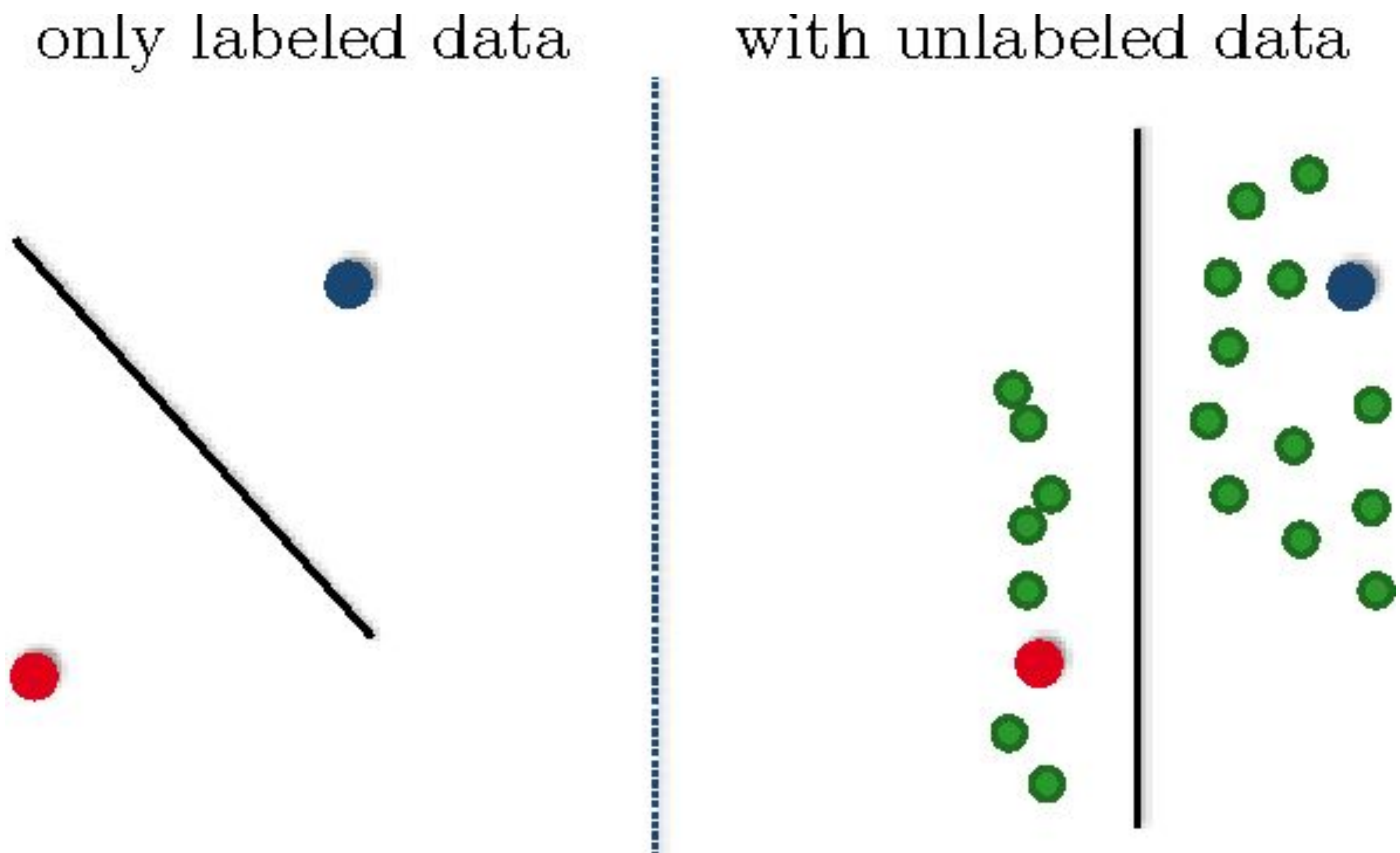
Dimensionality Reduction or Feature Extraction

Learning algorithm

Results

Semi-supervised learning

Getting partial information from the unlabelled sample

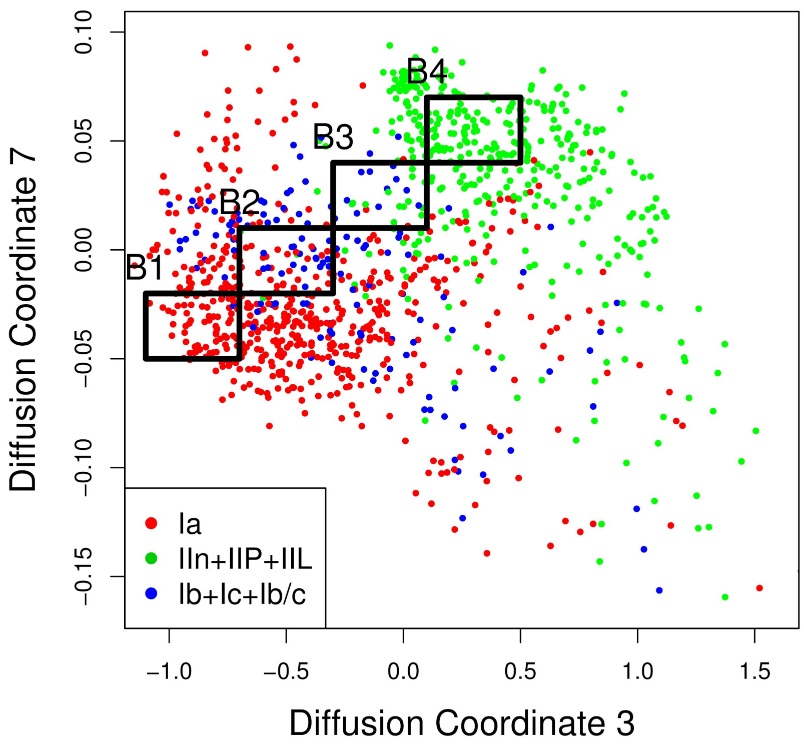


Valko et al., Semi-Supervised Apprenticeship Learning, 2012, JMLR,

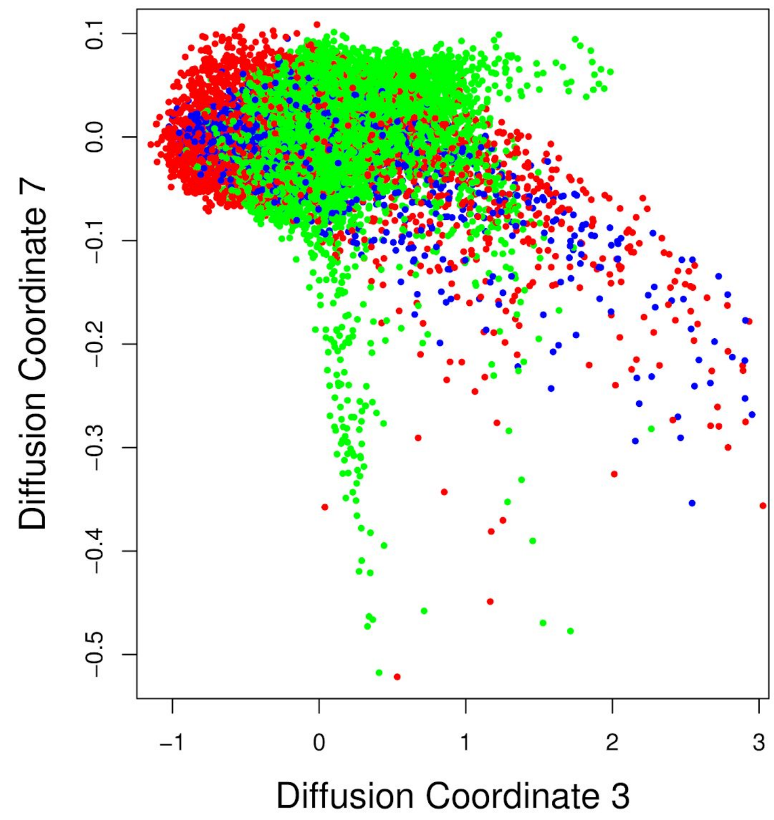
Semi-supervised learning

For Supernova Photometric Classification

Spectroscopic only (training)



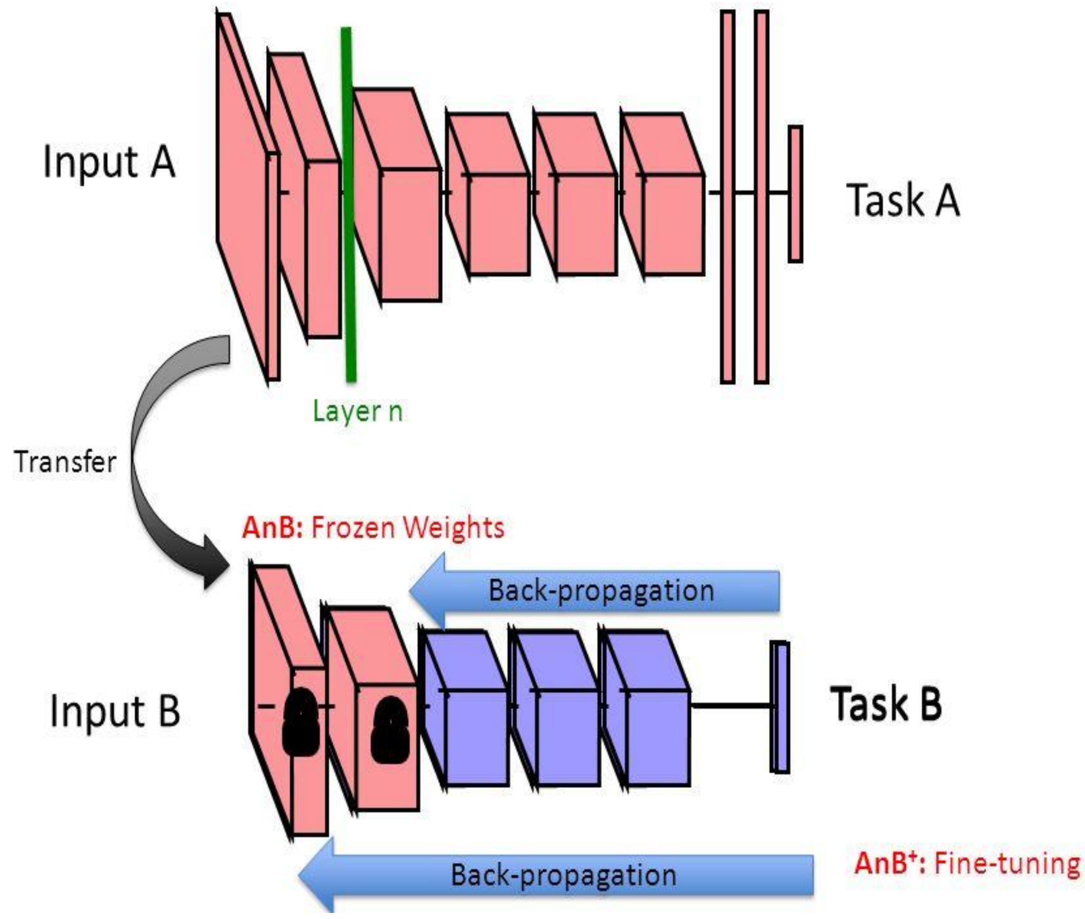
All available data



Richards et al., 2011, MNRAS

Transfer Learning

Borrowing information from somewhere else



http://slideplayer.com/slide/8370683/

Transfer Learning

Exploiting information from various data sets

Use only the low dimensional representation of things you are interested in!

Dimensionality Reduction
or
Feature Extraction

Homogeneous matrix

Learning algorithm

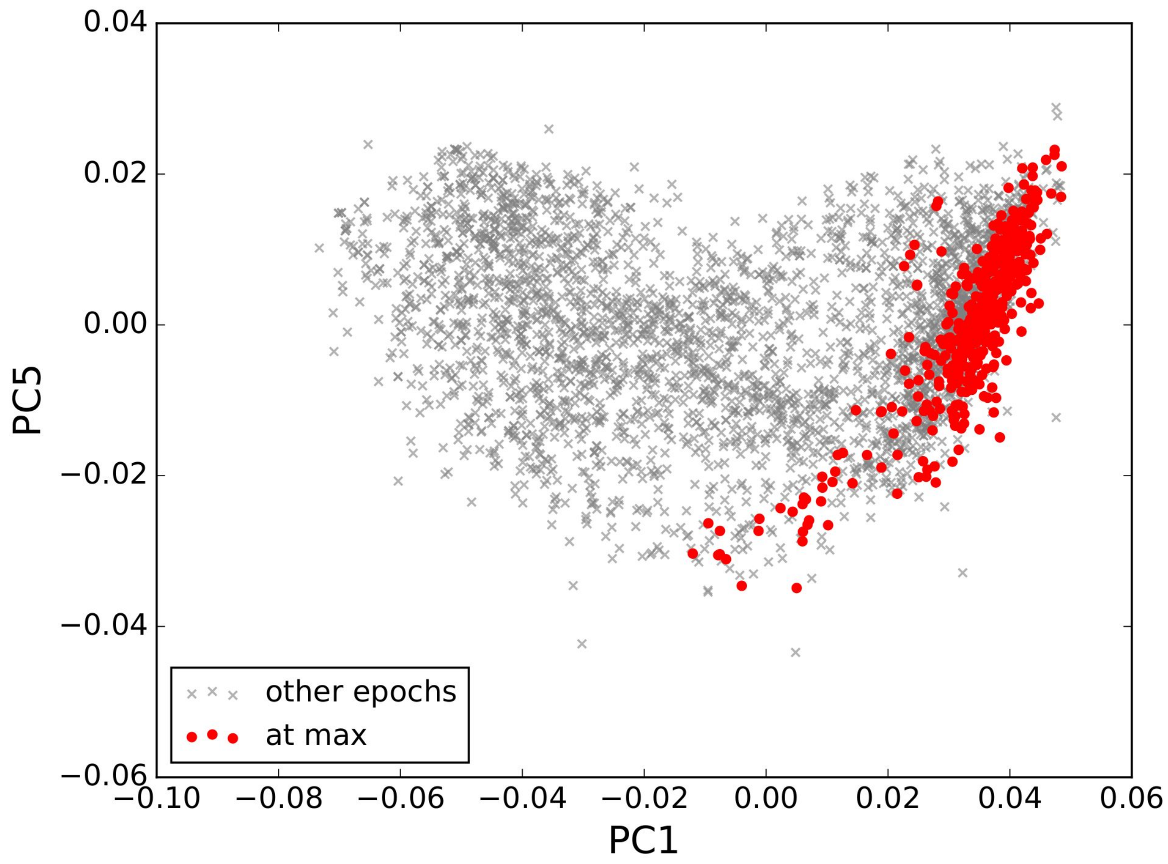
"Raw" data

Add observations from non-interesting sources

Results

Transfer Learning

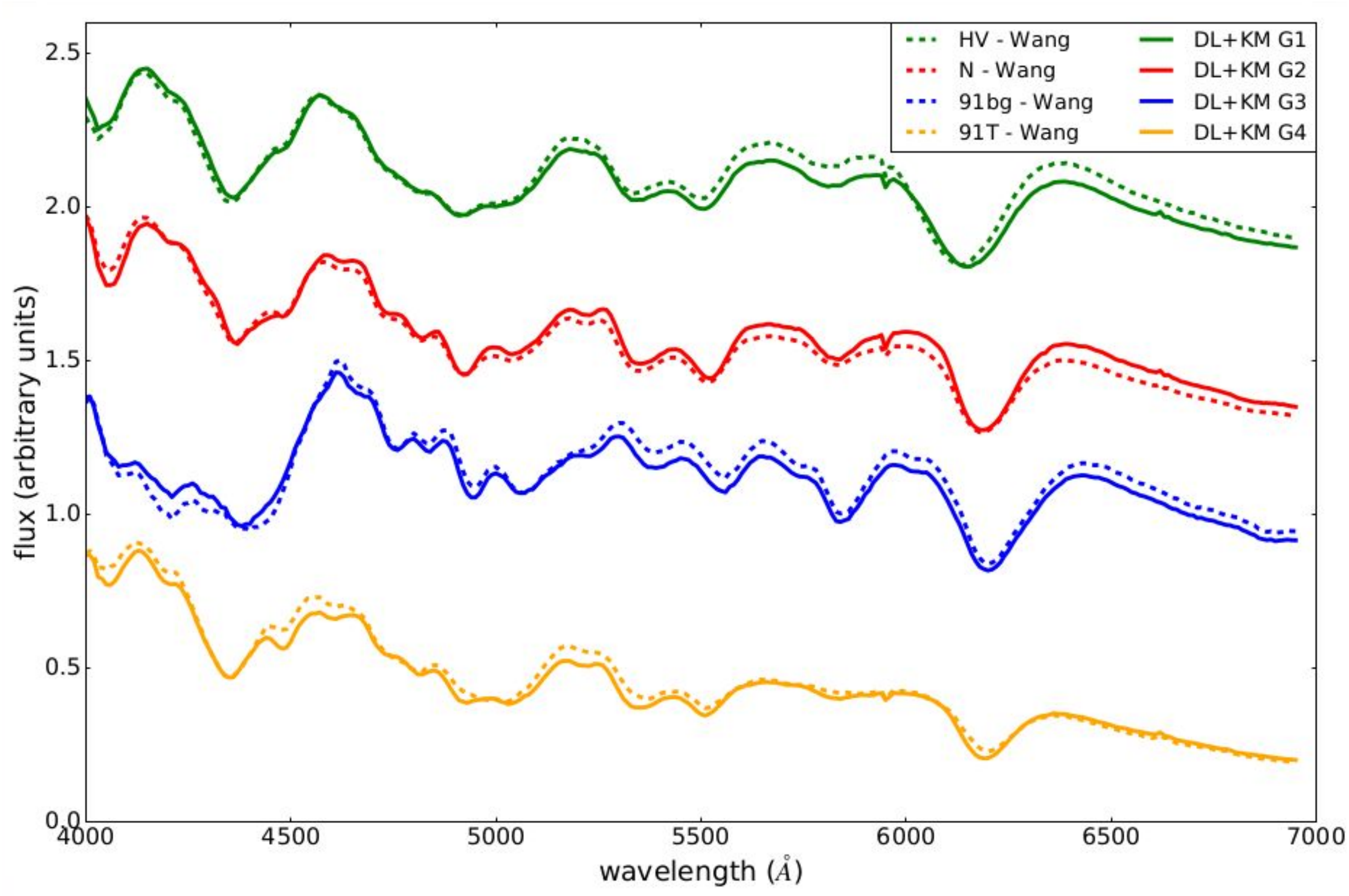
In Astronomy



*Sasdelli, **Ishida** et al., 2016, MNRAS, 461, Issue 2, p.2044, from **CRP #2***

Unsupervised Clustering

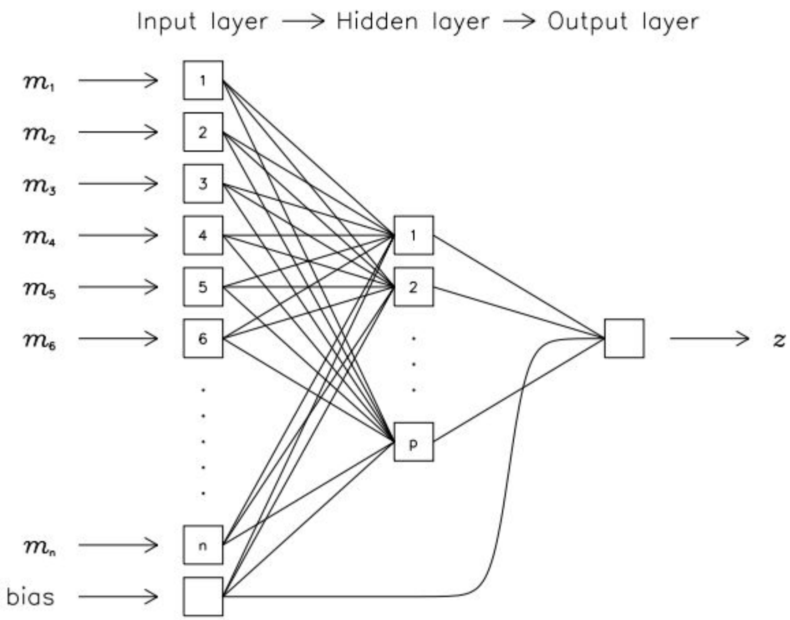
In Astronomy



*Sasdelli, **Ishida** et al., 2016, MNRAS, 461, Issue 2, p.2044, from **CRP #2***

Neural Network

In astronomy: photometric redshift estimation



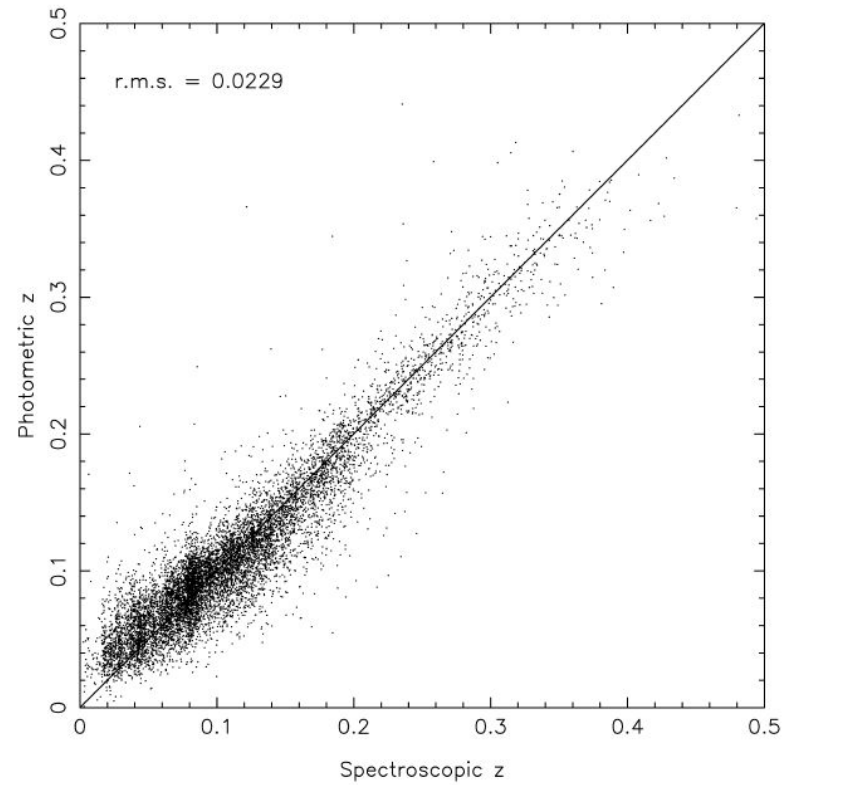
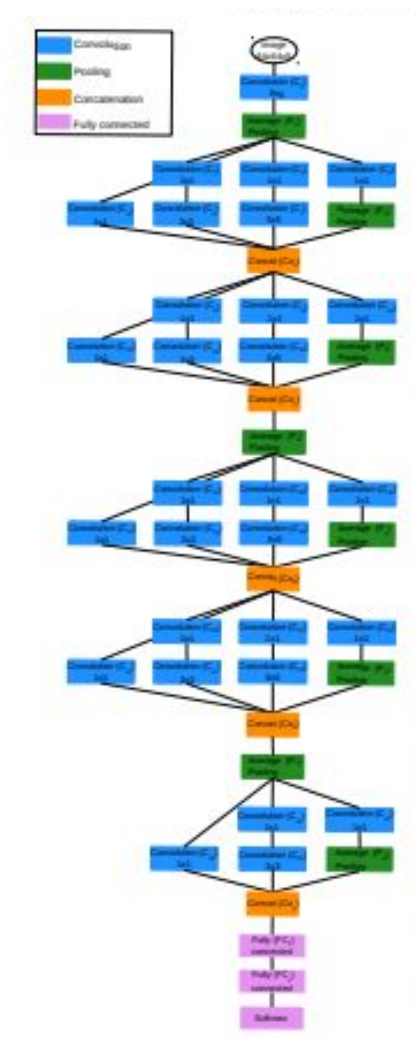


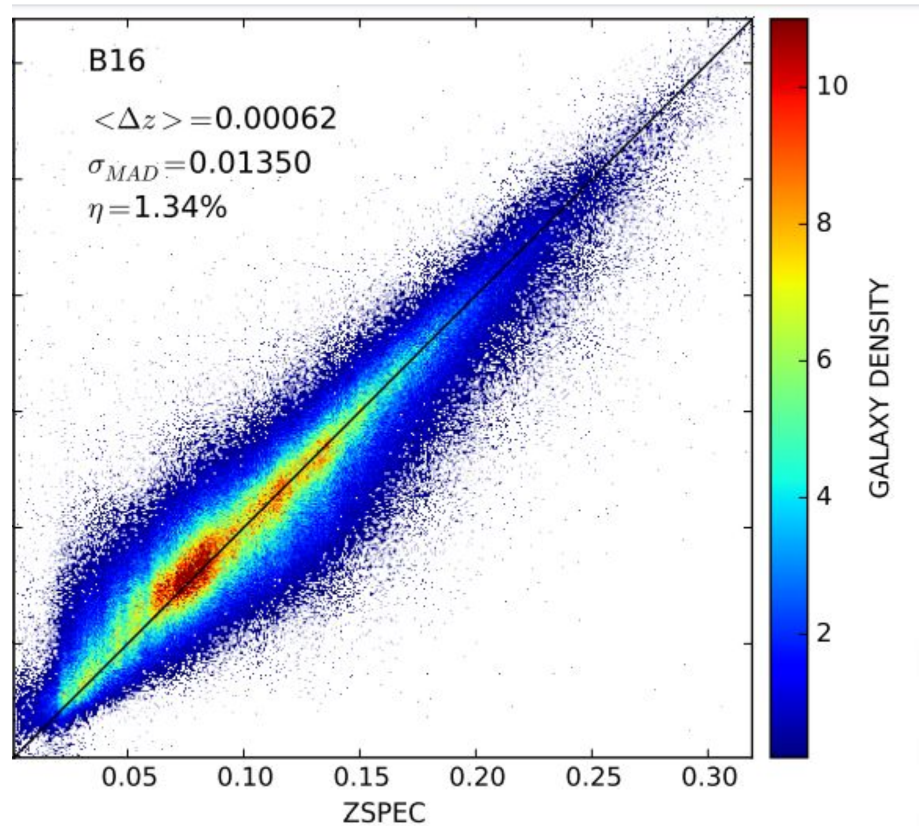
FIG. 2.— Spectroscopic *vs.* photometric redshifts for ANNz applied to 10,000 galaxies randomly selected from the SDSS EDR.

Collister & Lahav, 2003

Neural Network

In astronomy: photometric redshift estimation





Pasquet et al., 2019, A&A

Symbolic Regression

Supervised Learning: an extreme regression example

Mathematical atoms:

+, - , x , / , pow

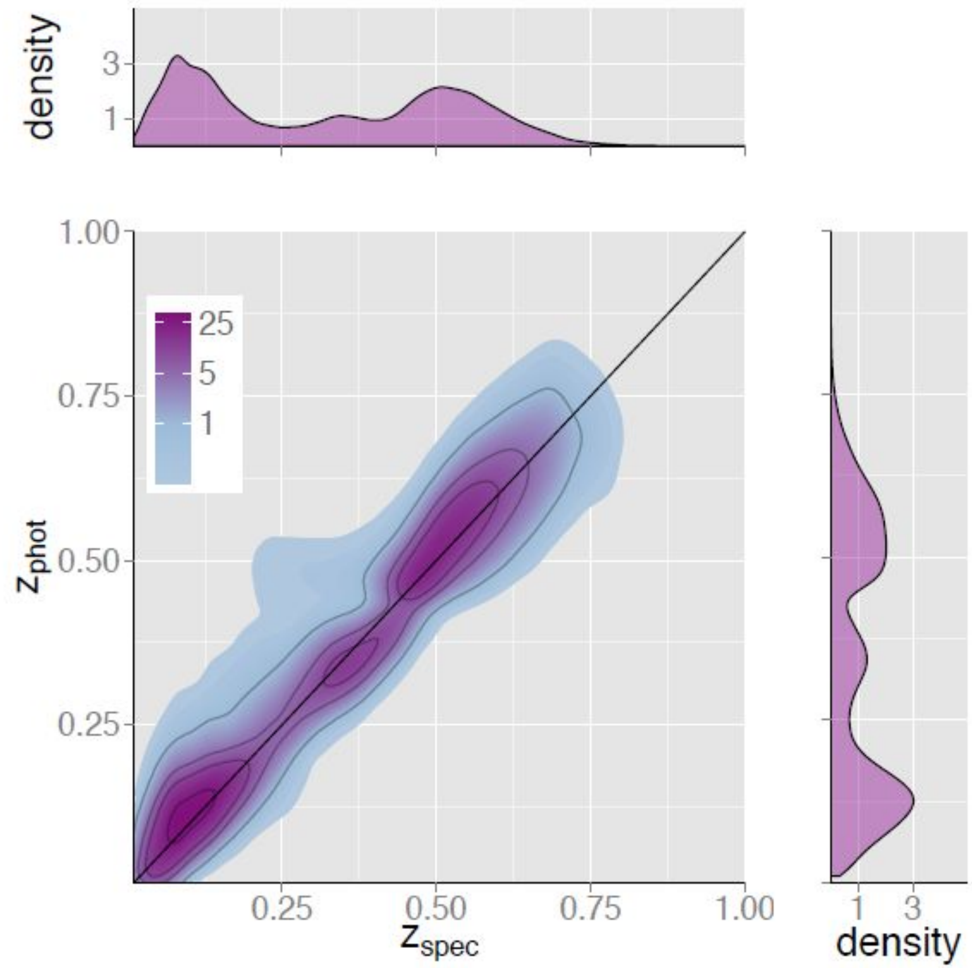
1 - Random construction of an analytical expression

2 - find the best parameters

3 - if result is better than previous keep it, otherwise discard it

Final expression:

$$z_{\text{phot}} = \frac{0.4436r - 8.261}{24.4 + (g-r)^2(g-i)^2(r-i)^2 - g} \\ +0.5152(r-i).$$



Pre-COIN paper:
*Krone-Martins, **Ishida** & de Souza, MNRASL 443 (2014)*

All of the above relies on representativeness…

This is a very strong assumption

All of the above relies on representativeness…

How often does this hypothesis hold in astronomy?

All of the above relies on representativeness…

How often does this hypothesis hold in astronomy?

What happens when it breaks?

Notebook: Regression2.ipynb

Labels are expensive

We need recommendation systems

Labels are expensive

How to construct optimal training samples?

Labels are expensive

How to construct optimal training samples?

Labels are expensive

How to construct optimal training samples?

Labels are expensive

How to construct optimal training samples?

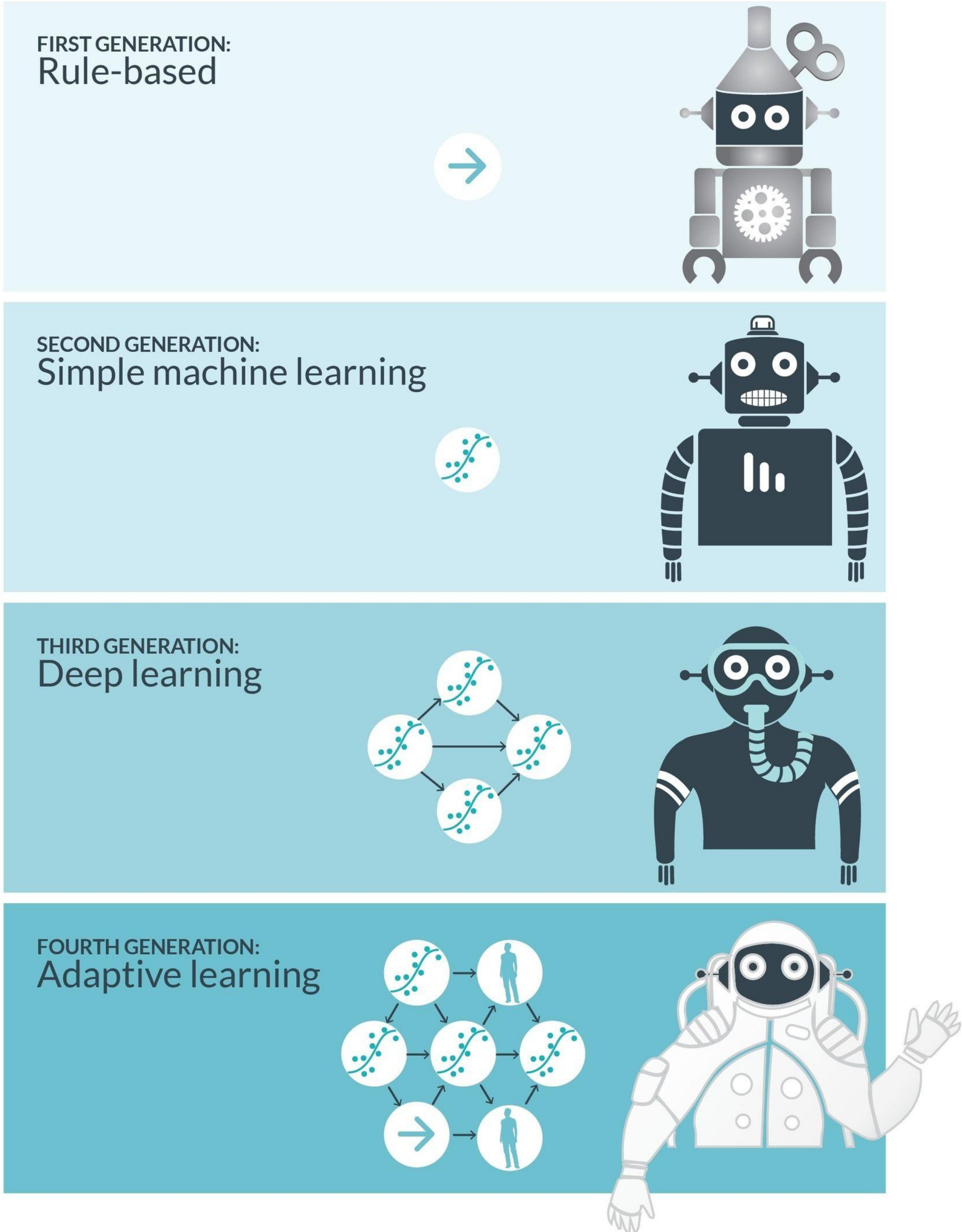
amazon

35% of Amazon's revenue are generated by it's recommendation engine.

NETFLIX

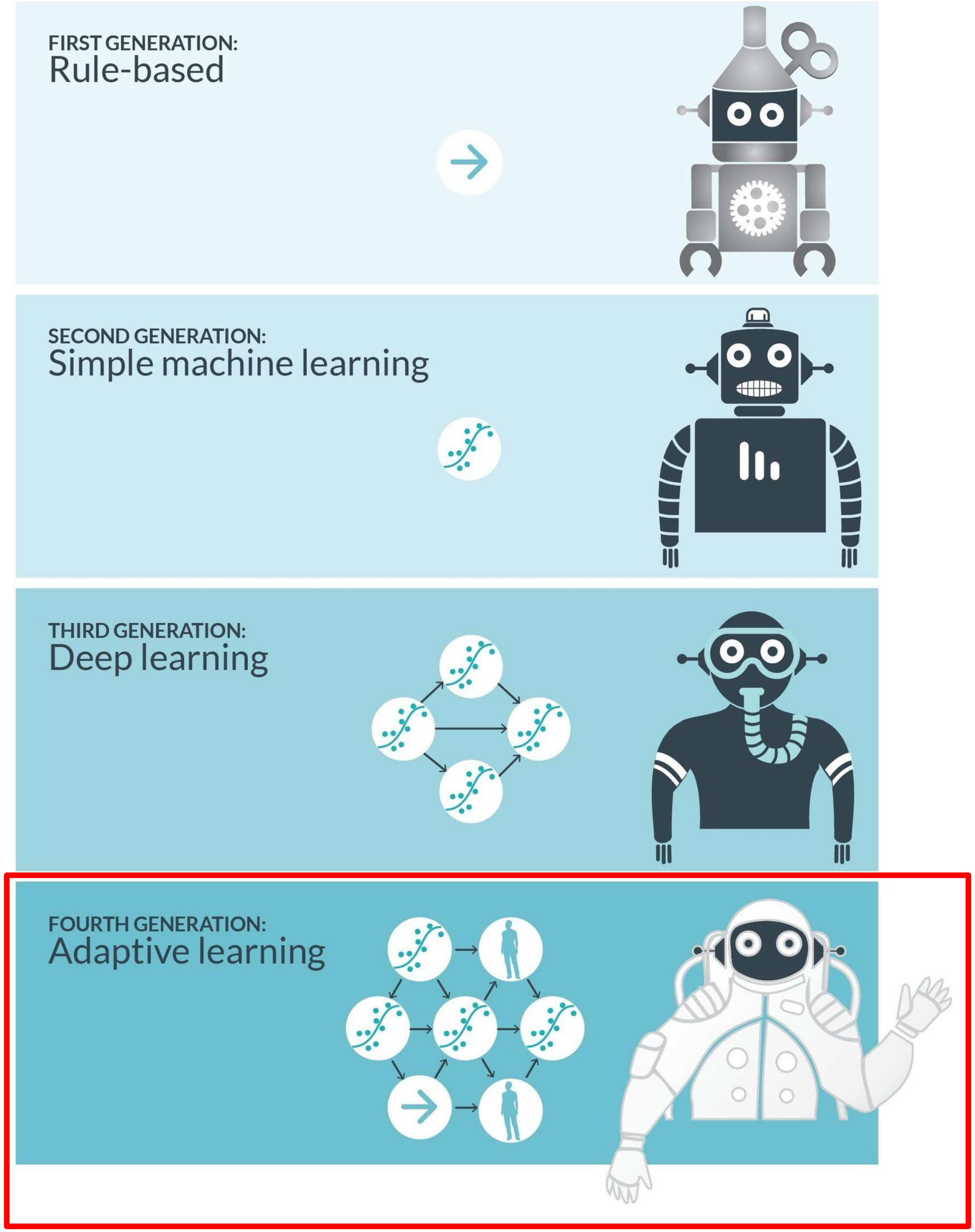
75% of users select movies based on Netflix's recommendations.

Can machines learn **better**, with **fewer** labelled examples, if they are carefully chosen?



Machines need to evolve…

so they need to adapt!

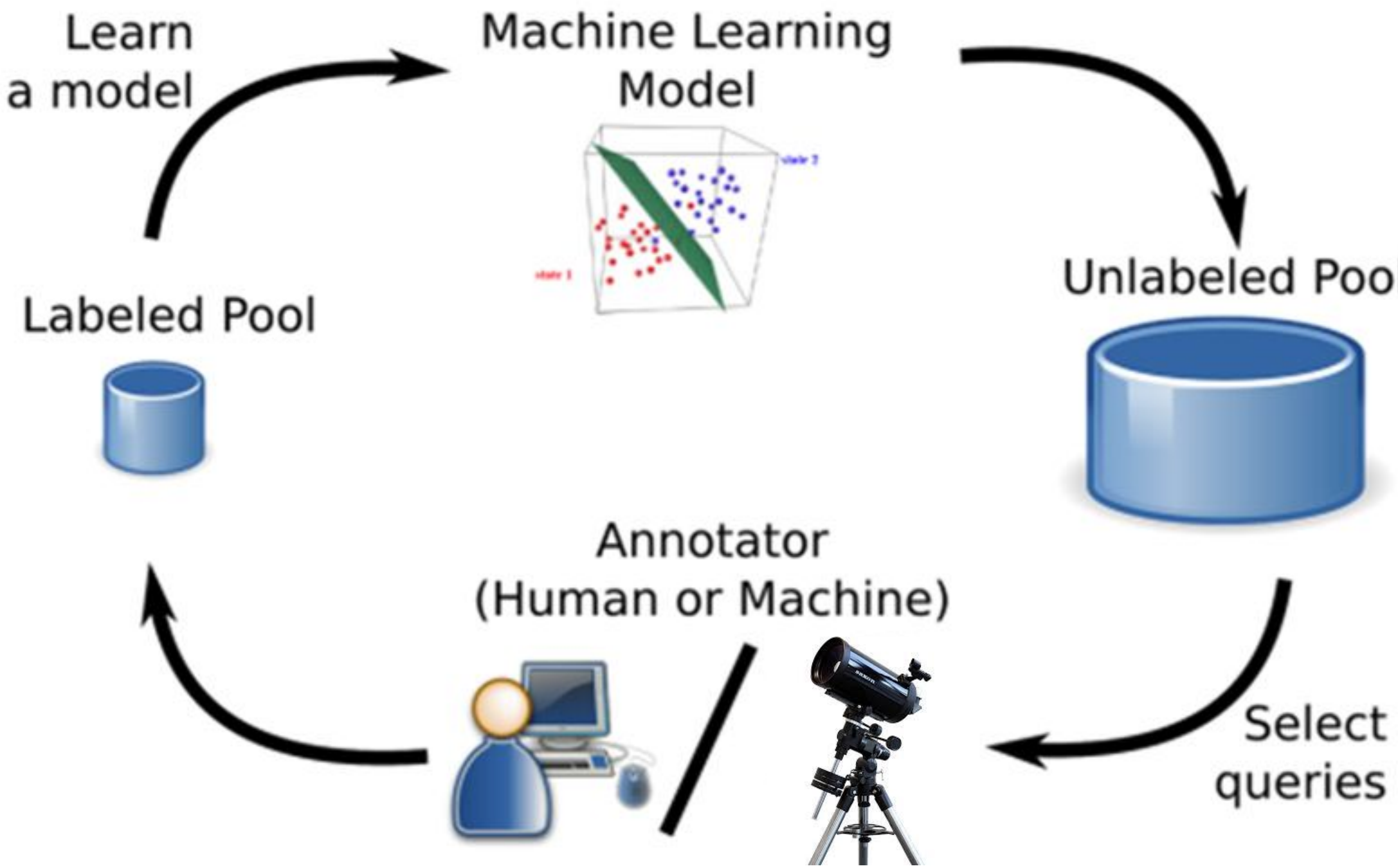


Machines need to evolve…

so they need to adapt!

Active Learning

Optimal classification, minimum training



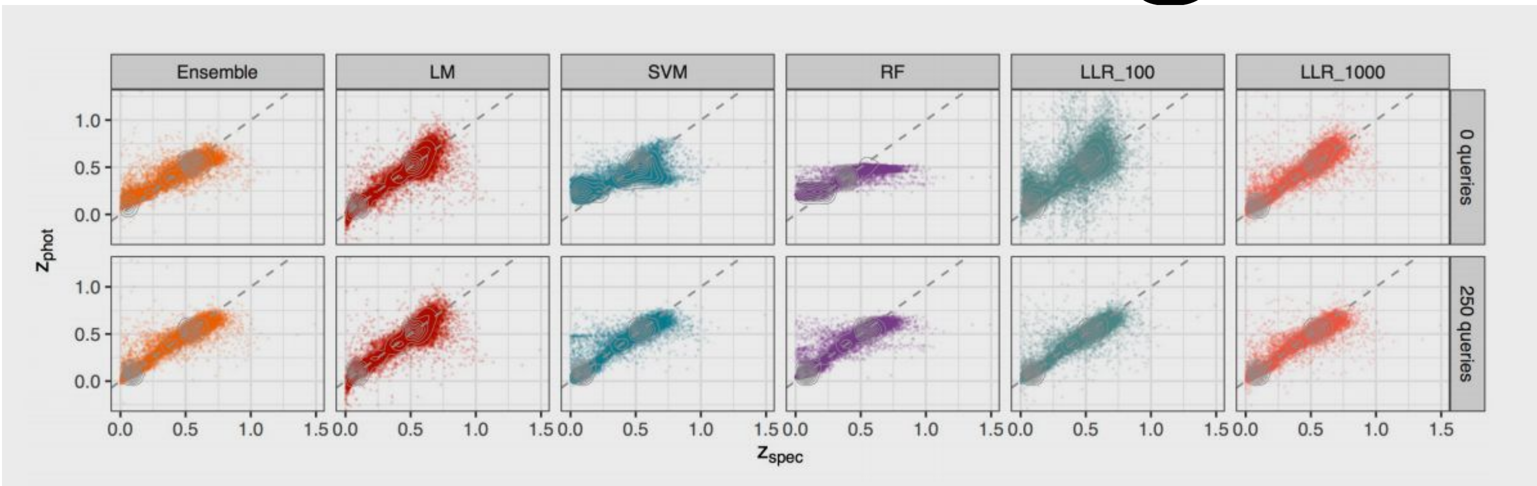
Optimal Experiment Design

In Statistics literature

$$PQ_{data,f}(x) \propto P_{x\sim data}(h_{train}(x) \neq f(x) \mid \textit{previous results})$$

- Pool based
- Generative
- Sequential

Active Learning for



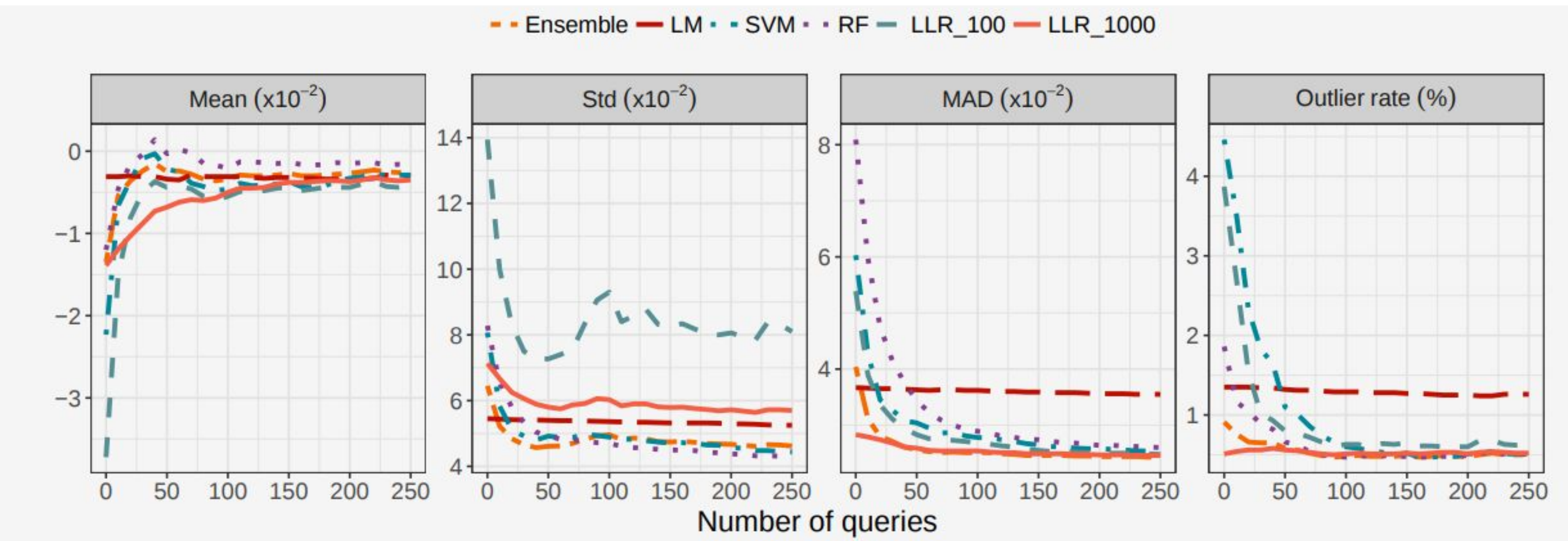


Figure 4. An assessment of the performance of the ensemble model and its constituent models using active learning. Performance diagnostics are shown as a function of the number of queries.

*Vilalta, **Ishida** et al., 2017 IEEE Symposium Series on Computational Intelligence (SSCI)*

AL for Supernova classification

A strategy

Parametric Fit
5 parameters

$$f(t) = A \frac{e^{-(t-t_0)/\tau_f}}{1 + e^{(t-t_0)/\tau_r}} + B,$$

Random Forest

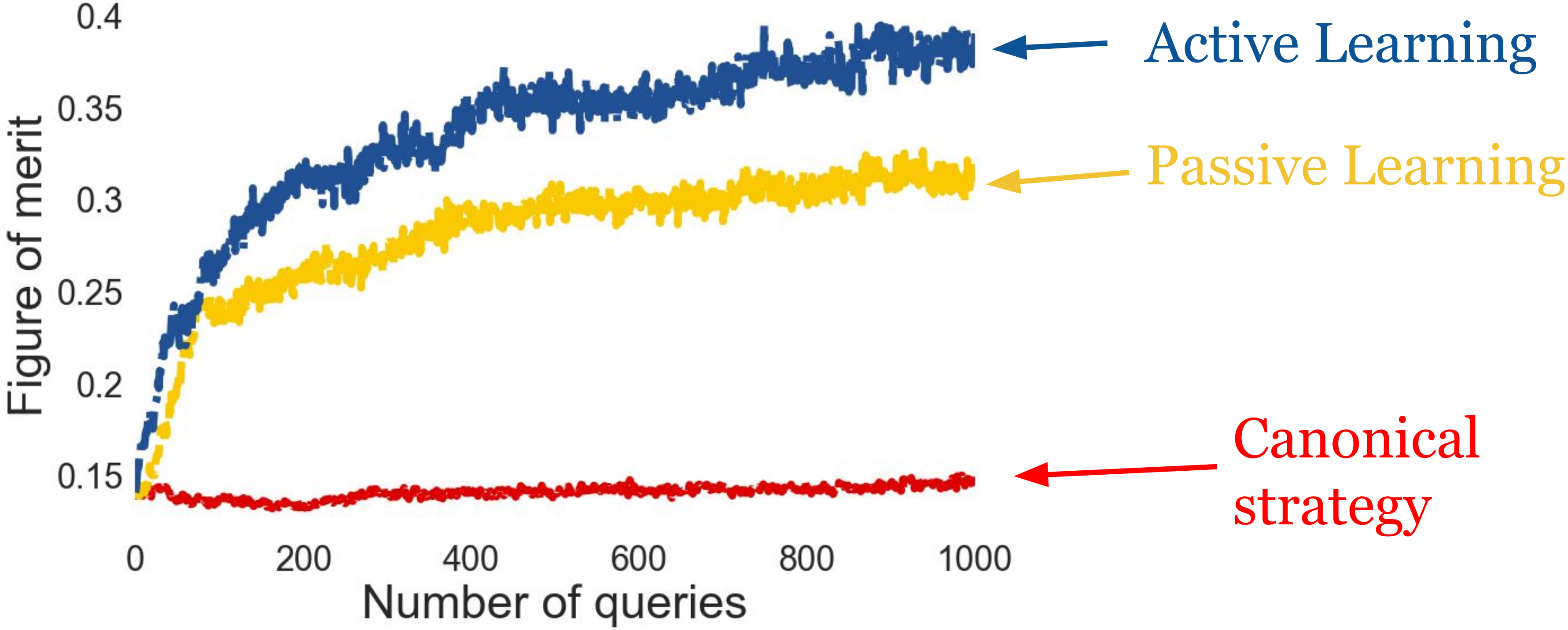
Photo-class uncertainty

Query the SN with highest uncertainty

LC+labels
LC

AL for SN classification

Static results

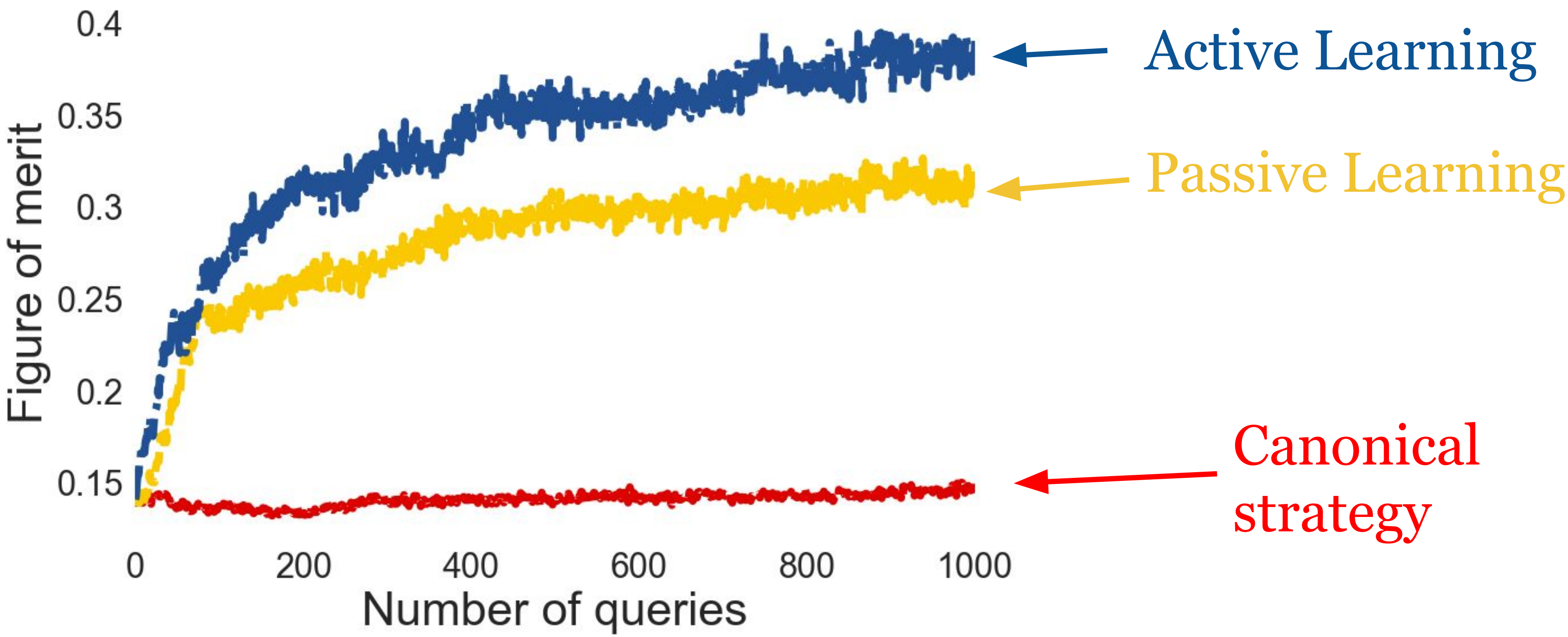


From COIN Residence Program #4, ***Ishida*** *et al., 2019, MNRAS, 483 (1), 2–18*

AL for SN classification

Static results

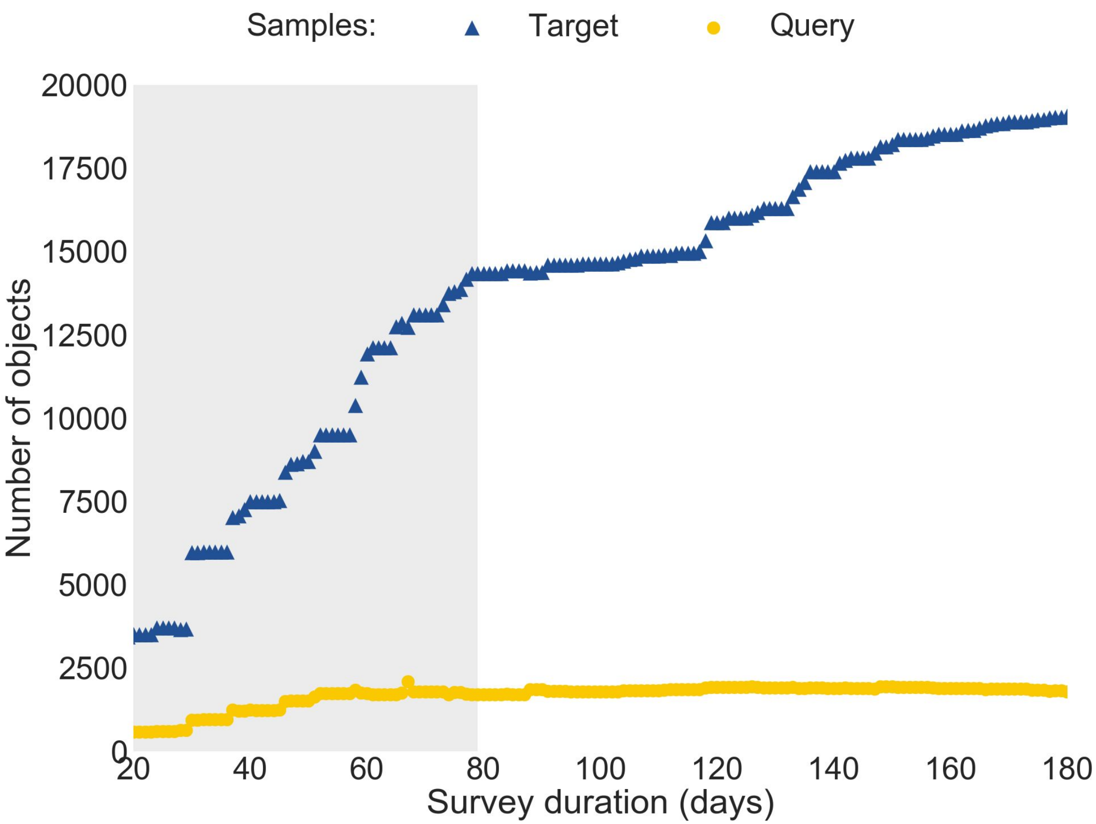
What astronomical aspect make this setting non-realistic?



From COIN Residence Program #4, ***Ishida*** *et al., 2019, MNRAS, 483 (1), 2–18*

SN are transients

Window of opportunity



1. Feature extraction done daily **with available observed epochs until then.**

2. Query sample is also re-defined daily: objects with **r-mag < 24**

From COIN Residence Program #4, ***Ishida*** *et al., 2019, MNRAS, 483 (1), 2–18*

Does this solve the problem completely?

No, it is just the best you can do!

Does this solve the problem completely?

No, it is just the best you can do!

Is this the only way of doing it?

No!, it is only one exciting possibility

Summary

*“**How** do we optimize machine learning results for astronomical purposes?”*

Adaptive Learning

designed for astronomical data

Final thoughts ….

We will have to adapt!

We are getting there...

Developments in human learning

Developments in human learning

Community code development

Developments in human learning

Community code development

Data challenges

kaggle

RAMP

Developed by Paris-Saclay Center for Data Science

Developments in human learning

Community code development

Data challenges

kaggle

RAMP

Developed by Paris-Saclay Center for Data Science

Large collaborations

euclid

LSST
Large Synoptic Survey Telescope

CERN

Good things happen when brains are connected properly...

COIN

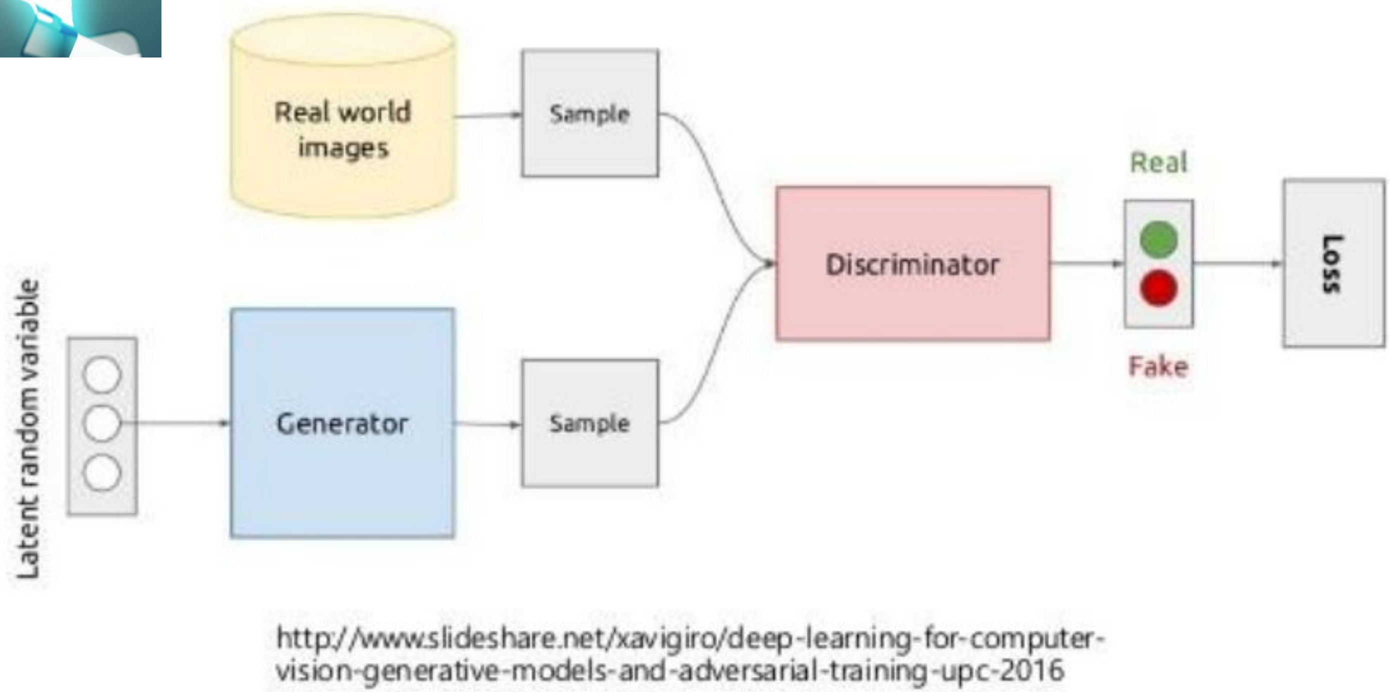
Cosmostatistics Initiative

https://cosmostatistics-initiative.org/

Extra slides

Adversarial Learning

The benefits of a worthy opponent



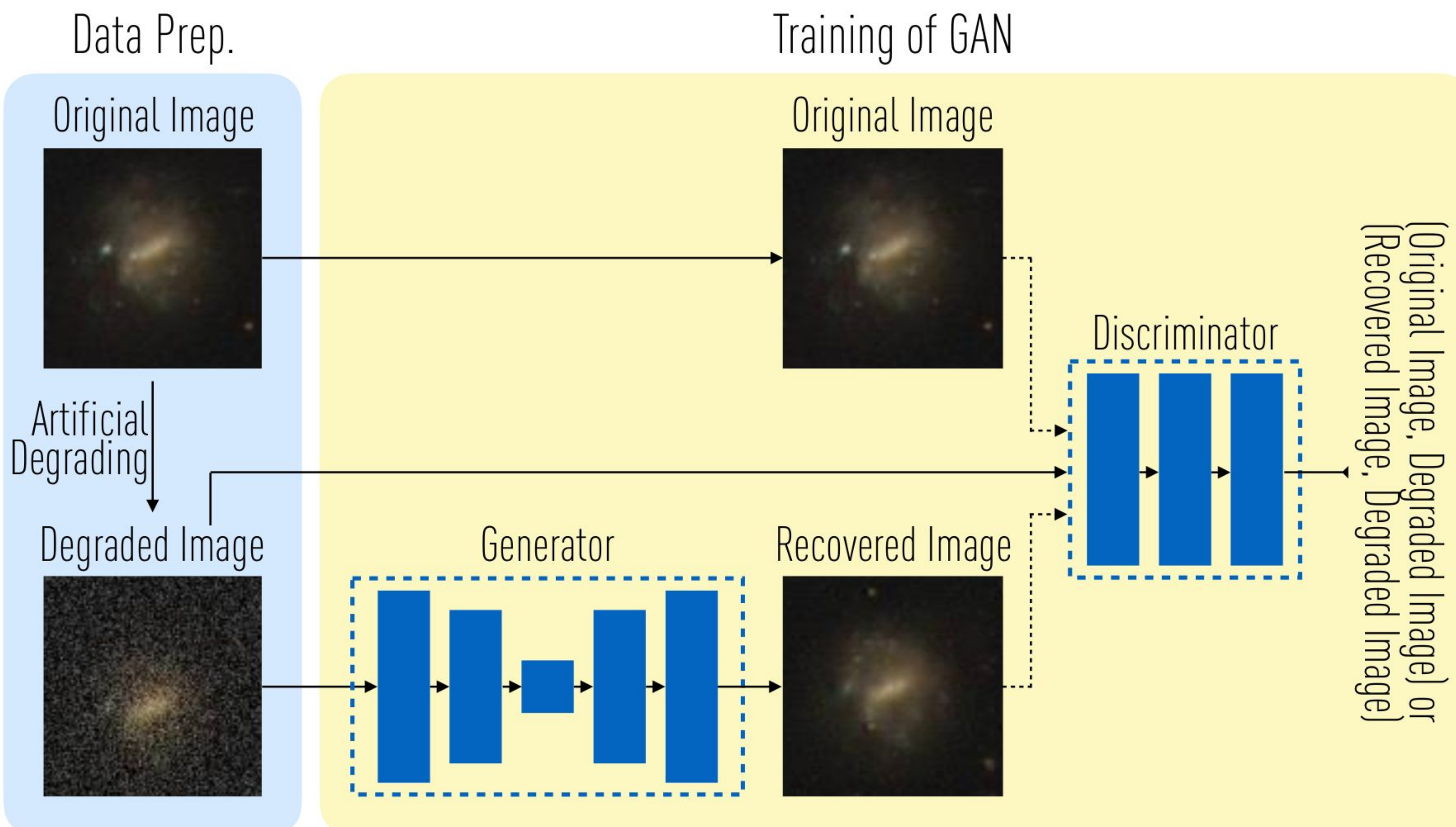
https://mascherari.press/introduction-to-adversarial-machine-learning/

Adversarial Learning

The benefits of a worthy opponent

K. Schawinski et al, 2017

In Astronomy

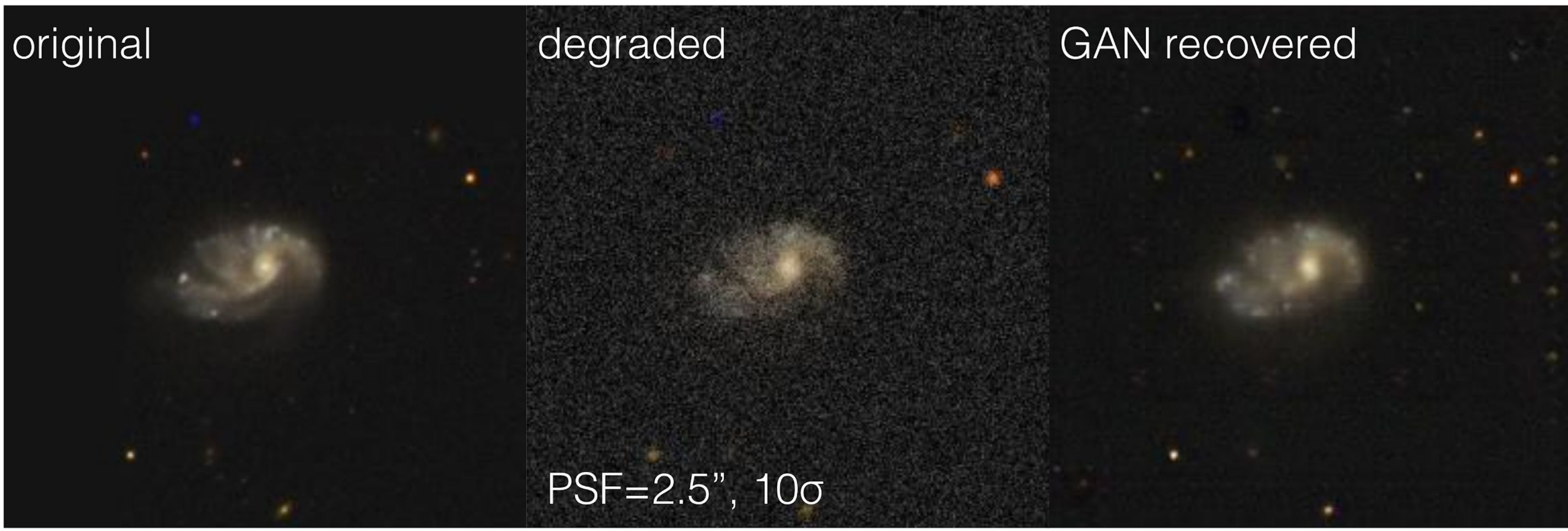


Adversarial Learning

The benefits of a worthy opponent

K. Schawinski et al, 2017

In Astronomy



Adversarial Learning

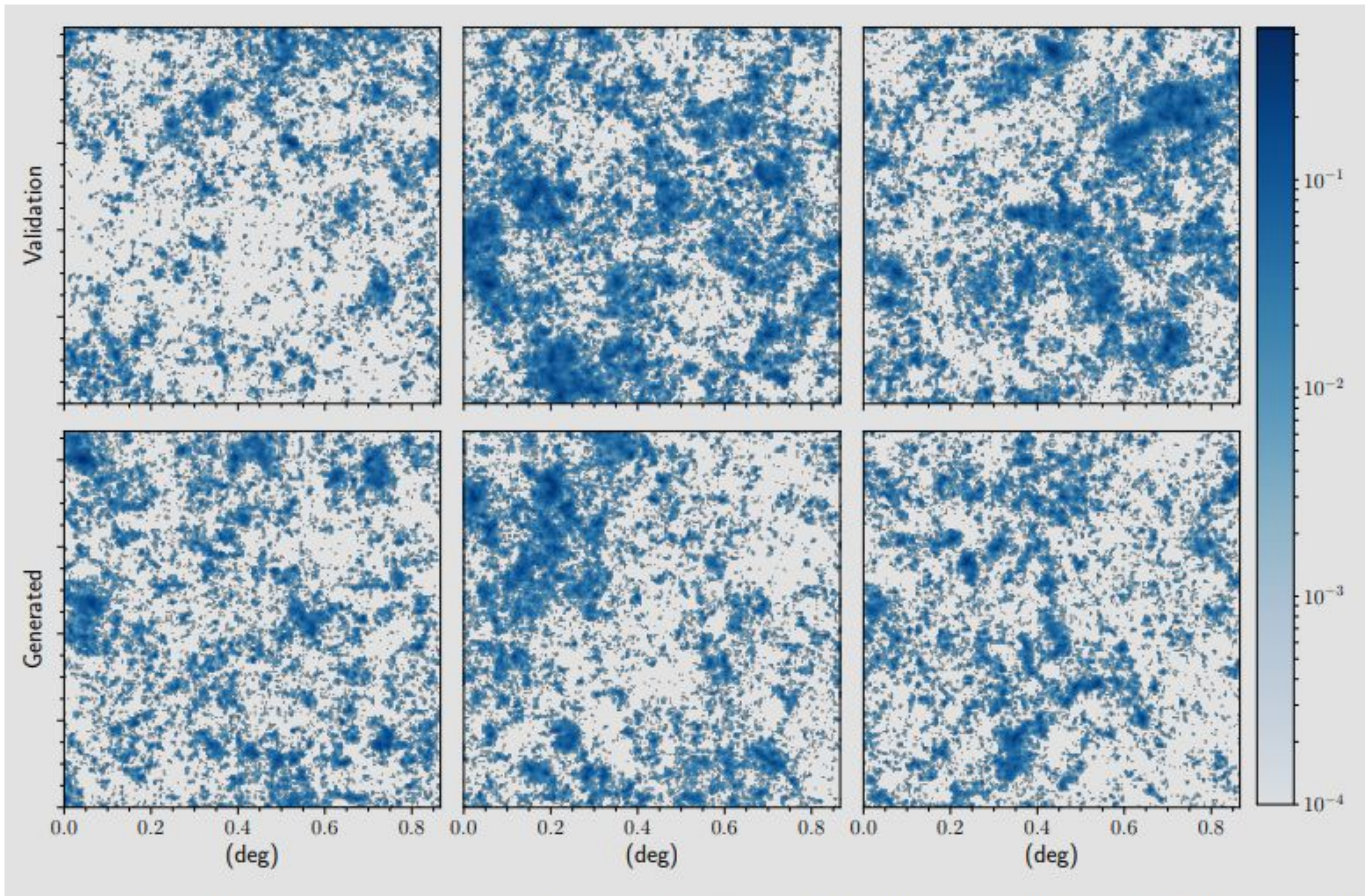
The benefits of a worthy opponent

In Cosmology

Mustafa et al, 2017 - CosmoGAN
Generating cheap WL maps

Original

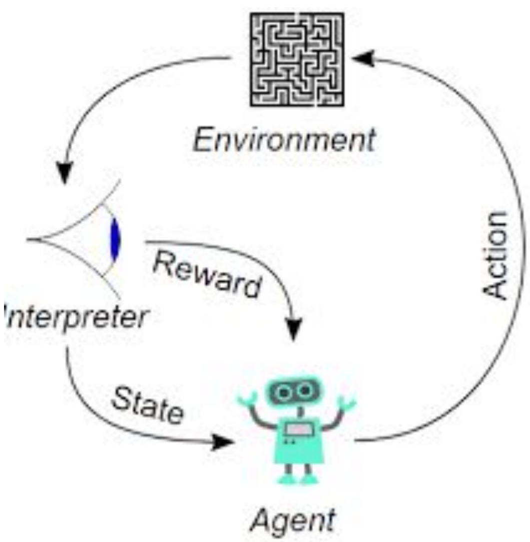
Generated

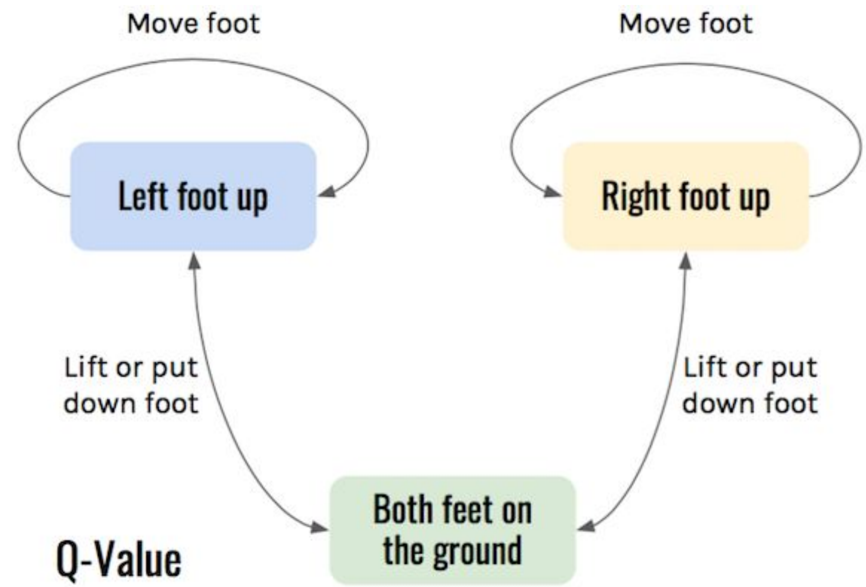


https://youtu.be/t81QhHaMS7w

Reinforcement Learning

The importance of feedback



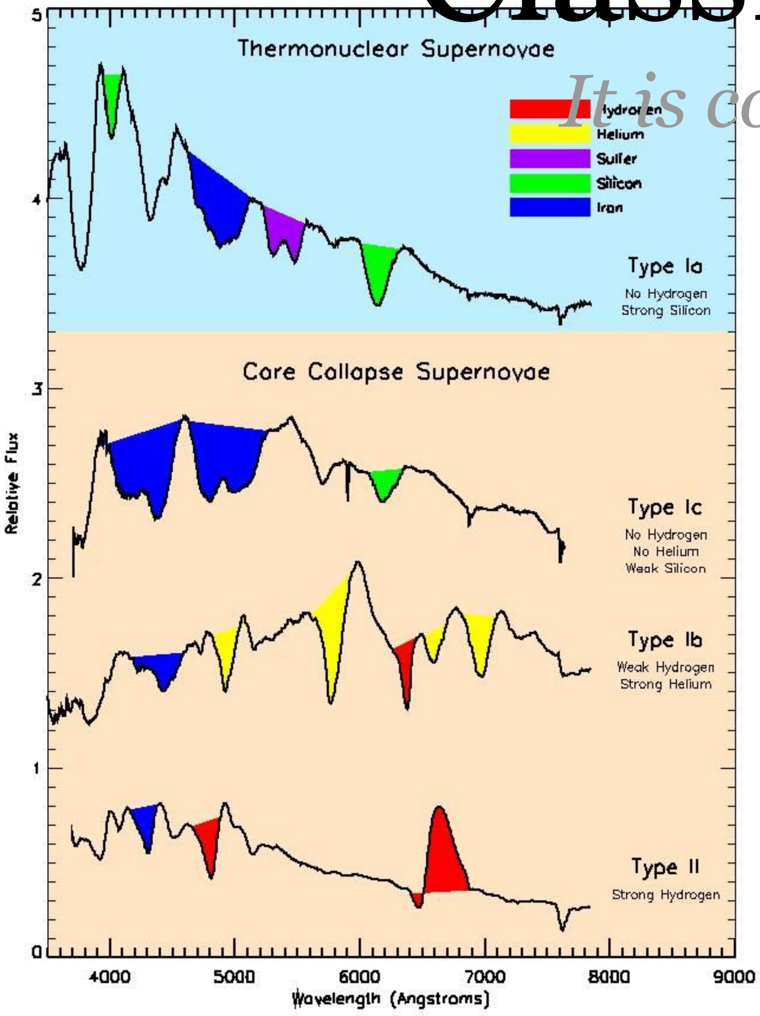


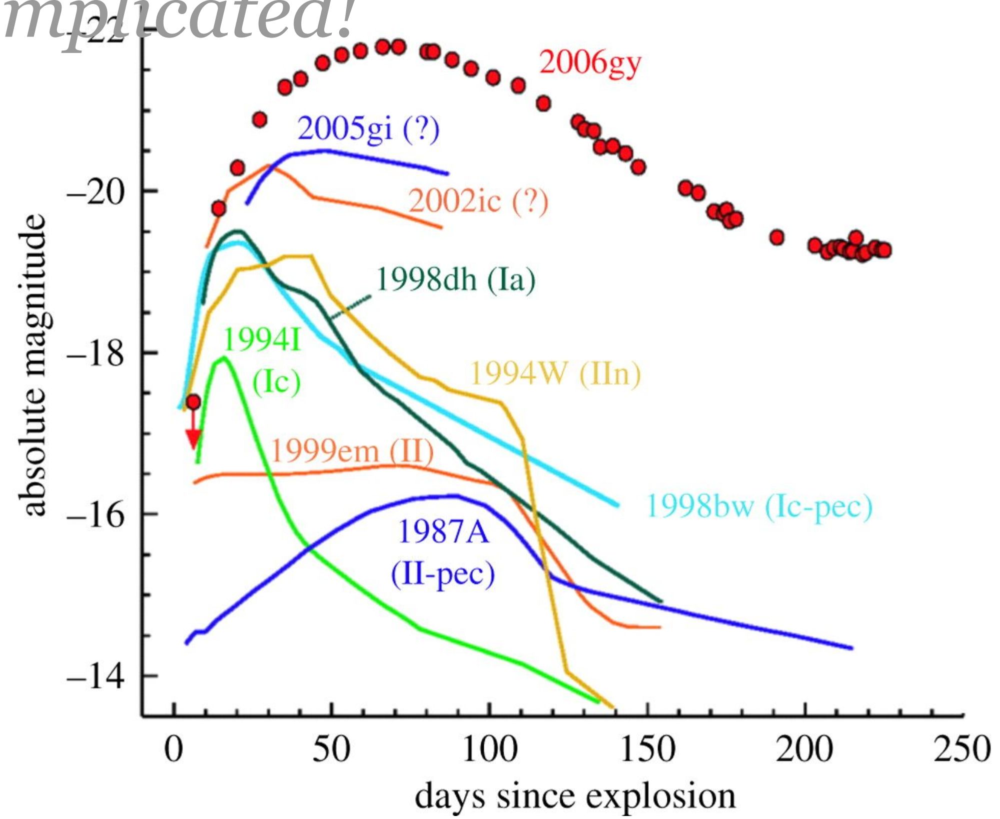
State	Action	Q-Value
Left foot up	Move foot forward	+ 0.5
Left foot up	Put foot down	+ 0.0
Left foot up	Move foot backward	- 0.5
...	...	...

https://hackernoon.com/using-q-learning-to-teach-a-robot-how-to-walk-a-i-odyssey-part-3-5285237cc3b1

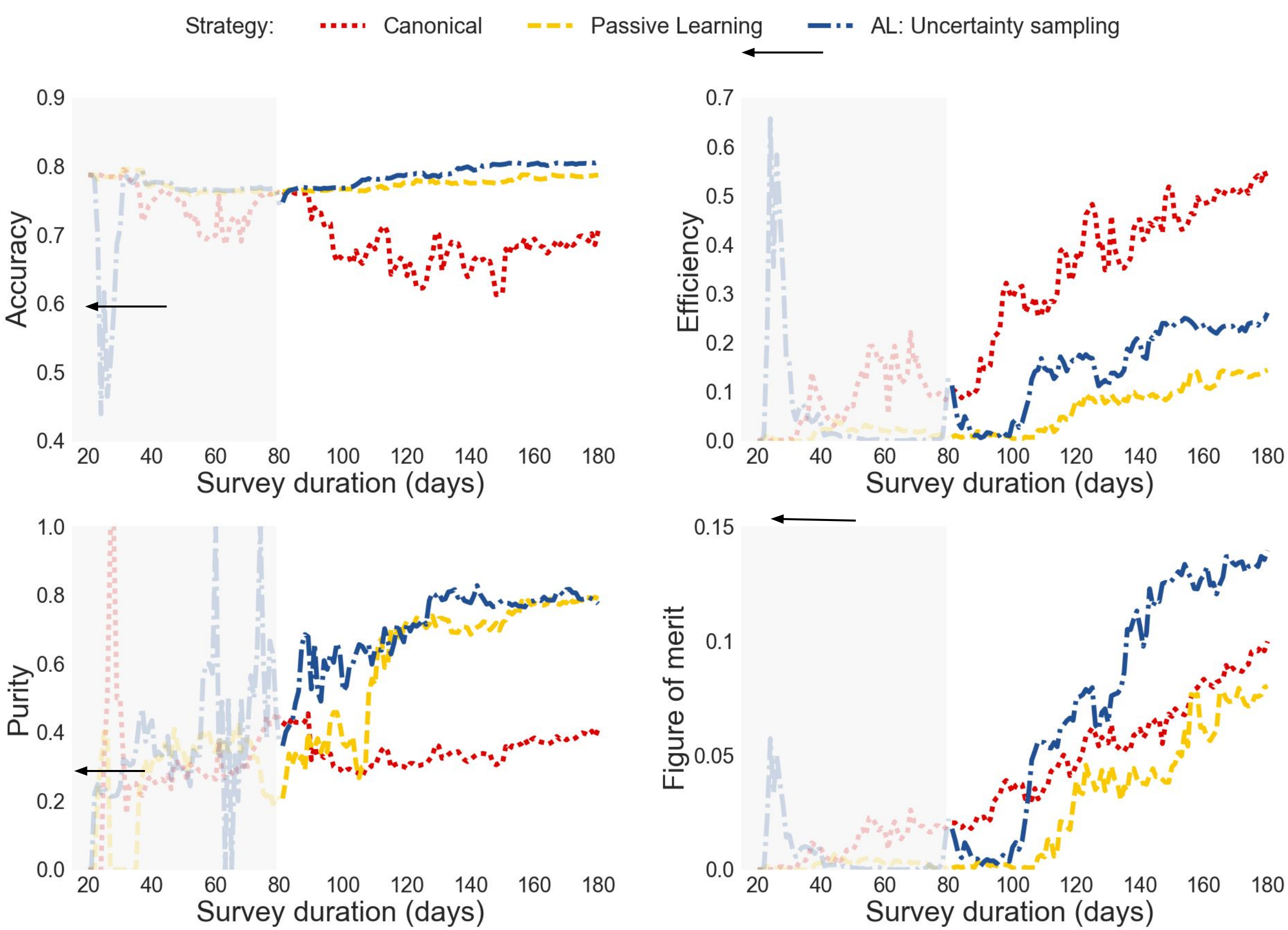
SN Photometric Classification

It is complicated!





No initial training



The arrow shows <u>traditional</u> Full light-curve results with full SNPCC spec

From COIN Residence Program #4, ***Ishida*** *et al., 2019, MNRAS, 483 (1), 2–18*

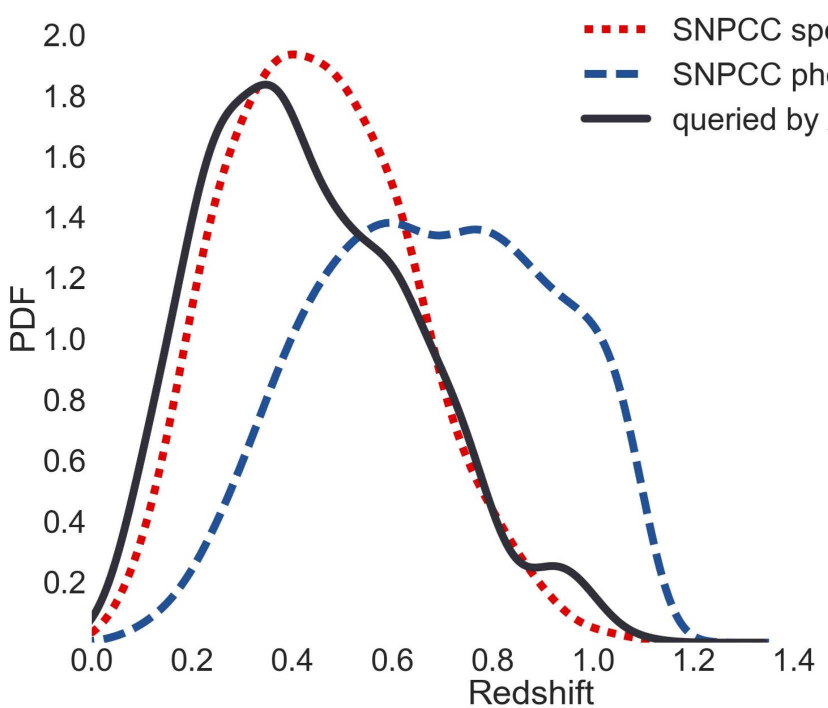
The queried sample

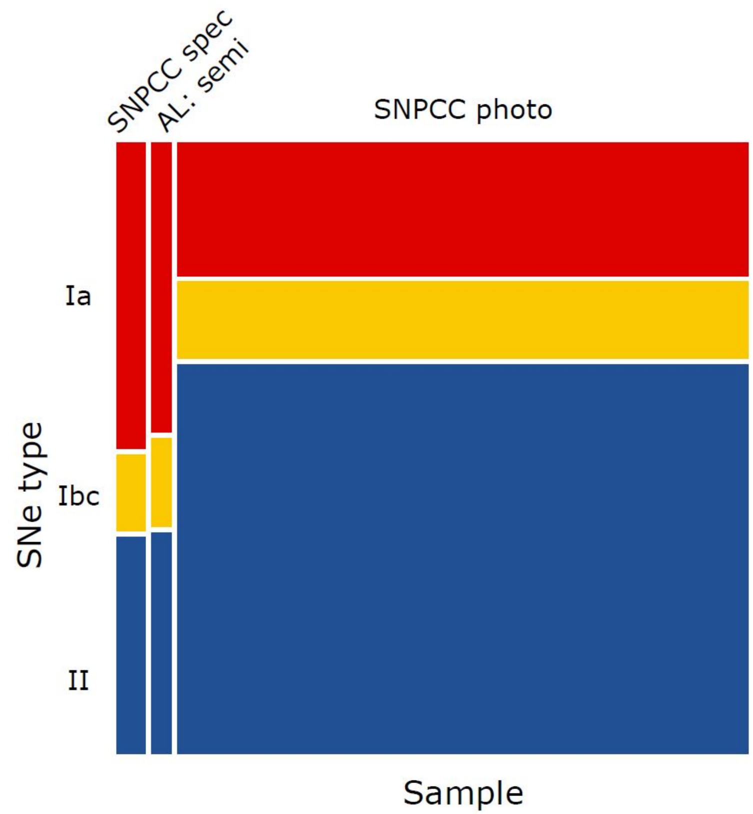
Partial LC, no training, time domain, batch

SNPCC spec:
1103 objects

Queried sample:
800 objects

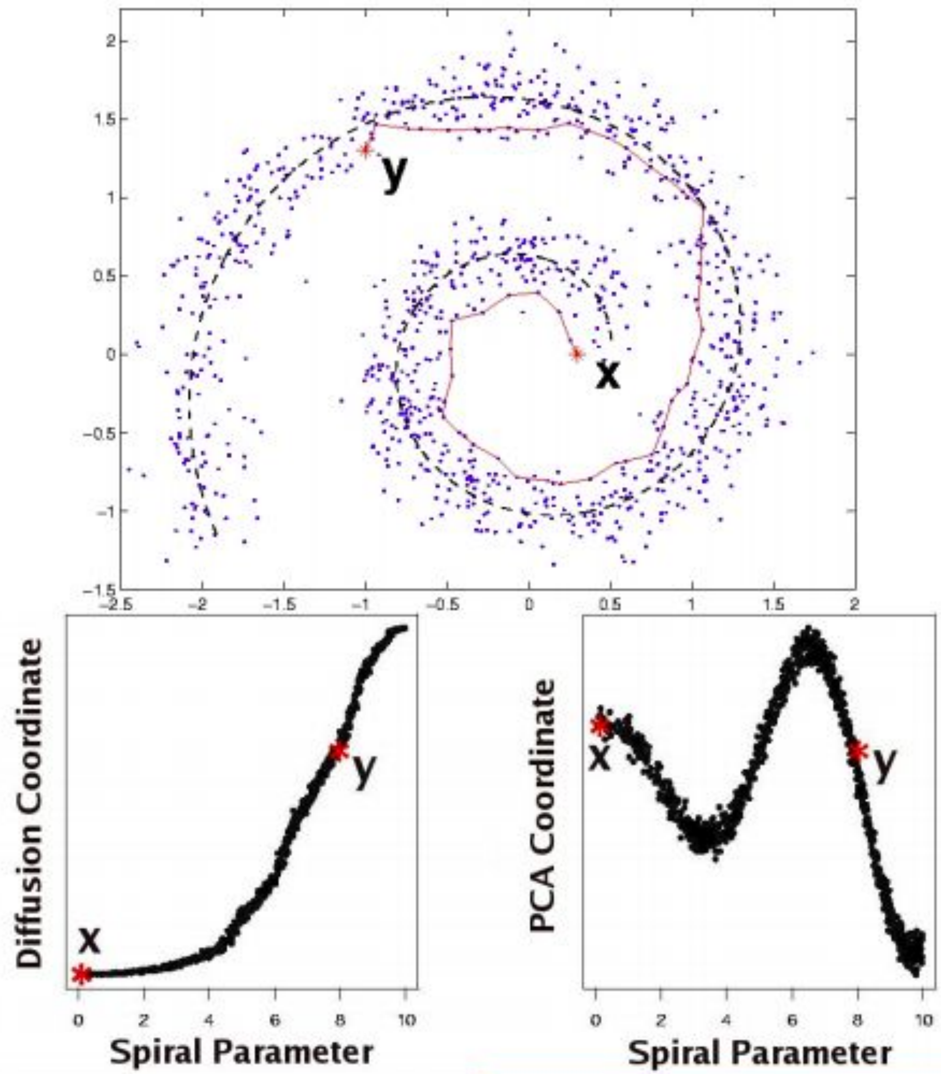
Telescope time:
Queried/spec = 0.999



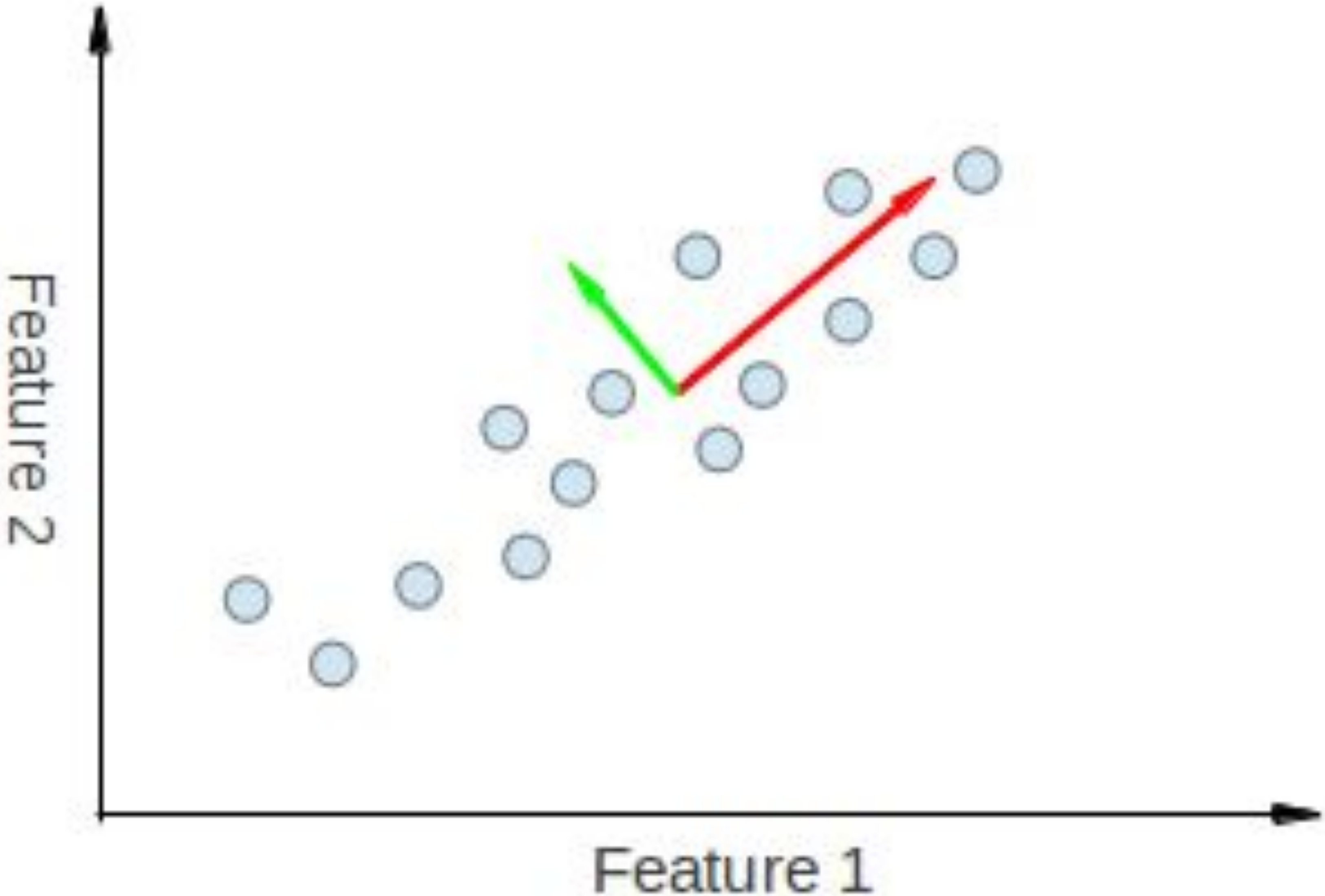


From COIN Residence Program #4, ***Ishida*** *et al., 2019, MNRAS, 483 (1), 2–18*

Diffusion Map: Spiral Example



Principal Component Analysis



Acknowledgement

- H2020-Astronomy ESFRI and Research Infrastructure Cluster (Grant Agreement number: 653477).