Public version

Data Science in Astronomy

Feel free to ask questions. Comments are also welcome.

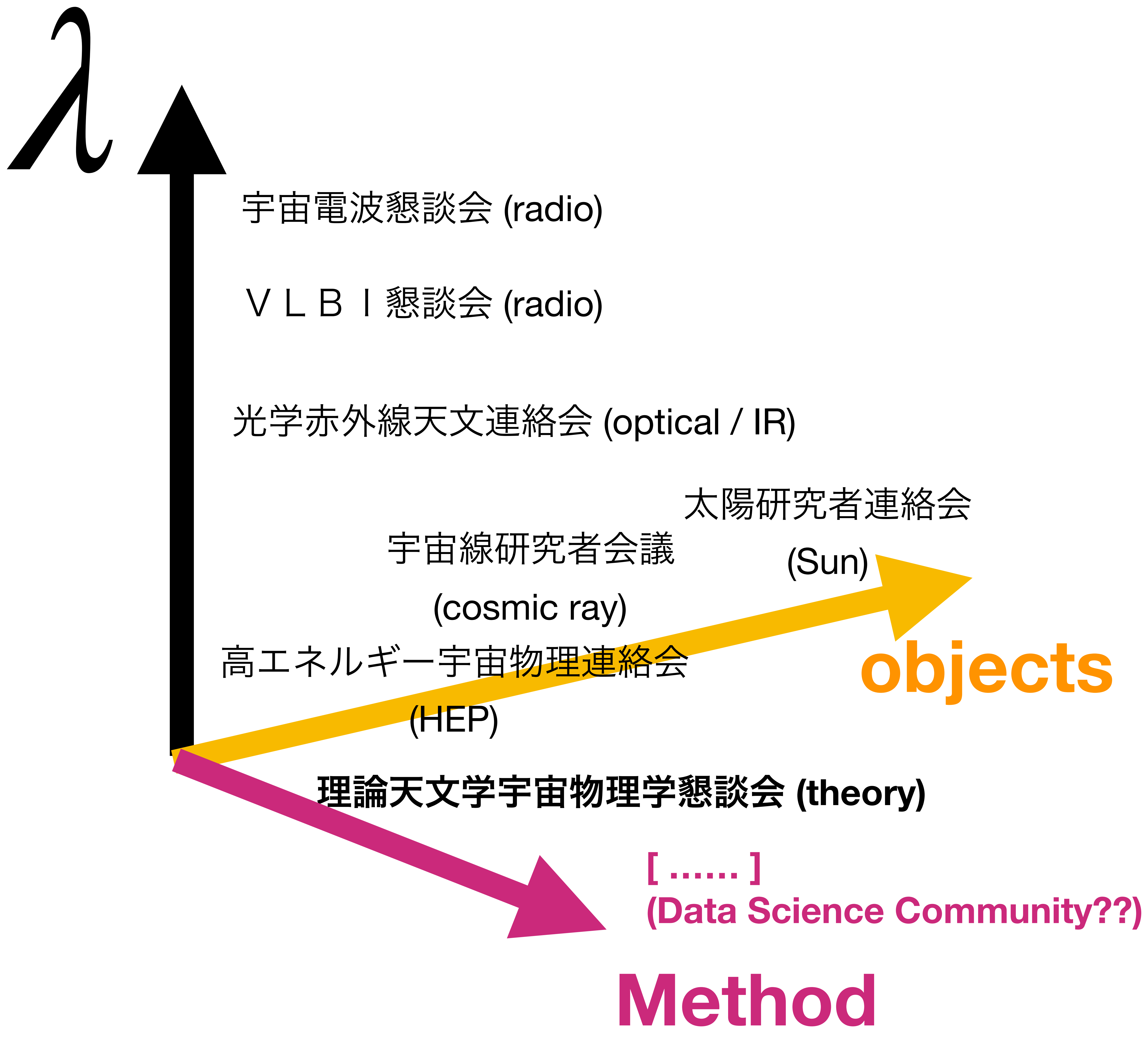
khattori@ism.ac.jp

Kohei Hattori (服部公平)

国立天文台 NAOJ National Astronomical Observatory of Japan

統計数理研究所 The Institute of Statistical Mathematics

No community of Data Science !?



2021年 春季年会

年会プログラム

於　東京工業大学 (オンライン開催)

2021 年 3 月 16 日 (火) ～ 3 月 19 日 (金)

Data Science Session
(1st time in history, I guess.)

Z1. 天文データ科学

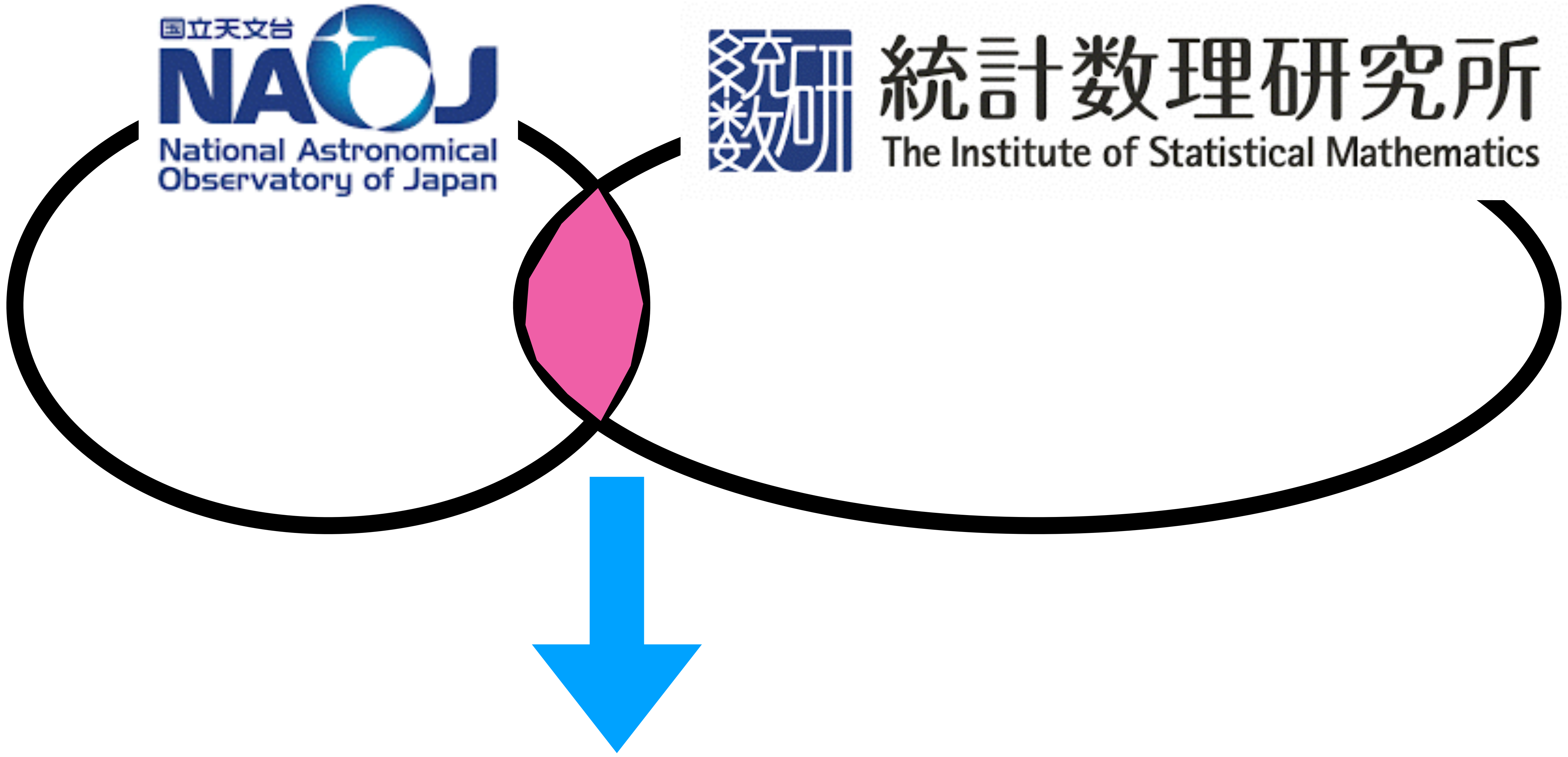
Instrument Sessions
(long history)

V1. 観測機器 (電波)
V2. 観測機器(光赤・重)
V3. 観測機器(X線・γ線)

3月16日 (火)	A	Z3. 計算宇宙惑星
	B	W. コンパクト天体
	C	S. 活動銀河核
	D	V2. 観測機器(光赤・重)
	E	V3. 観測機器(X線・γ線)
	F	P2. 原始惑星系円盤
	G	Q. 星間現象
	H	
3月17日 (水)	A	Z3. 計算宇宙惑星
	B	W. コンパクト天体
	C	R. 銀河
	D	V2. 観測機器(光赤・重)
	E	V3. 観測機器(X線・γ線)
	F	P2. 原始惑/P1.星形成
	G	Q. 星間現象
	H	Y. 教育・広報・他
3月18日 (木)	A	Z1. 天文データ科学
	B	W. コンパクト天体
	C	X. 銀河形成・進化
	D	V1. 観測機器 (電波)
	E	T. 銀河団
	F	P1. 星形成
	G	M. 太陽
	H	
3月19日 (金)	A	Z2. ngVLA の天文学
	B	N. 恒星進化
	C	X. 銀河形成・進化
	D	V1. 観測機器 (電波)
	E	
	F	
	G	M. 太陽
	H	P3. 惑星系

Institute of Statistical Mathematics (ISM = 統計数理研究所)

三鷹　　立川



- **Masato Shirasaki**
- **Kohei Hattori (myself)**

We are trying to connect astronomers and statisticians.

ISM webpage — https://www.ism.ac.jp

An ideal workplace for astro data scientists.

Astronomers can use various resources.

・共同研究スタートアップ

Data analysis consulting. **(First consulting is free.)**
Any questions from astronomers are welcome!


大学共同利用機関法人 情報・システム研究機構
統計数理研究所 統計思考院

トップ | 統計思考院とは | 統計思考院の事業活動 | 関連する各種提言

思考院トップ > 統計思考院の事業活動 > 共同研究スタートアップ

共同研究スタートアップ

「統計的かもしれない」疑問にお答えする場を提供します。

新型コロナウイルスの感染拡大防止のため、オンラインでのご相談もお受けしています。

データ解析・統計分析に関わる問題でお悩みの方々のために『共同研究スタートアップ』プログラムを用意しています。

本プログラムの命名には、日常的な相談の中に統計学研究のテーマを見つけ出し、共同研究に発展させたいという期待が込められていますが、一般的な統計リテラシー普及の役割も担っています。統計数理研究所までお越しいただいて直接お話を伺うことを原則としておりますが、混雑時には面会の約束がかなり先になることを予めご了承ください。

助言の対象範囲には、取得済みのデータに関する解析法だけでなく、データの収集法、調査の方法、実験の計画といった事項も含まれます。

https://www.ism.ac.jp/shikoin/startup/index.html
https://www.ism.ac.jp/shikoin/overview/index.html

・Seminars

- 天文観測におけるビッグデータ解析と宇宙論パラメータの推定 **(hosted by M. Shirasaki)**
- 統計物理と統計科学のセミナー **(hosted by ISM people)**

・受託研究員制度

(Need to pay. Visiting scholar / visiting PhD student)

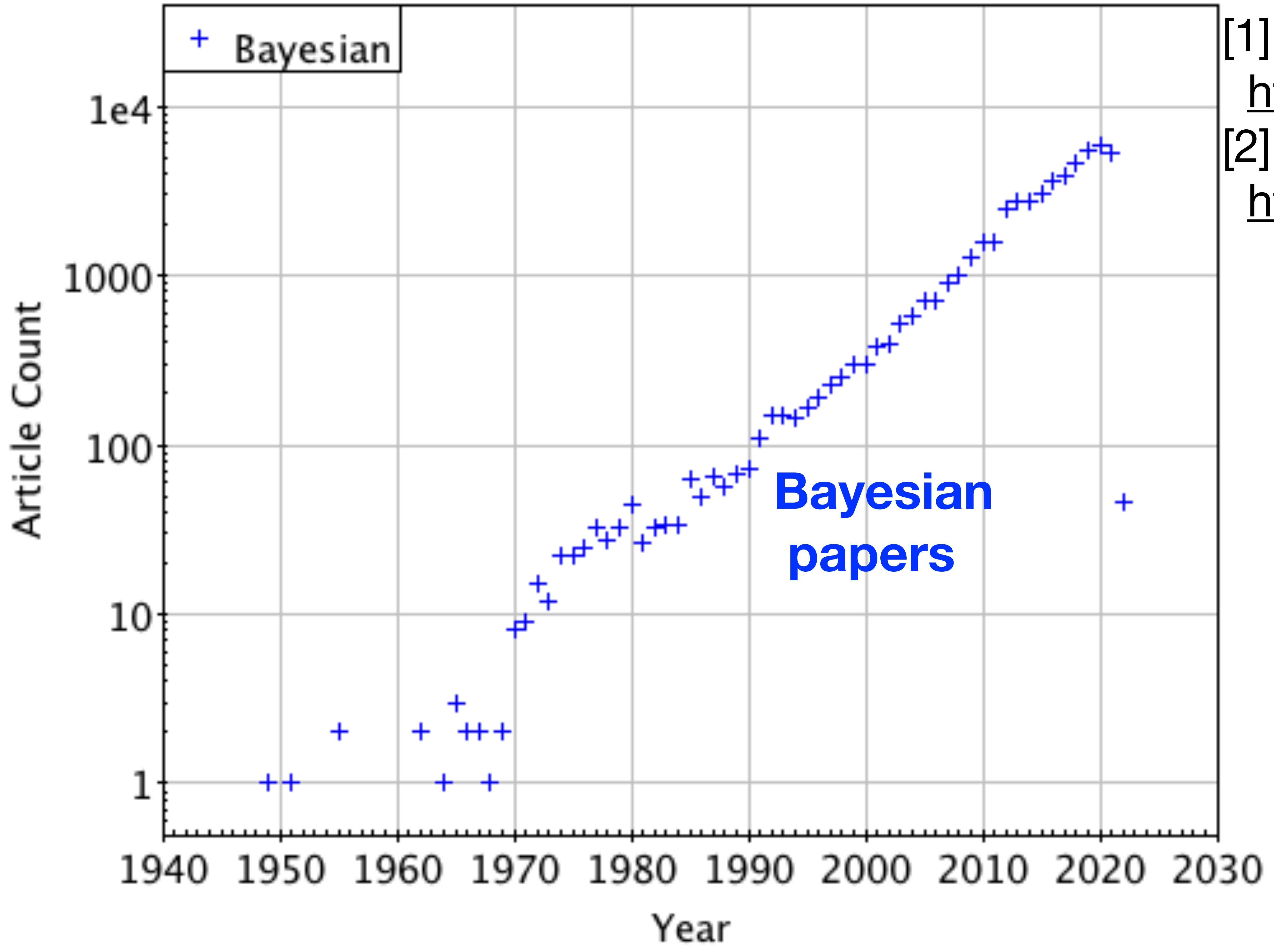
(I started mentoring an astro PhD student.)

Most importantly, people at ISM
are supportive, and
are HIGHLY interested in ASTRONOMY data.

Trends in ADS papers

ADS papers with “Bayesian” in the abstract.

Exponential growth. Time scale ~7 years



Unfamiliar with Bayes? See, e.g.,

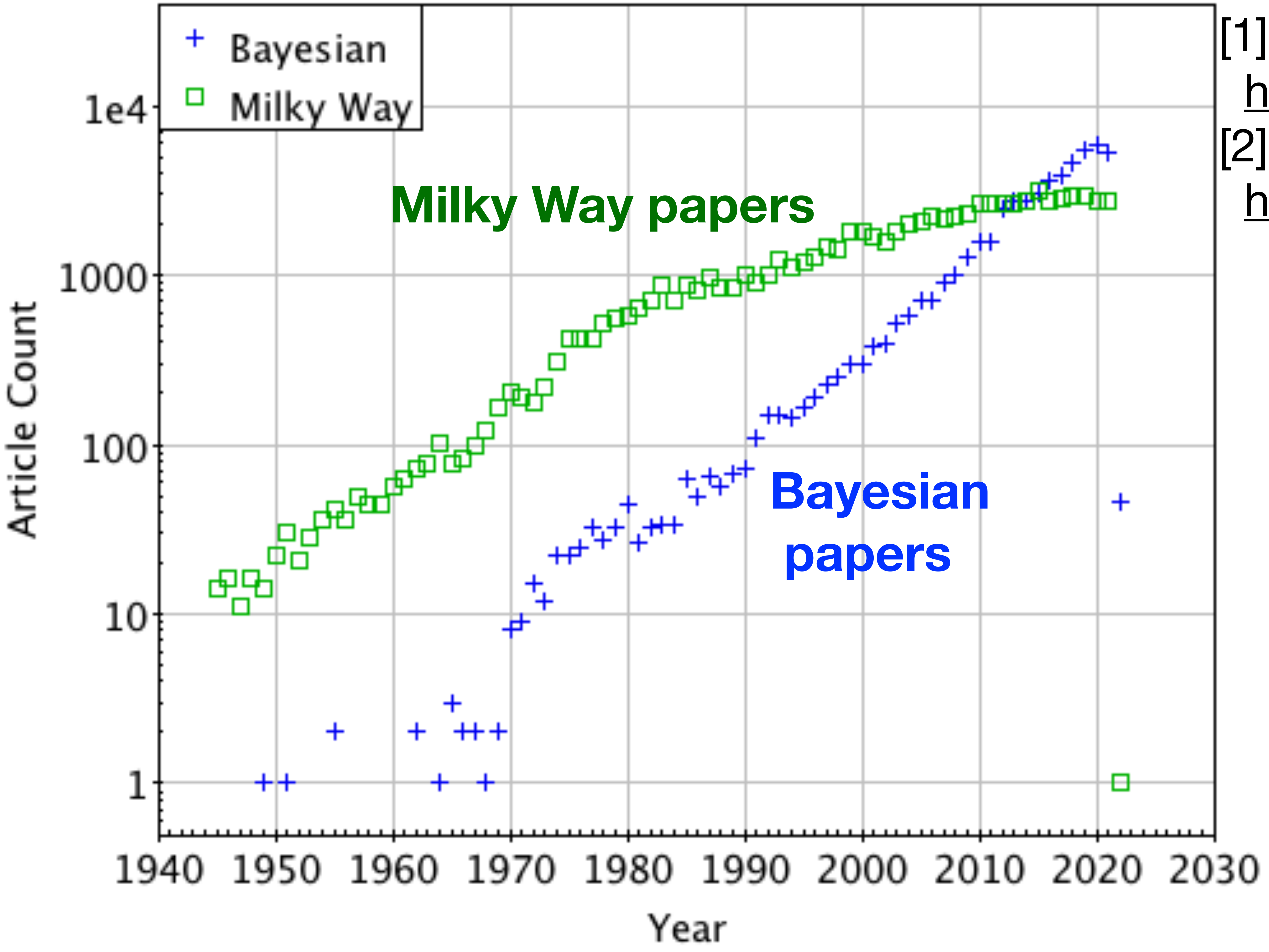
[1] Data analysis recipes: **Fitting a model to data**
https://arxiv.org/abs/1008.4686

[2] Data analysis recipes: Probability calculus for inference
https://arxiv.org/abs/1205.4446

Trends in ADS papers

ADS papers with "Bayesian" in the abstract.
ADS papers with "Milky Way" in the abstract.

Exponential growth. Time scale ~7 years



Unfamiliar with Bayes? See, e.g.,

[1] Data analysis recipes: **Fitting a model to data**
https://arxiv.org/abs/1008.4686

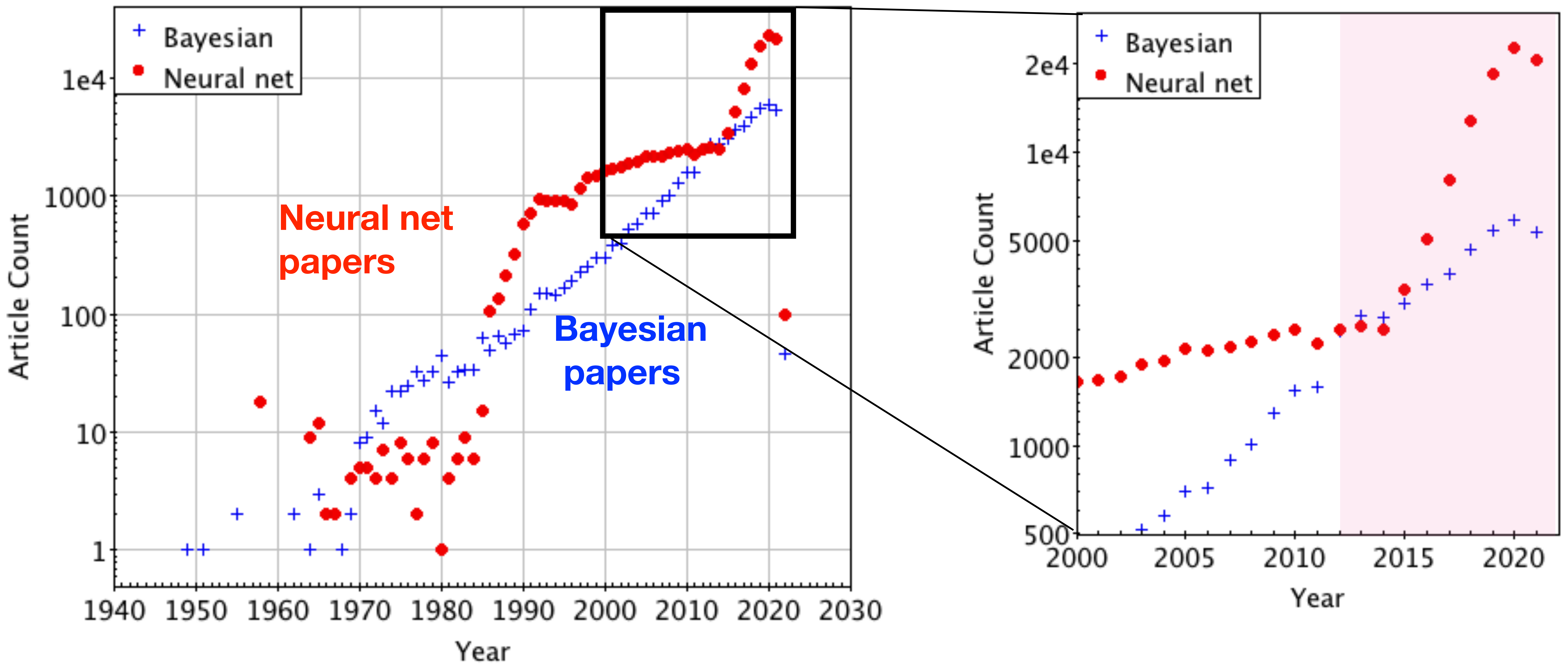
[2] Data analysis recipes: Probability calculus for inference
https://arxiv.org/abs/1205.4446

astrophysics data system

Trends in ADS papers

ADS papers with **“Bayesian”** in the abstract. **Exponential growth. Time scale ~7 years**

ADS papers with **“neural net”** in the abstract. **Exponential growth.**
After 2015, the time scale is ~2.5 years

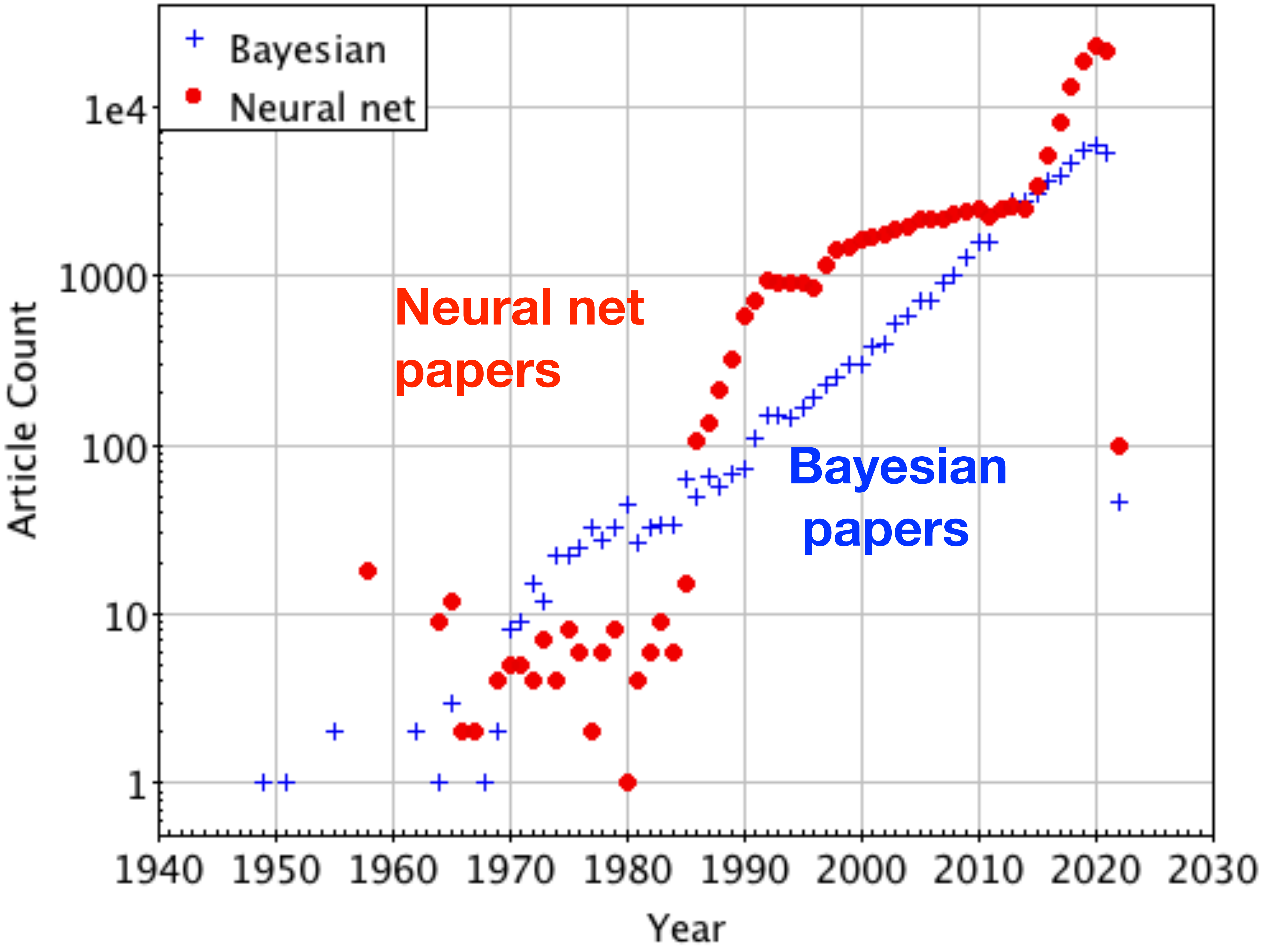


Trends in ADS papers

ADS papers with "Bayesian" in the abstract.

ADS papers with "neural net" in the abstract.

Exponential growth.
After 2015, the time scale is ~2.5 years



Inflation of ML universe.

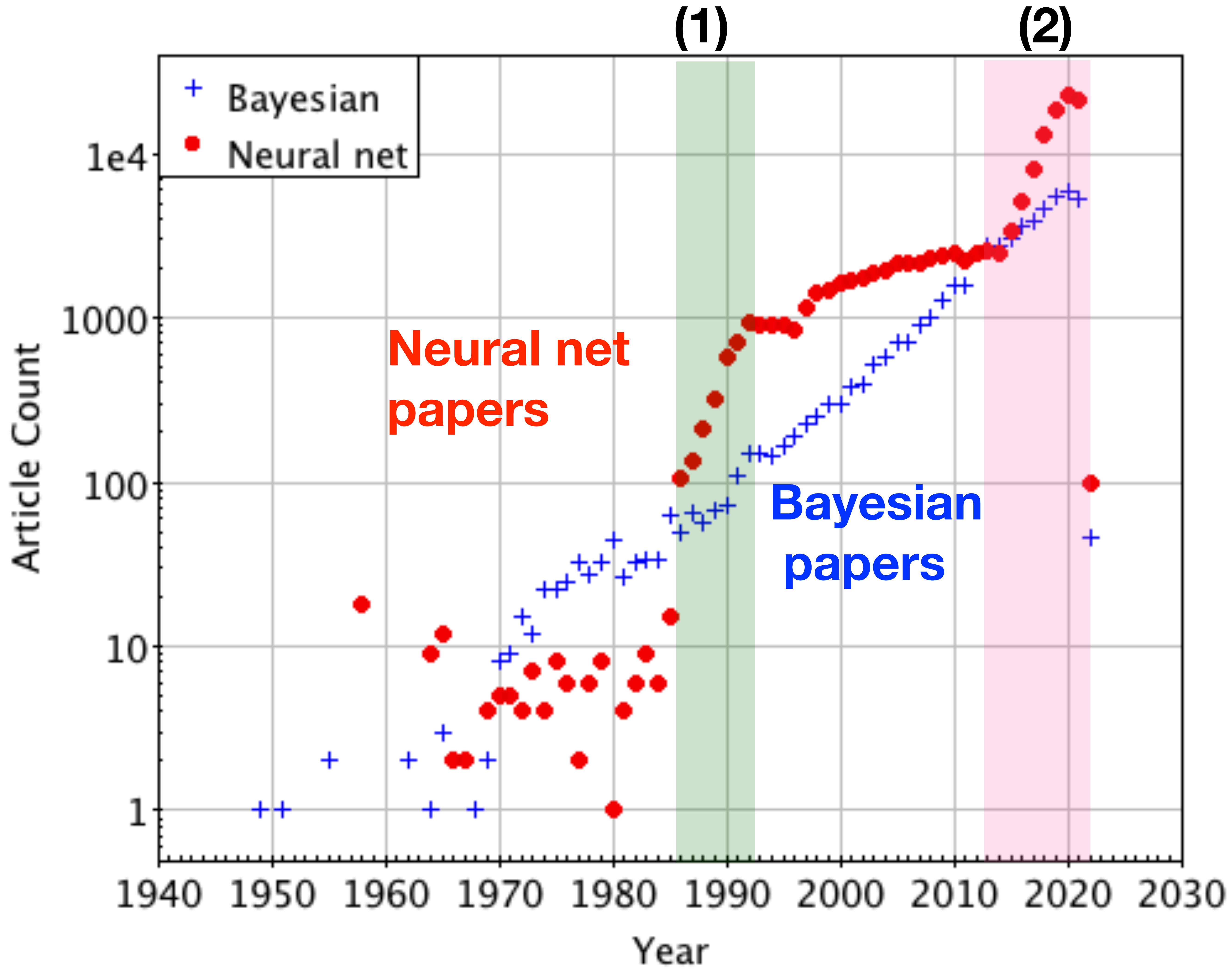
Japanese PhD course ~ 5 years
Are we ready to train PhD student
in this expansion of the ML world?

Trends in ADS papers

ADS papers with "Bayesian" in the abstract.

ADS papers with "neural net" in the abstract.

Exponential growth.
After 2015, the time scale is ~2.5 years

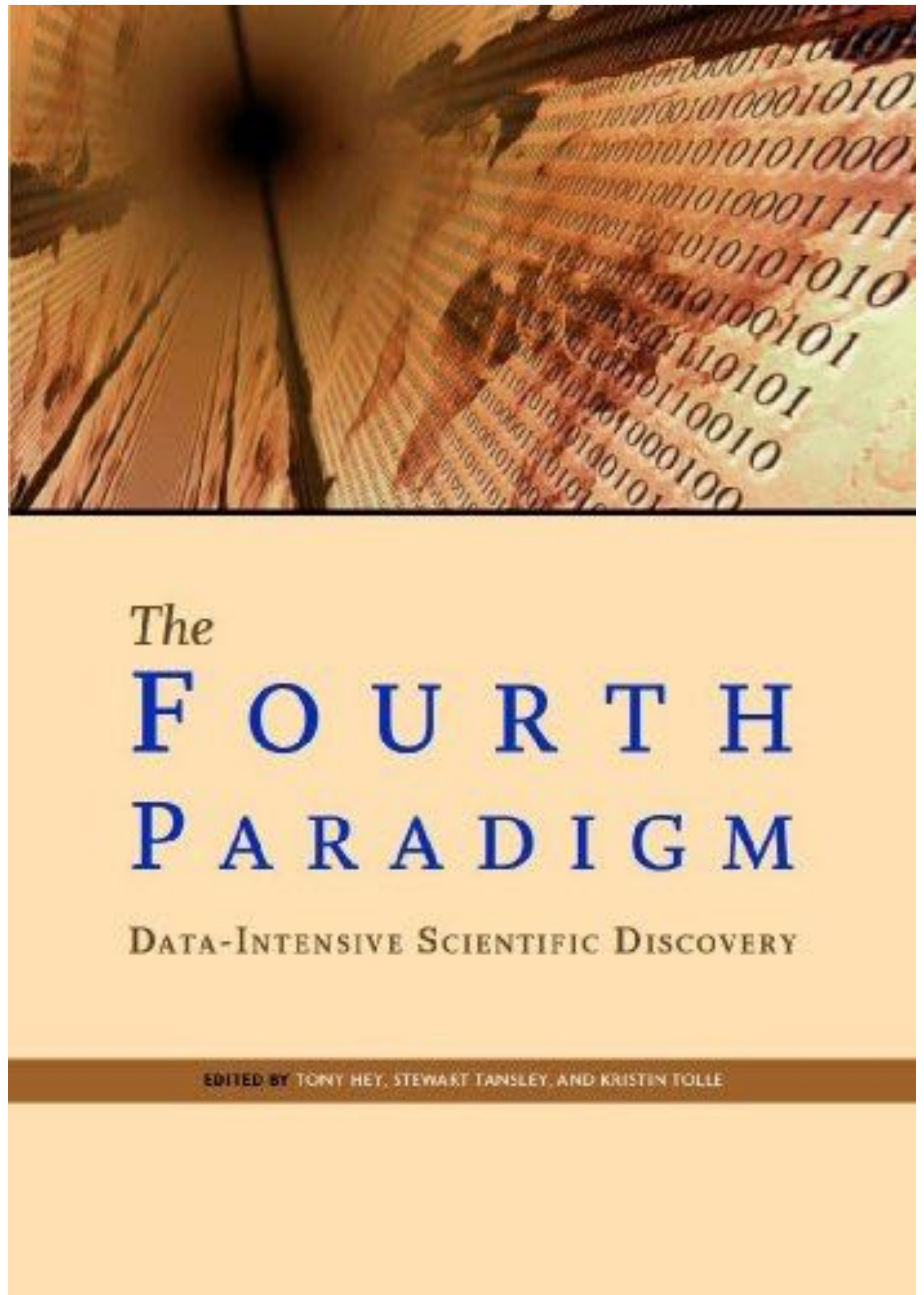


(1) Back propagation method [誤差逆伝播法]

(2) Deep learning (AlexNet) [深層学習]

New techniques → **Speed-up of science**

History of science



1. Empirical science
   in the last ~1000s of years

2. Theoretical science
   in the last ~100s of years

3. Computational science
   in the last several decades

4. **Data exploration**
   Now

Today's talk

- (1) Era of big data
- (2) Dimensionality reduction
- (3) Sparsity
- (4) Bayesian analysis
- (5) Machine learning
- (6) Neural network
- (7) Data challenge

- (1) Era of big data
- (2) Dimensionality reduction
- (3) Sparsity
- (4) Bayesian analysis
- (5) Machine learning
- (6) Neural network
- (7) Data challenge

Era of Big Data

Keywords for astro observations

- wide-field
- deep
- high-resolution
- high-dimensional (many bands)
- time-domain
- high-precision
- large data volume (e.g., radio telescope)

Data are becoming larger / more complex.
Most data will never be “seen” by eye.

Need to automate
data acquisition / reduction / analysis
… **Data science challenge**

Main part

- (1) Era of big data

Trends in statistical mathematics

- (2) Dimensionality reduction
- (3) Sparsity
- (4) Bayesian analysis

Trends in machine learning

- (5) Machine learning
- (6) Neural network

Robustness / reliability of new methods

- (7) Data challenge

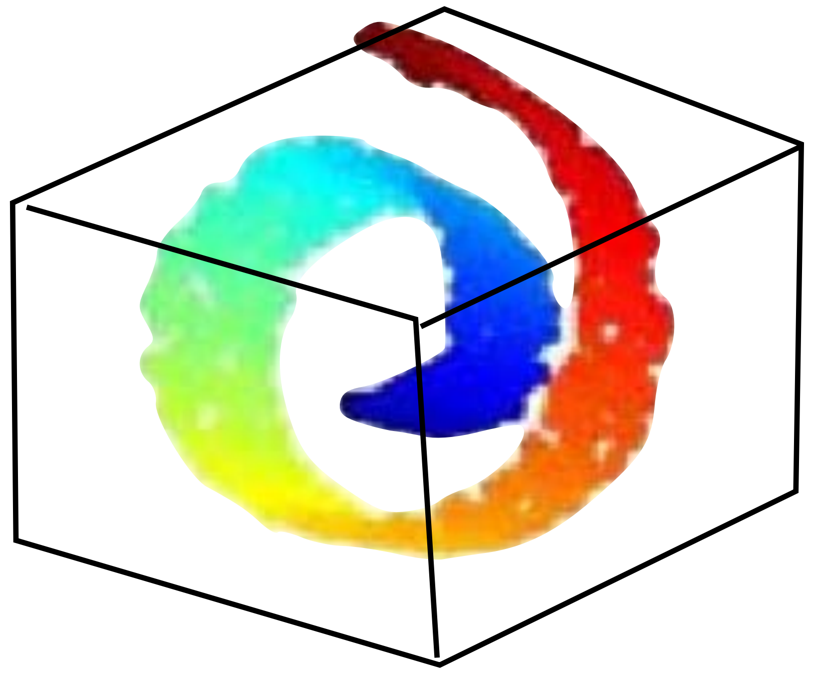
Trends in
statistical mathematics

- (1) Era of big data
- (2) Dimensionality reduction
- (3) Sparsity
- (4) Bayesian analysis
- (5) Machine learning
- (6) Neural network
- (7) Data challenge

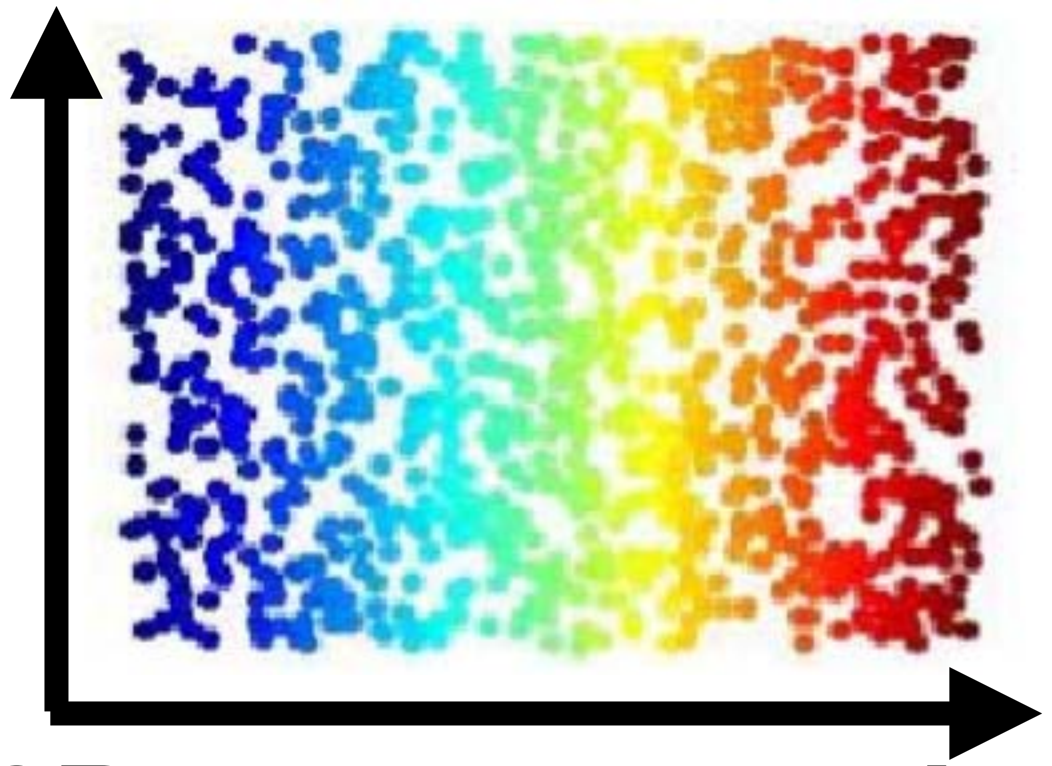
Dimensionality reduction

“特徴量”

N-dimensional “data” → low-dimensional “feature”

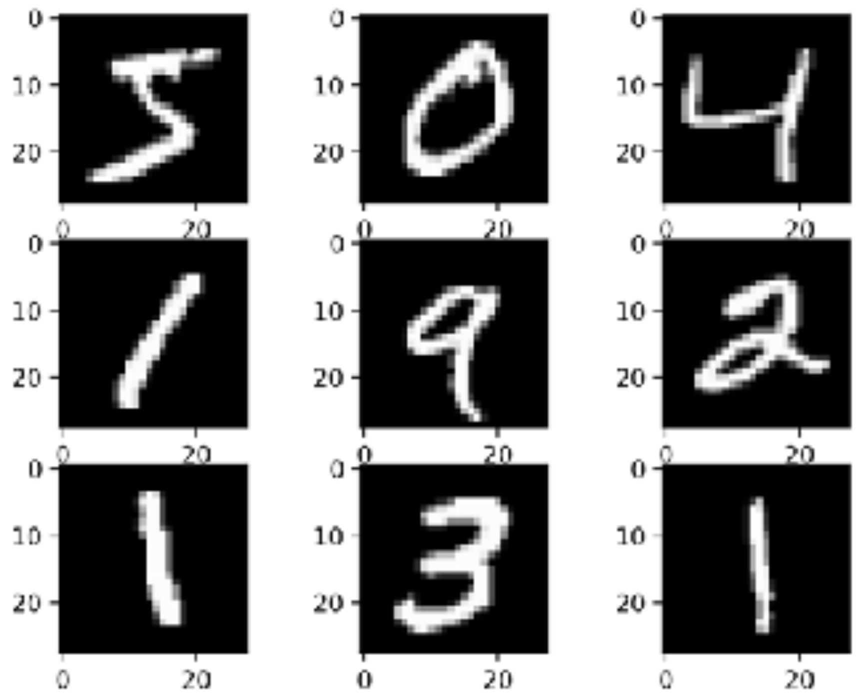


2D manifold in **3D-space**



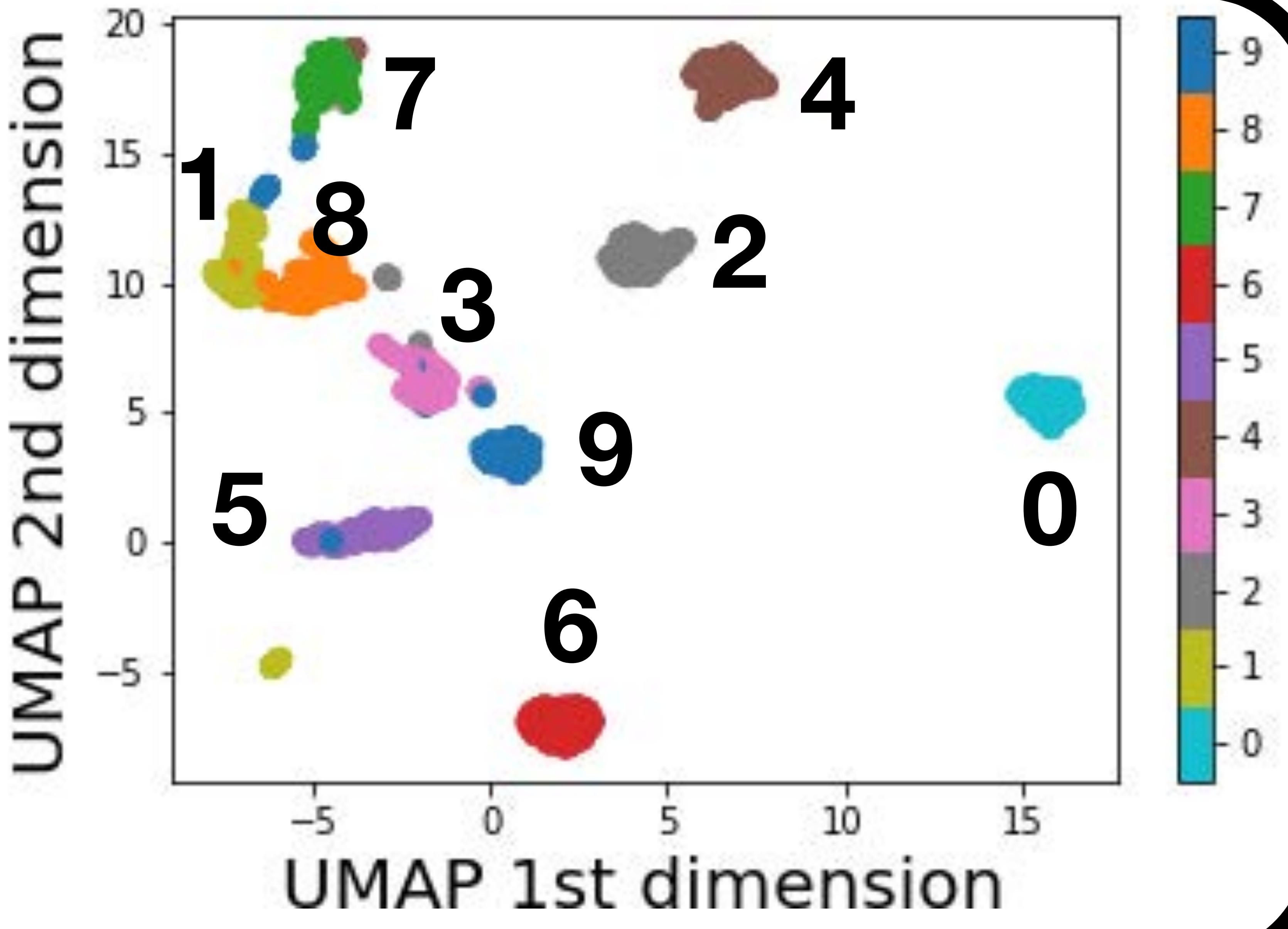
2D representation

Simpler description
“Manifold learning”



UMAP

28x28 = 784 pixels
… 784-dim data



Techniques

- **Self-Organizing Map** (SOM)
- **Isomap**
- **UMAP**

Trends in statistical mathematics

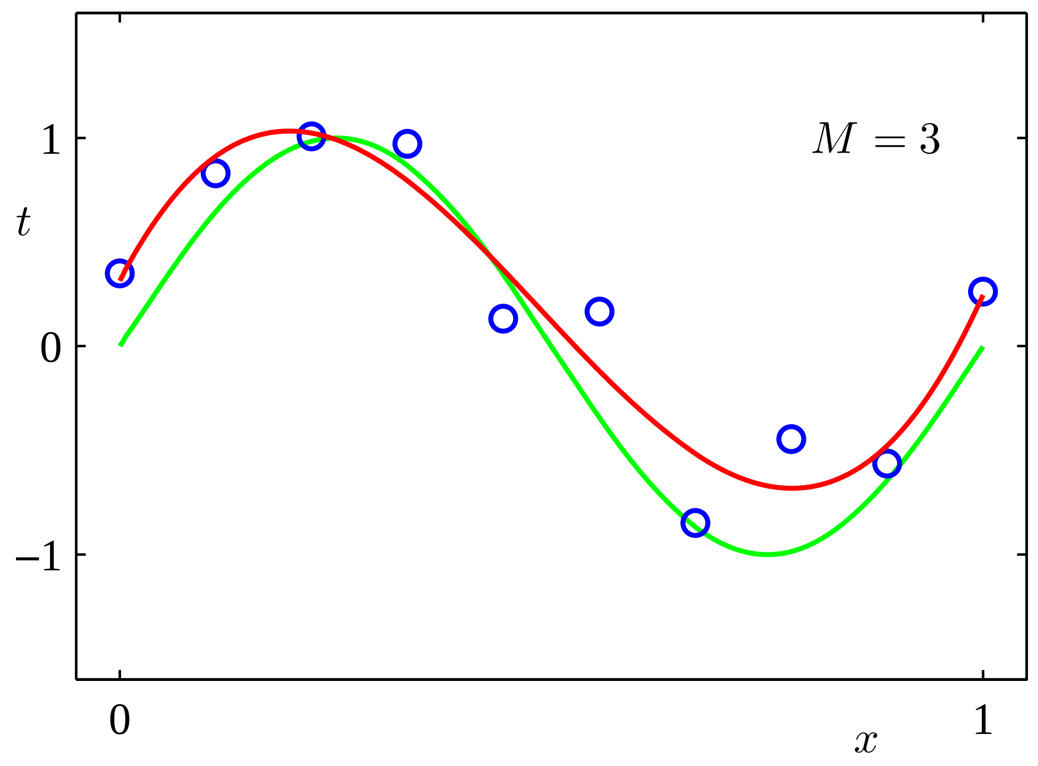
- (1) Era of big data
- (2) Dimensionality reduction
- **(3) Sparsity**
- (4) Bayesian analysis
- (5) Machine learning
- (6) Neural network
- (7) Data challenge

Sparse Modeling

Easy case:

(# of model params) $\leq$ (# of data points)

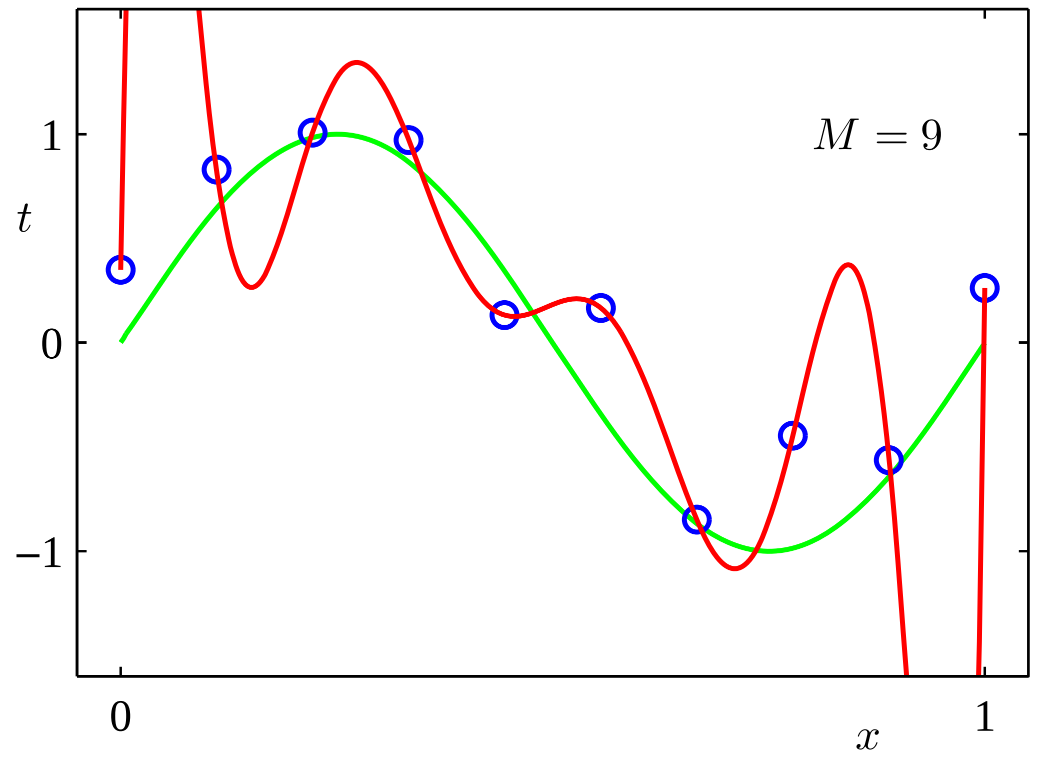
$$N \leq M$$



What if …

(# of model params) $>$ (# of data points)

$$N > M$$



In general, it fails (degenerate params).
But there is a hope, if the model is sparse.

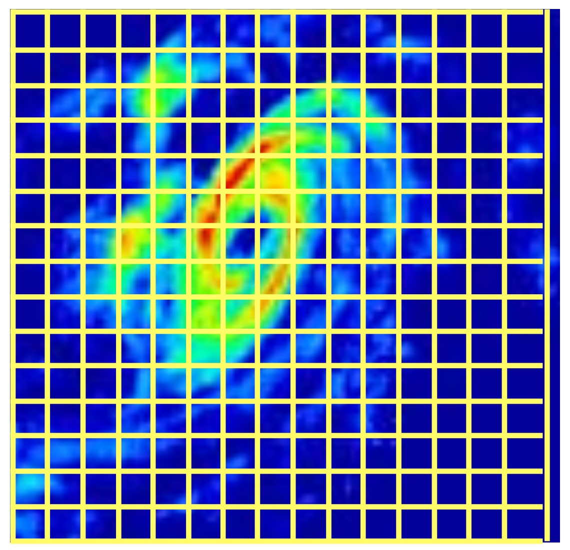
Tibshirani (1996)

Sparse Modeling for image data

Visibility

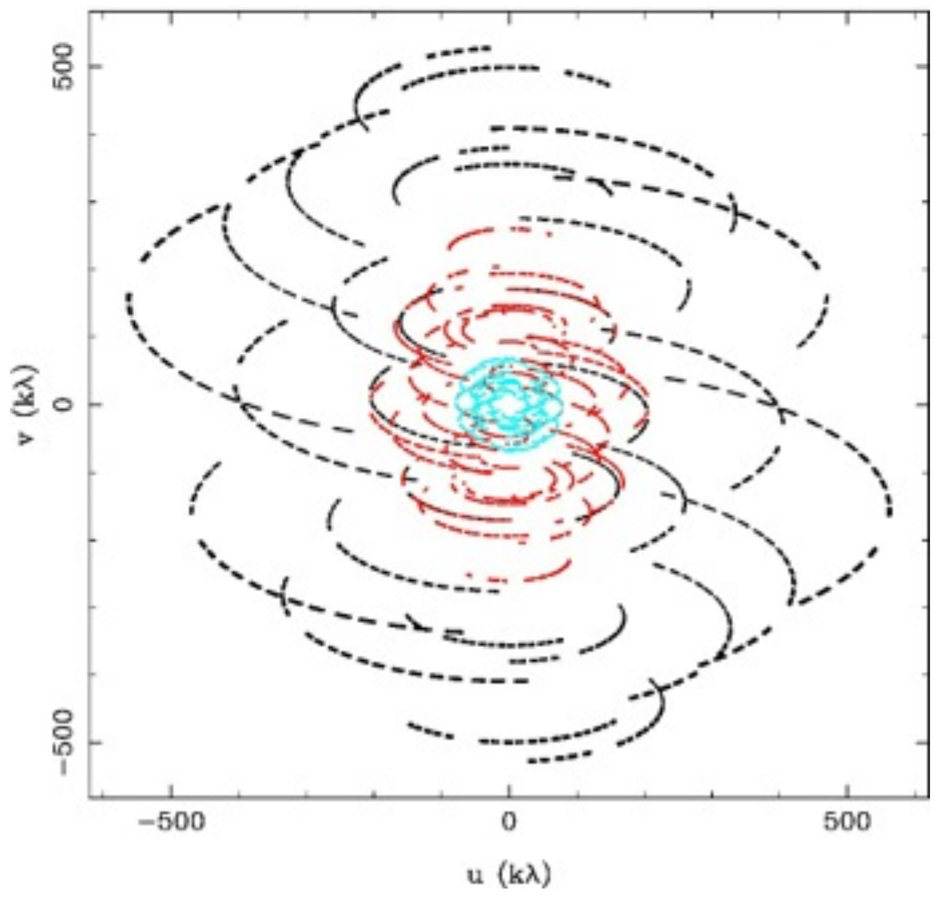
Fourier transformation of a radio image

$$v(U, V) = \mathcal{F}\left[\quad \right]$$



Model brightness (β) has N pixels.

incomplete sampling



***M* data points for the observed v(U,V)**

$$\begin{bmatrix} v_1 \\ \vdots \\ v_M \end{bmatrix} = \mathcal{F} \begin{bmatrix} \beta_1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \beta_N \end{bmatrix}$$

Often, we have M < N.
No unique solution for β.

Sparse Modeling

$$\begin{bmatrix} v_1 \\ \vdots \\ v_M \end{bmatrix} = \mathcal{F} \begin{bmatrix} \beta_1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \beta_N \end{bmatrix}$$

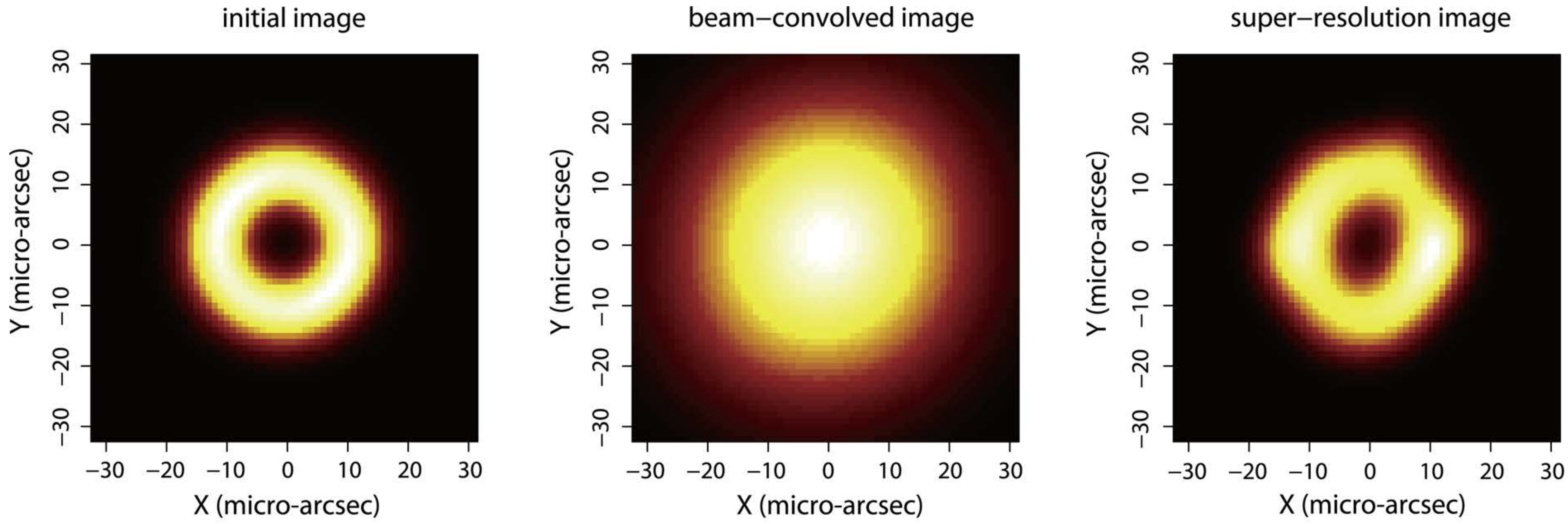
If we assume that the image is sparse (many zeros in beta), we can obtain a LASSO solution:

$$\hat{\beta} = \arg\min \left[\underbrace{\frac{1}{2} \|\mathbf{v} - F\beta\|_2^2}_{\text{residual}} + \underbrace{\lambda \sum_i |\beta_i|}_{\text{penalty}} \right]$$

By adopting a large lambda, we obtain a sparse solution.
(Adequate lambda depends on the science case. Need cross-validation.)

Sparse Modeling

Ground truth **LASSO**



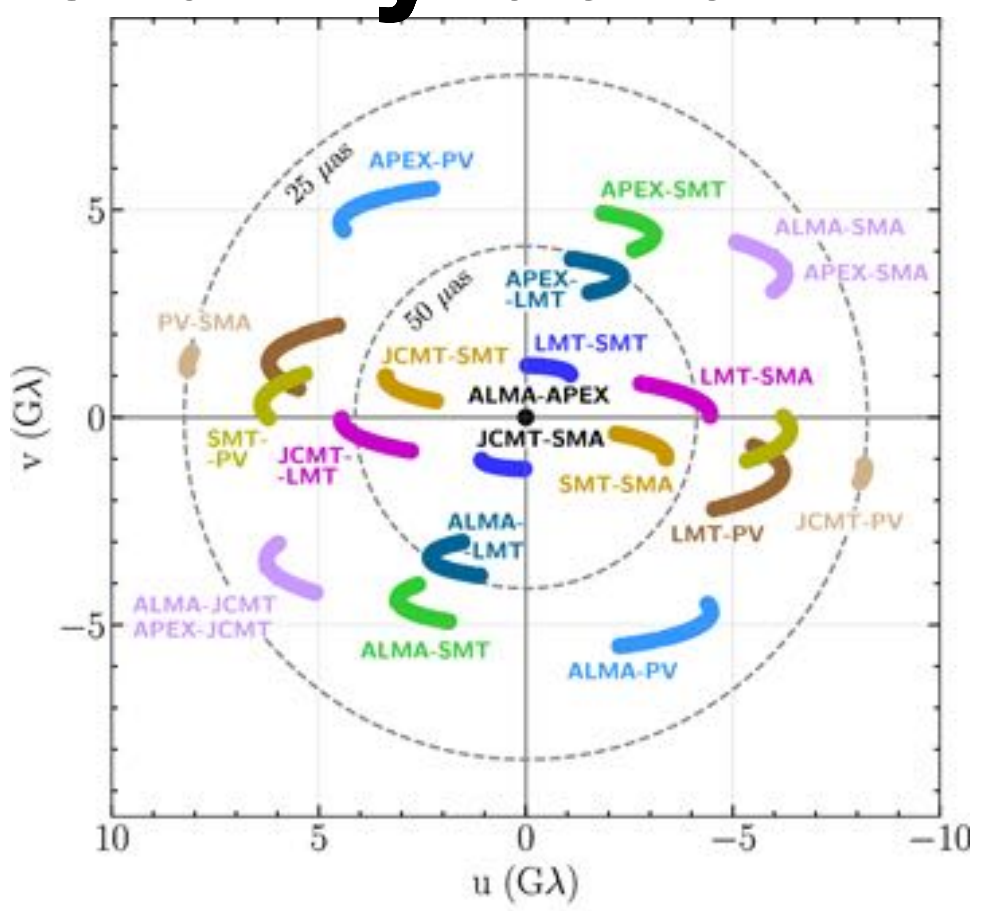
Honma, Akiyama, Uemura, Ikeda (2014)

Sparse Modeling

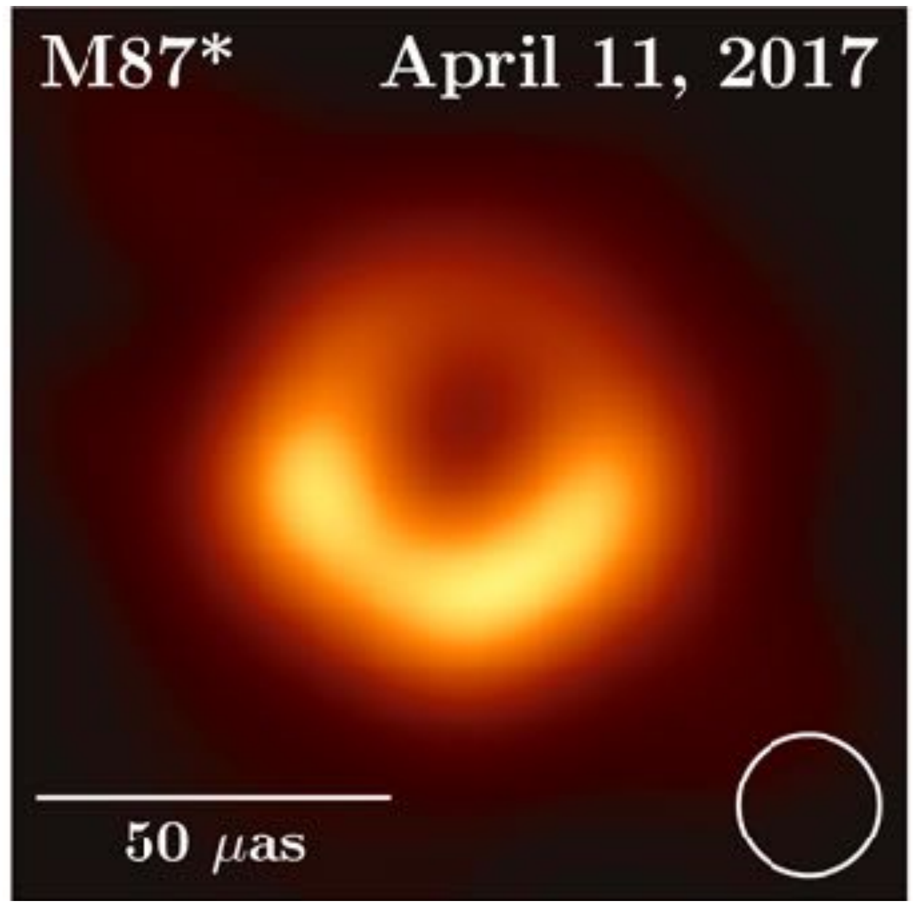
To believe or not to believe; *blind test* is important.

M87's black hole shadow

Visibility data



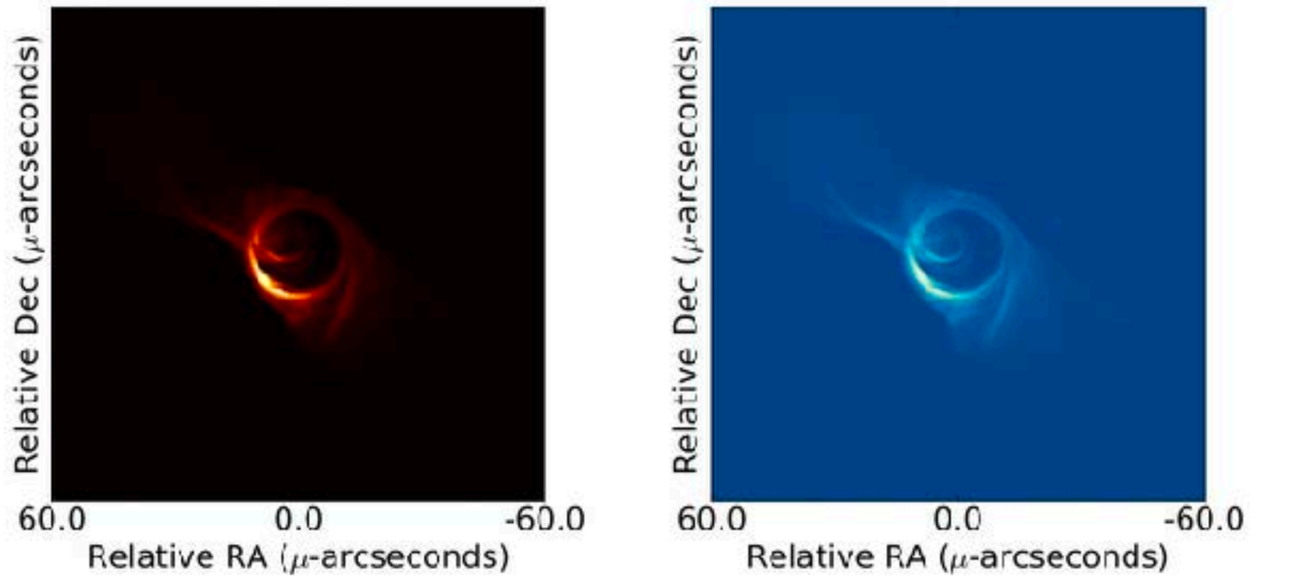
Reconstructed image



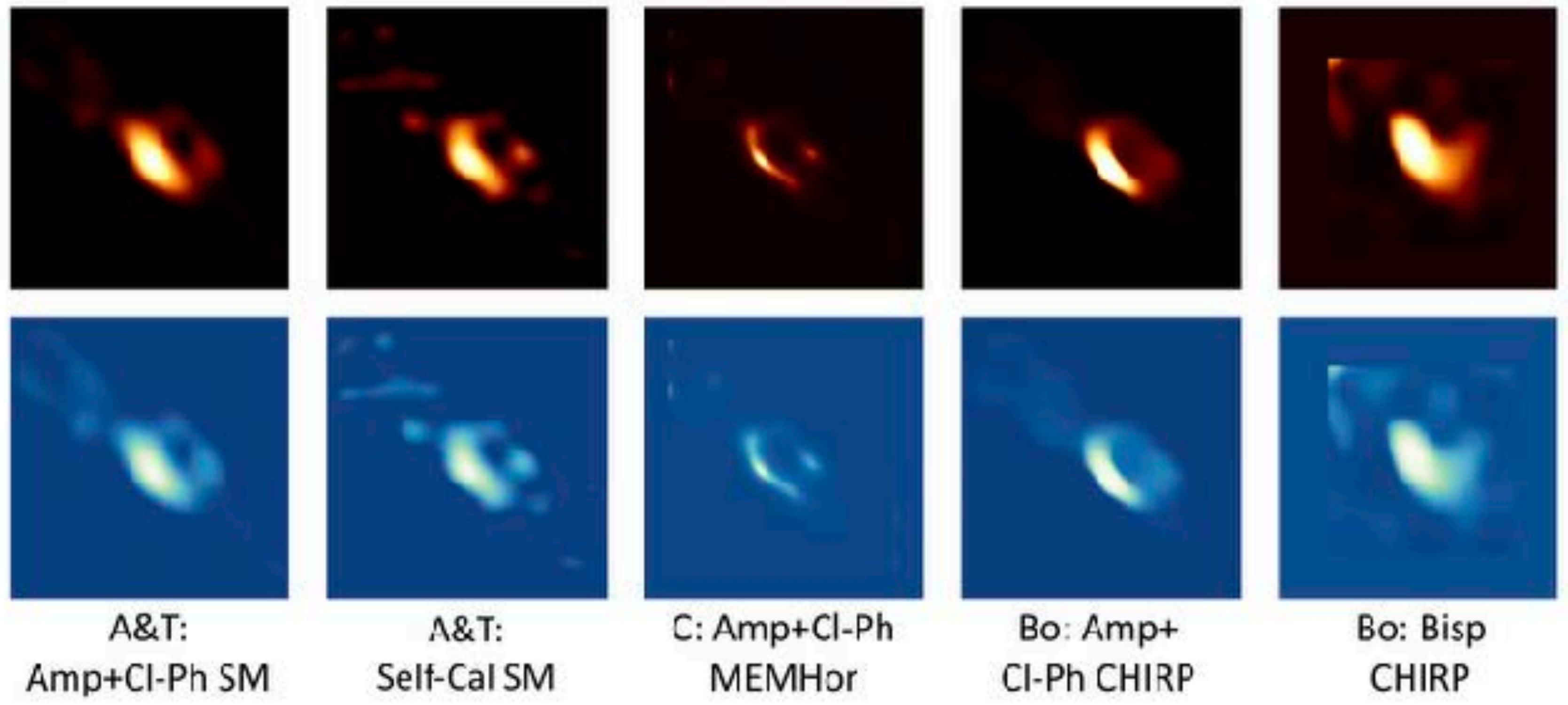
EHT Collaboration, Akiyama et al. (2019)

Data challenge (Mock analysis)

Ground truth



Reconstructed image



Katherine L. Bouman PhD thesis (2017)

Radio telescopes opened a new window
into SMBHs and possibly IMBHs (ALMA, SKA, ngVLA).

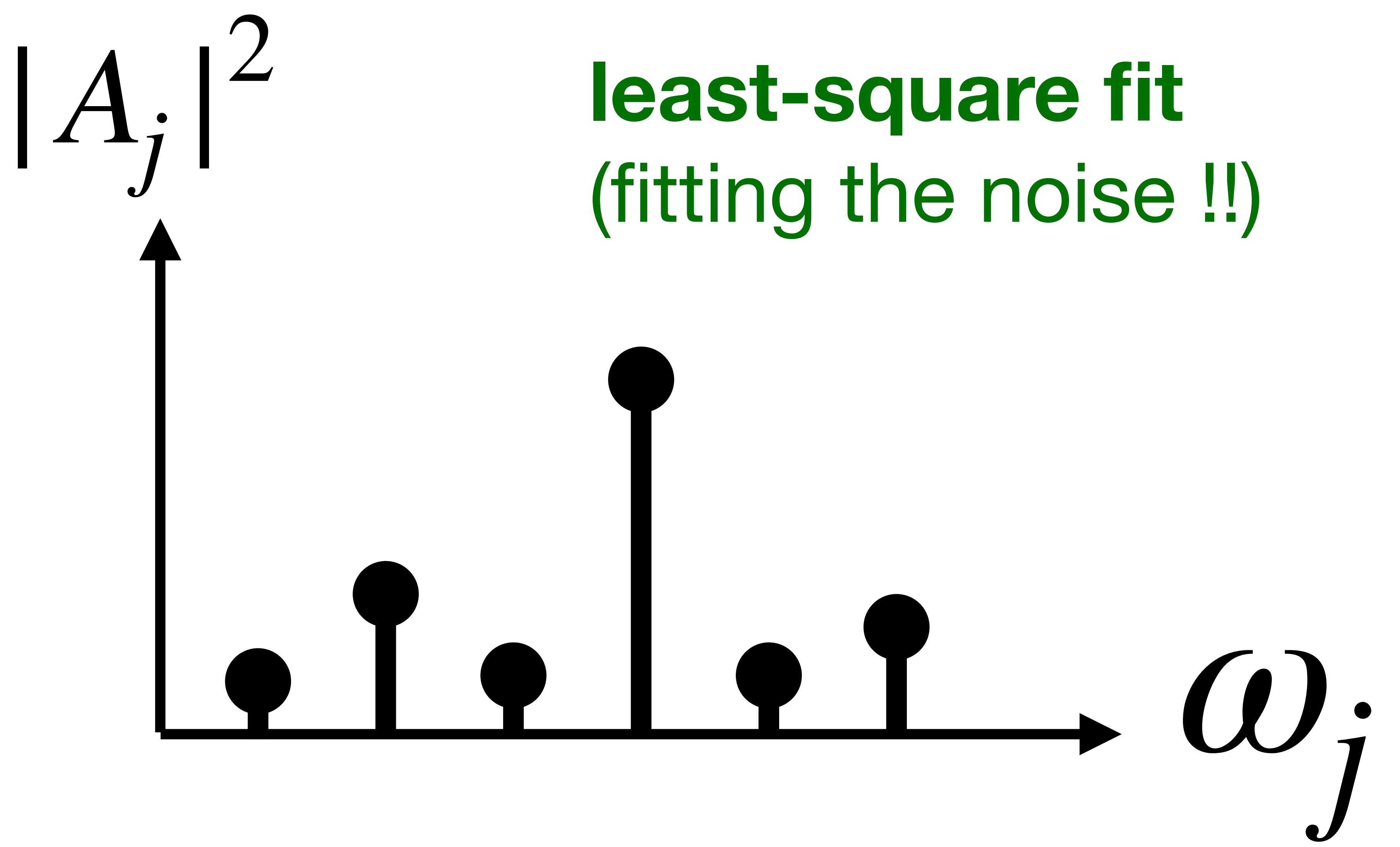
Sparse Modeling for time-series data

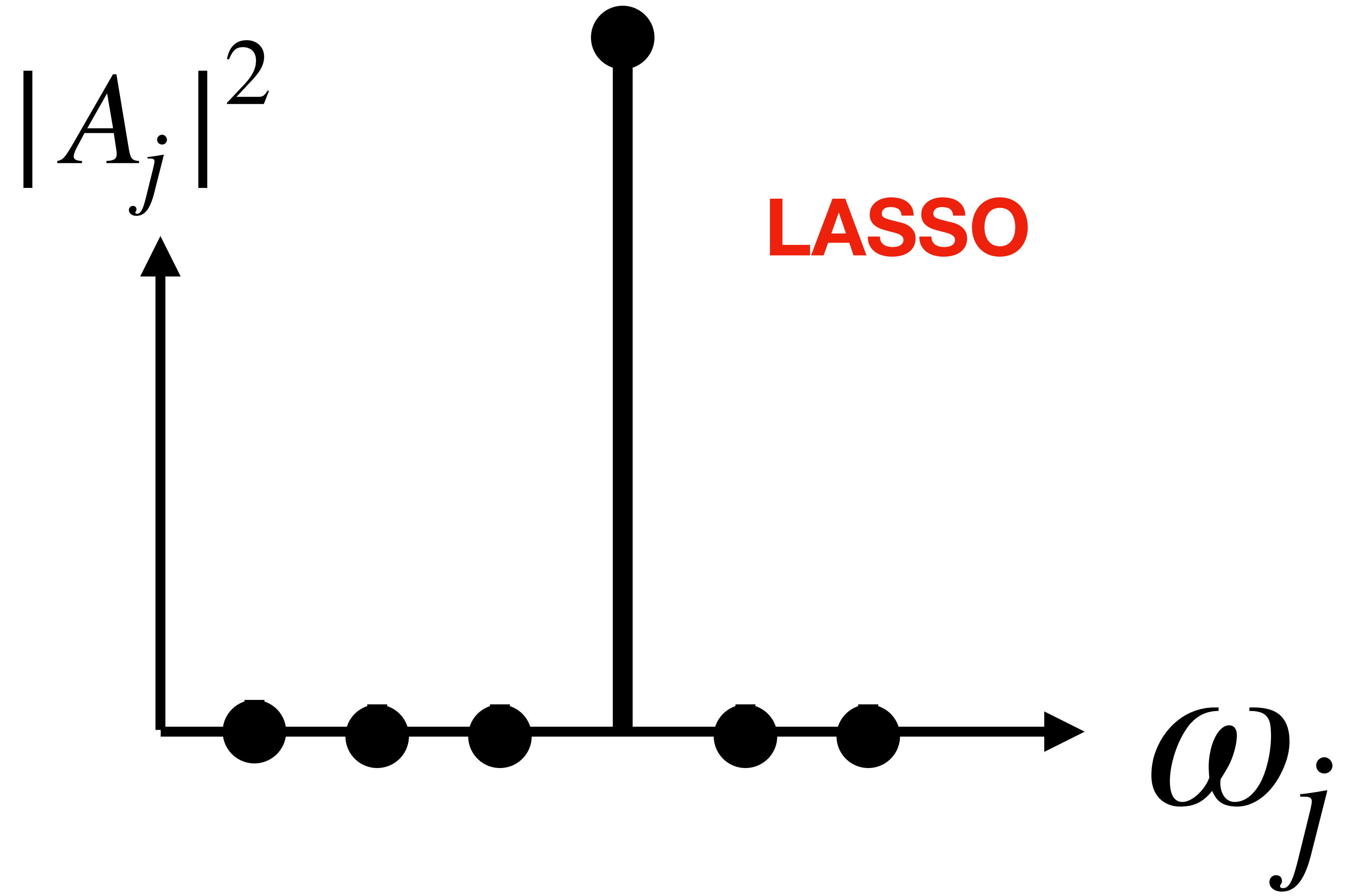
$$\textbf{flux: } y(t) = \sum_j A_j \exp(i\omega_j t)$$

Kato & Uemura (2012)*

Regular variable stars (e.g., Cepheids, RR Lyrae)
have only few frequencies with non-zero amplitude.

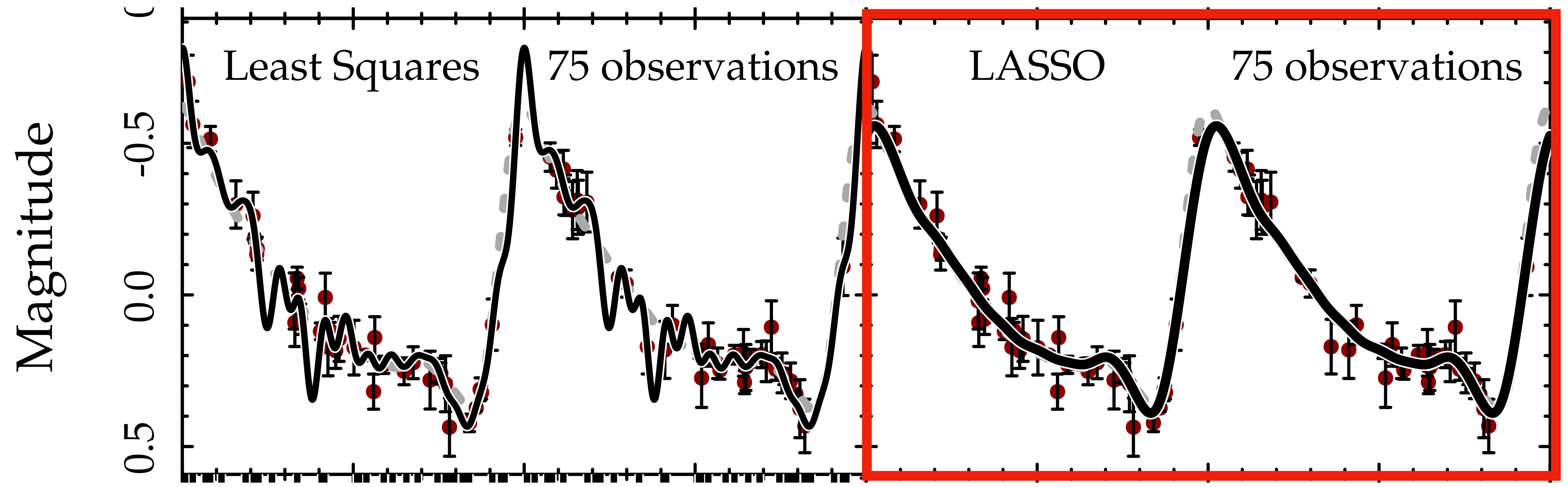
LASSO in power-spectrum space.
>> Simplest representation of the light curve.





Bellinger, Wysocki, Kanbur (2016)

Successful LASSO for RR Lyrae

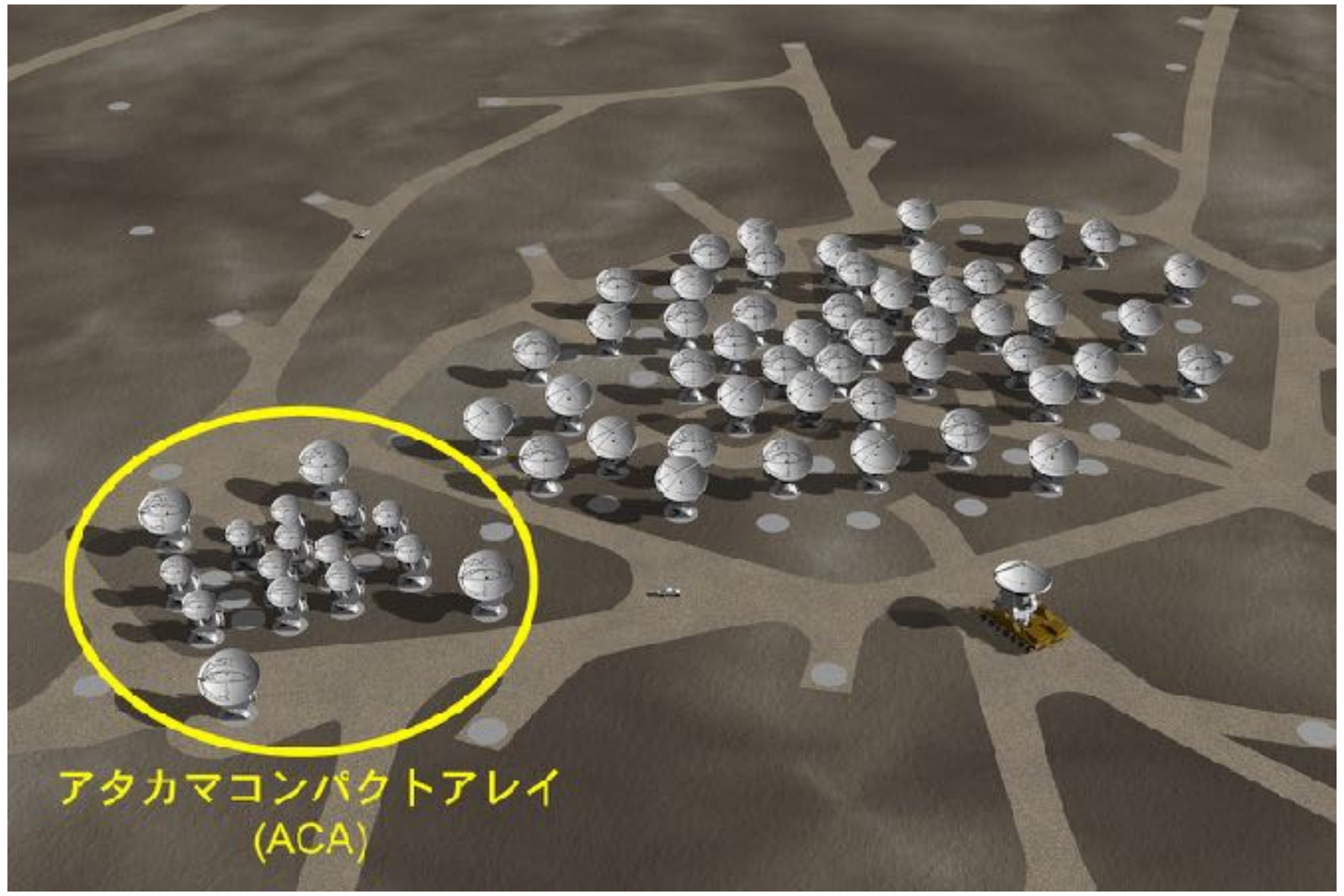


* Data taken from VSNET.

Sparse Modeling

森田耕一郎 教授

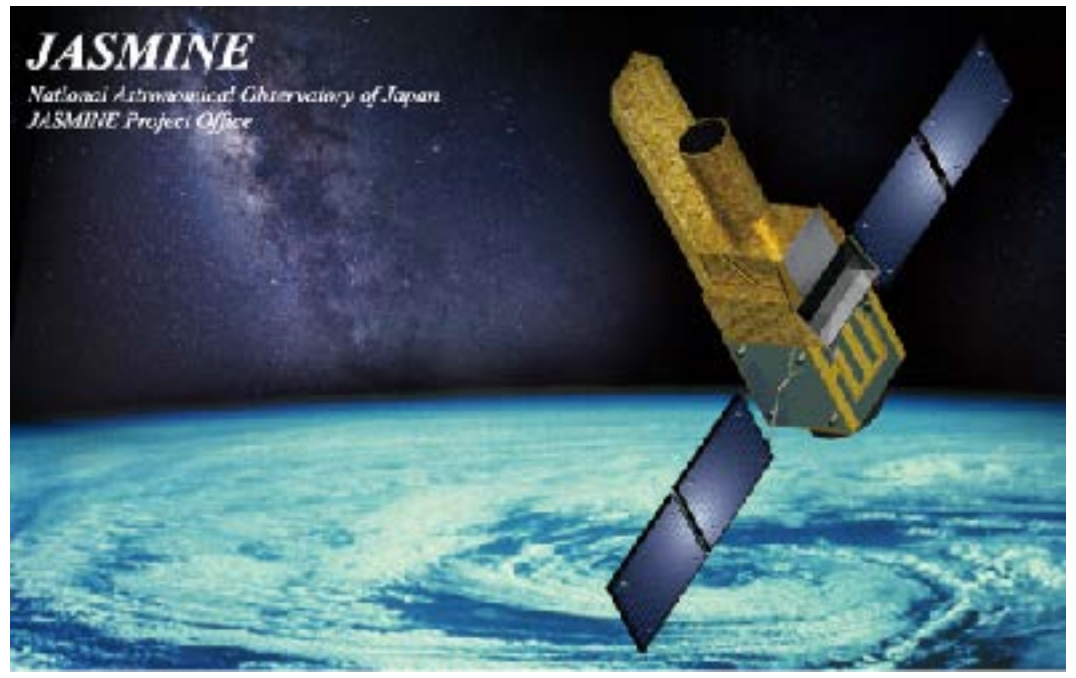
(1954 - 2012)



One of the first persons in astronomy who envisioned the importance of sparse modeling in radio interferometry.

Sparse matrix … Tough example

JASMINE will ***stare at*** the MW center region.
JASMINE will
- observe 1e5 stars
- take photo for 80 x 2000 times for each star (***Paparazzi !***)
- measure N=1e7 params (including satellite attitude)



If we are to make the most of JASMINE,
we need to solve a huge matrix inversion problem

$10^{10} \times 10^{7}$ **matrix**

$$10^{10} \quad \begin{pmatrix} o_1 \\ \vdots \\ o_\ell \\ \vdots \\ o_L \end{pmatrix} \simeq \begin{pmatrix} \hat{o}_1 \\ \vdots \\ \hat{o}_\ell \\ \vdots \\ \hat{o}_L \end{pmatrix} + \underbrace{\left.\begin{pmatrix} \frac{\partial}{\partial p_1} f_1^{\mathcal{P}} & \cdots & \frac{\partial}{\partial p_\lambda} f_1^{\mathcal{P}} & \cdots & \frac{\partial}{\partial p_\Lambda} f_1^{\mathcal{P}} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial p_\ell} f_\ell^{\mathcal{P}} & \cdots & \frac{\partial}{\partial p_\lambda} f_\ell^{\mathcal{P}} & \cdots & \frac{\partial}{\partial p_\Lambda} f_\ell^{\mathcal{P}} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial p_L} f_L^{\mathcal{P}} & \cdots & \frac{\partial}{\partial p_\lambda} f_L^{\mathcal{P}} & \cdots & \frac{\partial}{\partial p_\Lambda} f_L^{\mathcal{P}} \end{pmatrix}\right|_{\hat{\boldsymbol{p}}}}_{D^{\mathcal{P}}} \begin{pmatrix} p_1 - \hat{p}_1 \\ \vdots \\ p_\lambda - \hat{p}_\lambda \\ \vdots \\ p_\Lambda - \hat{p}_\Lambda \end{pmatrix} \quad 10^{7}$$

$$\boldsymbol{o} \simeq \hat{\boldsymbol{o}} + D^{\mathcal{P}} \Delta\hat{\boldsymbol{p}}$$

Although the design matrix is sparse, it will be tough to solve it quickly.
>> Need a Parallel Computing LSQR (Least Squares with QR-factorization**) method?**

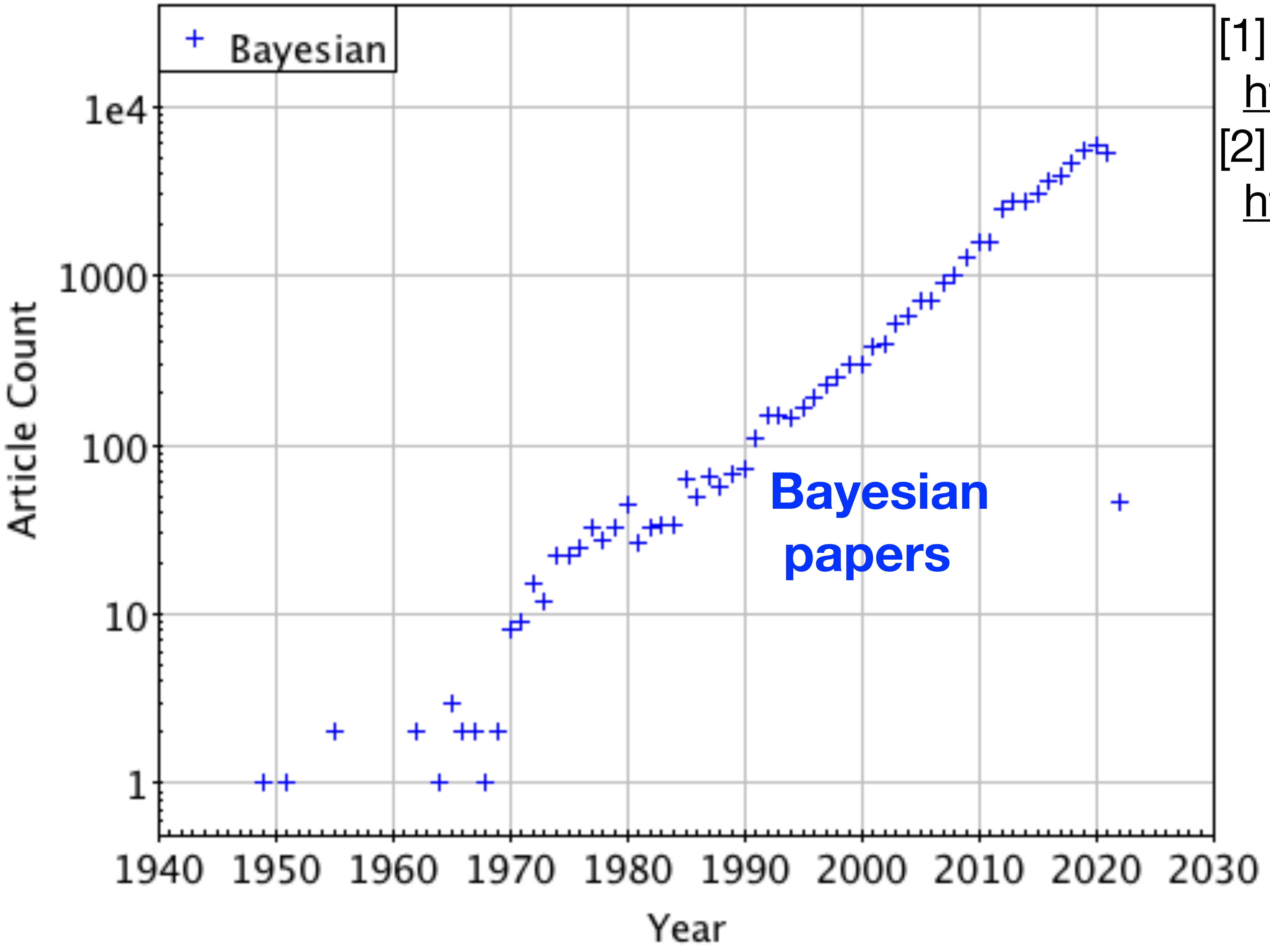
Trends in
statistical mathematics

- (1) Era of big data
- (2) Dimensionality reduction
- (3) Sparsity
- (4) Bayesian analysis
- (5) Machine learning
- (6) Neural network
- (7) Data challenge

Trends in ADS papers

ADS papers with "Bayesian" in the abstract.

Exponential growth. Time scale ~7 years



Unfamiliar with Bayes? See, e.g.,

[1] Data analysis recipes: **Fitting a model to data**
https://arxiv.org/abs/1008.4686

[2] Data analysis recipes: Probability calculus for inference
https://arxiv.org/abs/1205.4446

Bayesian analysis

Only 3 topics (to save time)**:**

(1) MCMC in high-dimensional space is tricky,
especially if the posterior distribution is multi-modal.
>> Try nested sampling (Skilling 2004). [e.g., Hikage et al. 2019]

(2) Bottleneck in MCMC is the computational cost in the likelihood.
>> Simplify the likelihood function (e.g., interpolation) [e.g., Nishimichi et al. 2019]
>> Reduce the *effective* data size [e.g., Hattori et al. 2021]

(3) "Likelihood" is sometimes hard to define (e.g., likelihood of **N-body** model?)
>> Try ABC "Approximate Bayesian Computation"
= MCMC-like analysis for the summary statistics

- (1) Era of big data
- (2) Dimensionality reduction
- (3) Sparsity
- (4) Bayesian analysis
- (5) Machine learning
- (6) Neural network
- (7) Data challenge

Trends in machine learning

Machine learning

Machine learning is the science of getting computers to act without being explicitly programmed.

— Andrew Ng

A set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty.

— Kevin P. Murphy

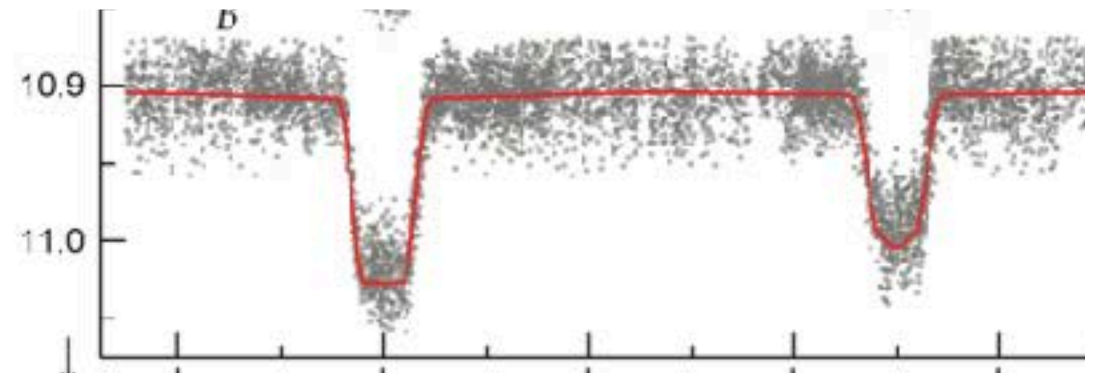
Machine learning is the study of computer algorithms that can improve automatically through experience and by the use of data.

— Wikipedia

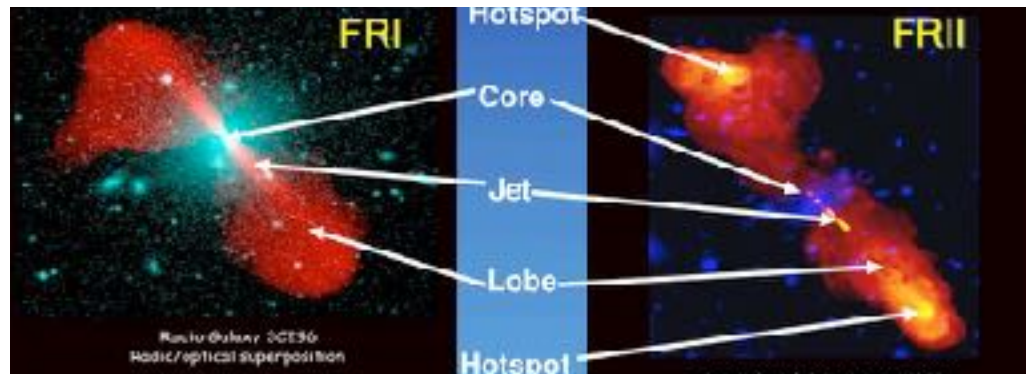
Machine learning

Supervised learning: Based on the "input-output" pairs (test data), find the function that maps input to output.

Classification



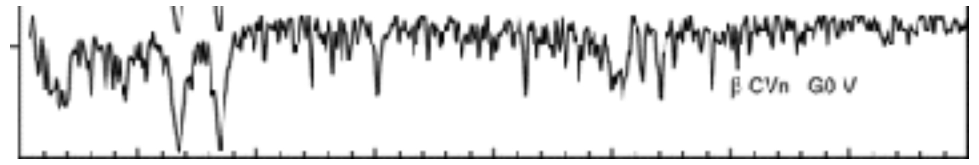
Transients / Variables / Flares / Micro-lensing / SNe



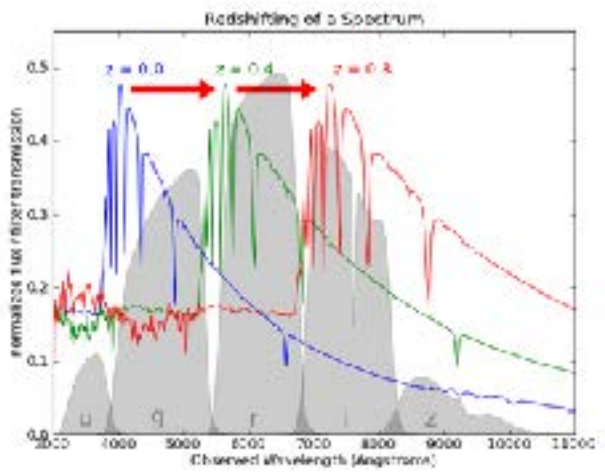
Radio galaxy types

PLASTiCC challenge

Parameter inference



Spectral analysis (Subaru PFS / GA)

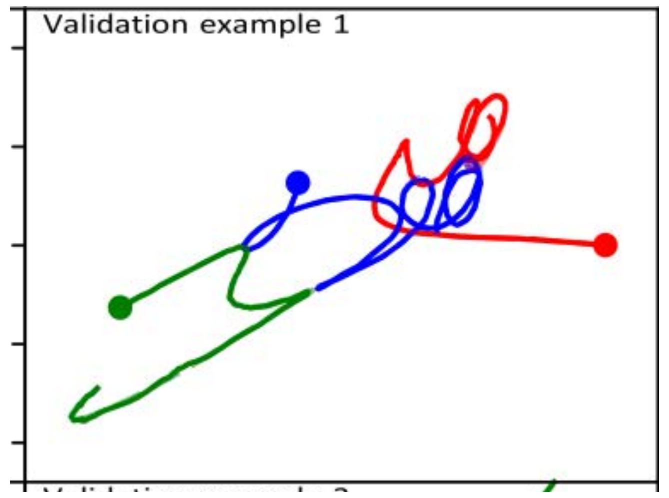


Photometric redshift

Identification

Find strong lensing images

Model generation



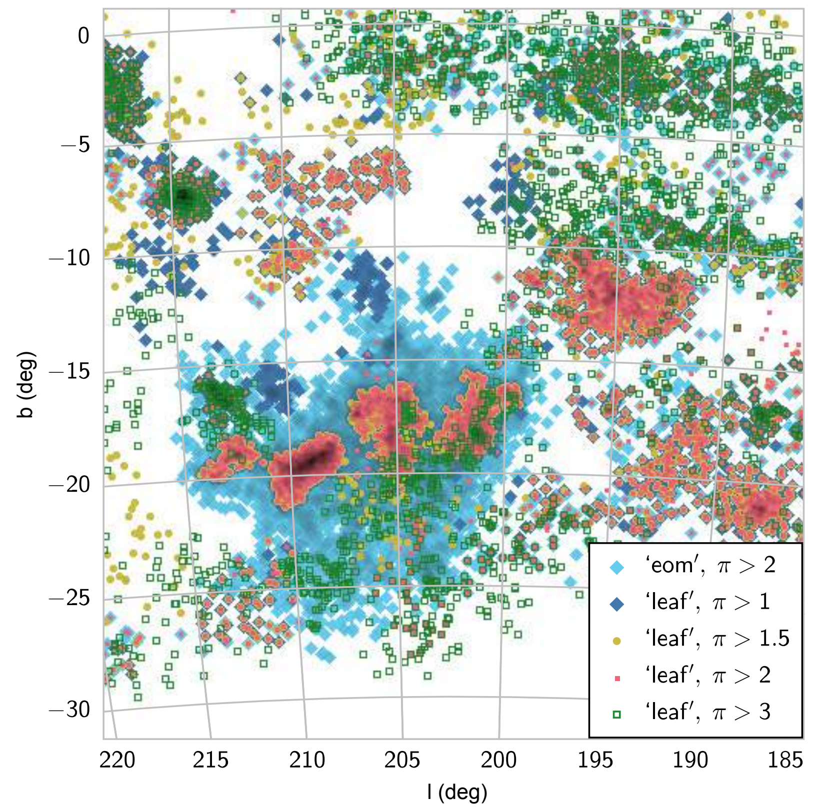
Chaotic 3-body model

... etc.

Machine learning

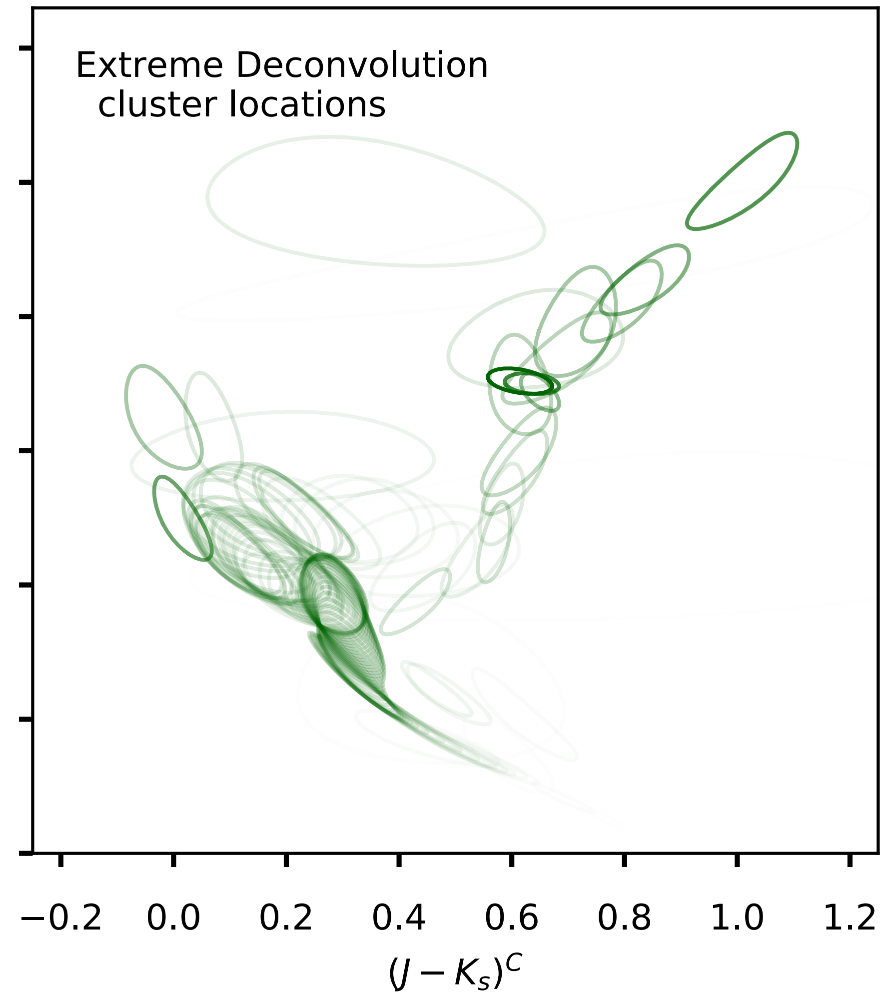
Unsupervised learning: Find the pattern in the data without external information.

Clustering



Find open clusters /
Over-density of galaxies

Density estimation



Deriving precise CMD /
Deriving 3D dust map

Anomaly detection

Pair-instability SNe /
Hypervelocity stars /
Gravitational lensing

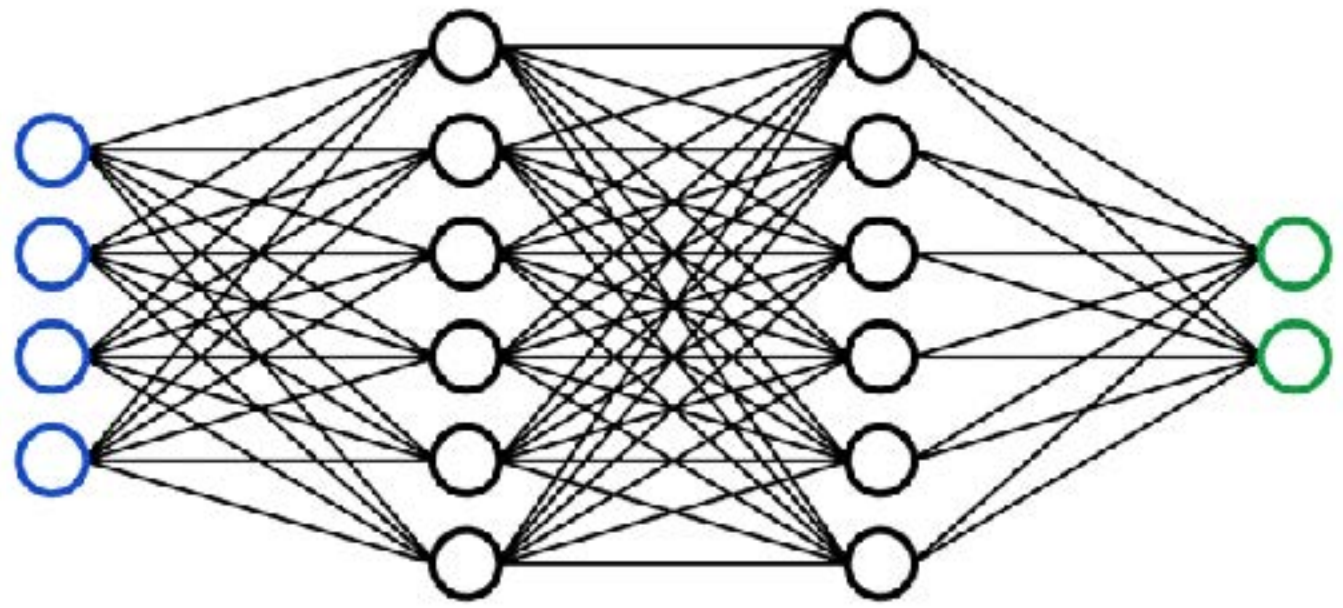
... etc.

Machine learning

Classical methods

K-means
Support Vector Machine (SVM)
Random Forest
Gaussian Mixture Model (GMM)

Neural network

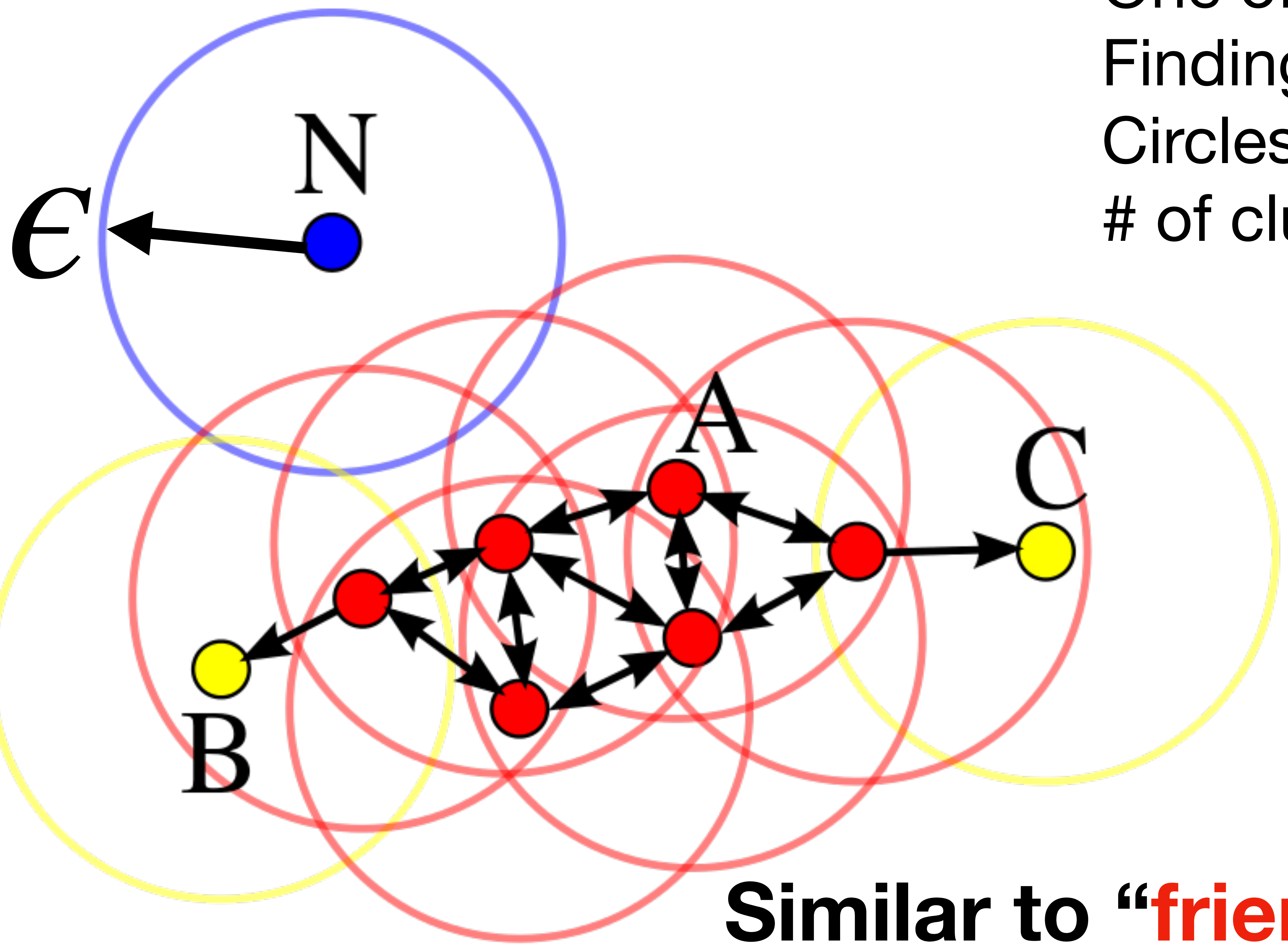


Clustering

Figure credit: https://en.wikipedia.org/wiki/DBSCAN

One of the fast / successful / easy-to-implement methods:

DBSCAN (Density-based spatial clustering of applications with noise)



One of the most widely used clustering method.
Finding data points within a given radius ϵ .
Circles which encloses less than N_min are ignored.
# of clusters is automatically determined.

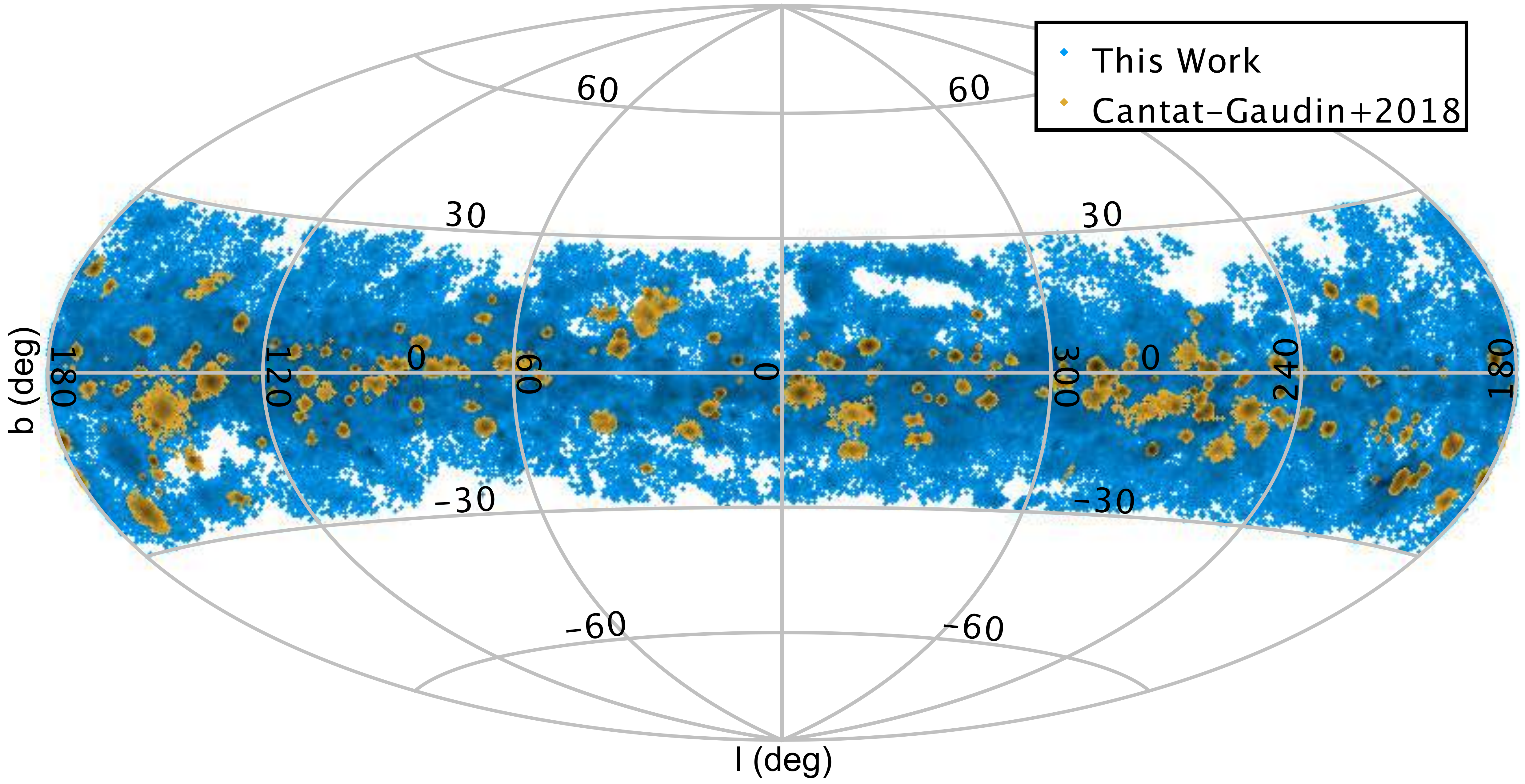
N_min = 4.
Only red circles form the cluster.

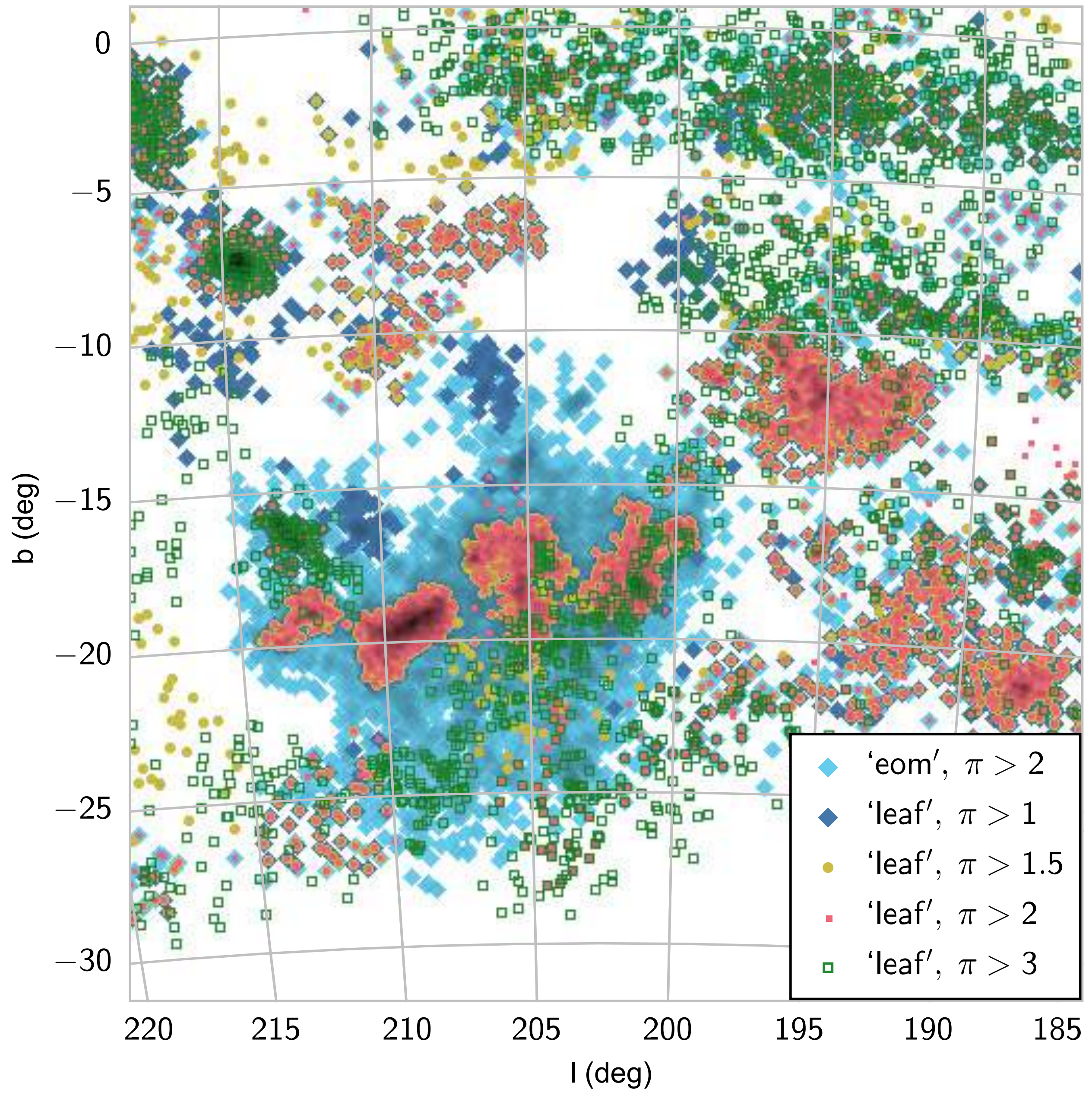
Similar to "friend-of-friend" method in N-body simulation.

Clustering

Discovering open clusters with Gaia data

Stars in an open cluster have similar position and velocity $(\mathbf{x}, \mathbf{v})$
>> DBSCAN can discover open clusters with Gaia's astrometric data.
>> There are ~2000 open clusters within 1 kpc from the Sun.

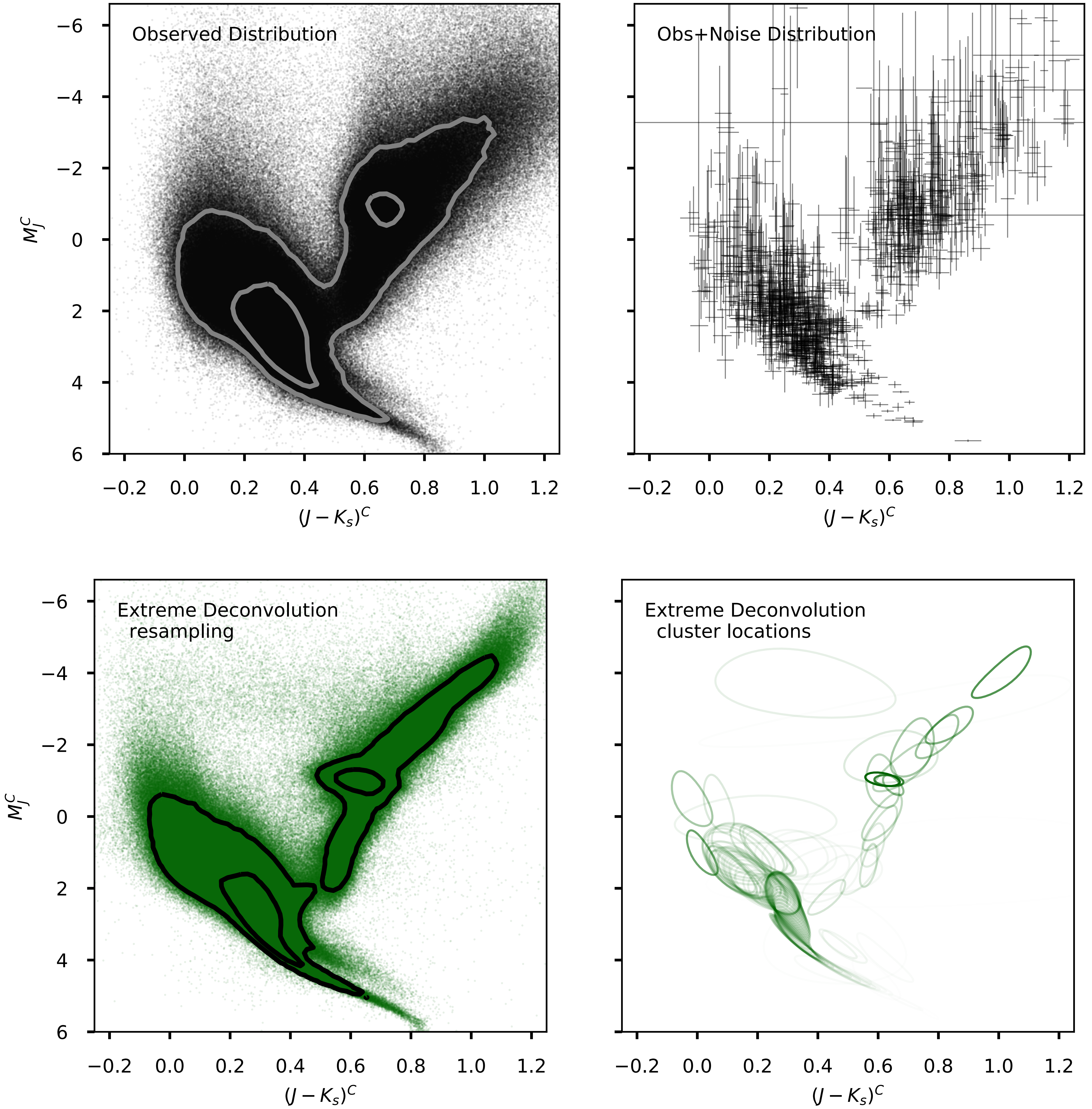




Kounkel & Covey (2019)

Density estimation

Color-magnitude diagram (CMD) of nearby stars in Gaia (DR1) data.



Raw data
… Blurred CMD due to distance error

Reconstructed CMD
… Gaussian Mixture model
after deconvolution of error

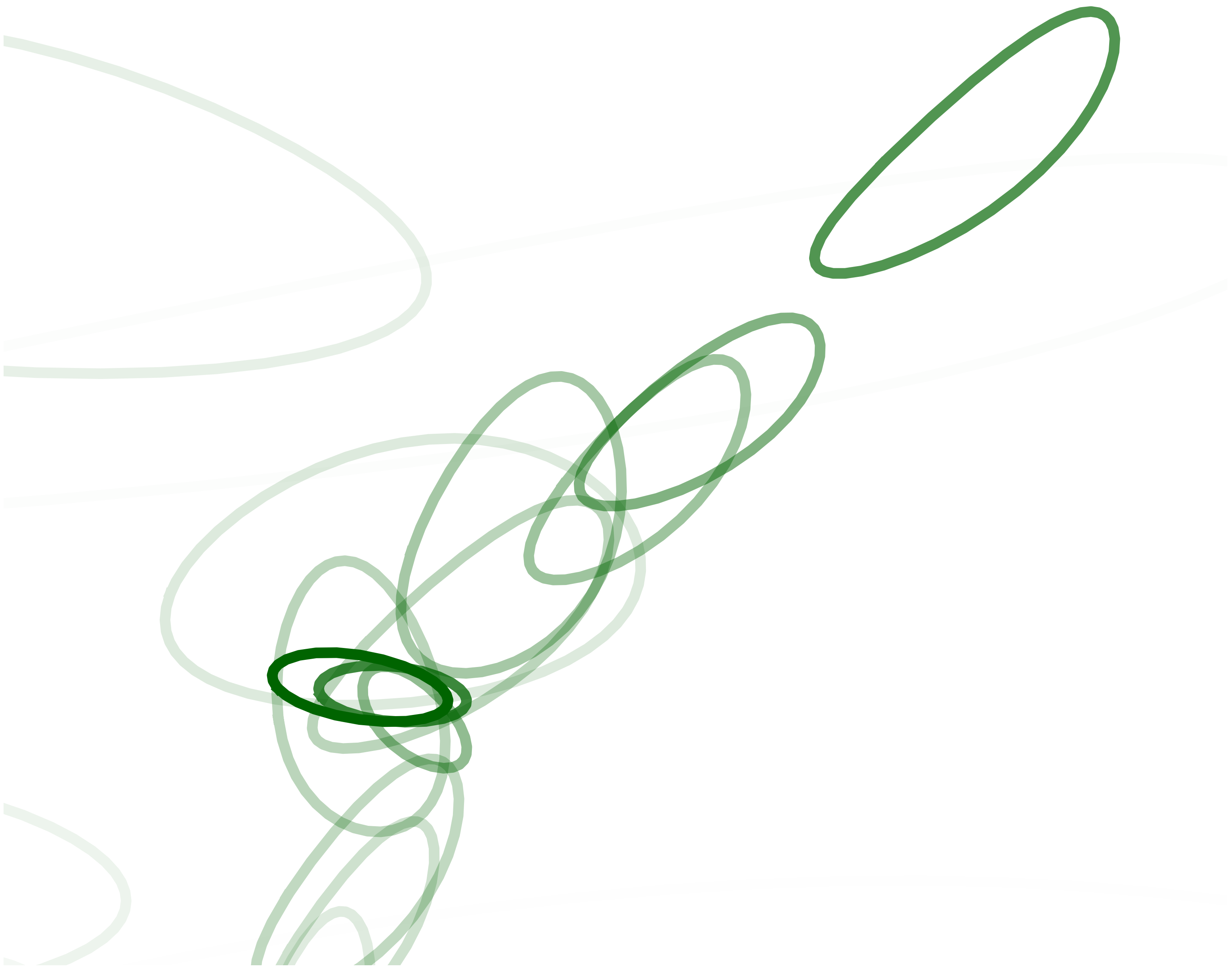
Anderson et al. (2017)
[see also Leistedt et al. 2017]

- (1) Era of big data
- (2) Dimensionality reduction
- (3) Sparsity
- (4) Bayesian analysis
- (5) Machine learning
- **(6) Neural network**
- (7) Data challenge

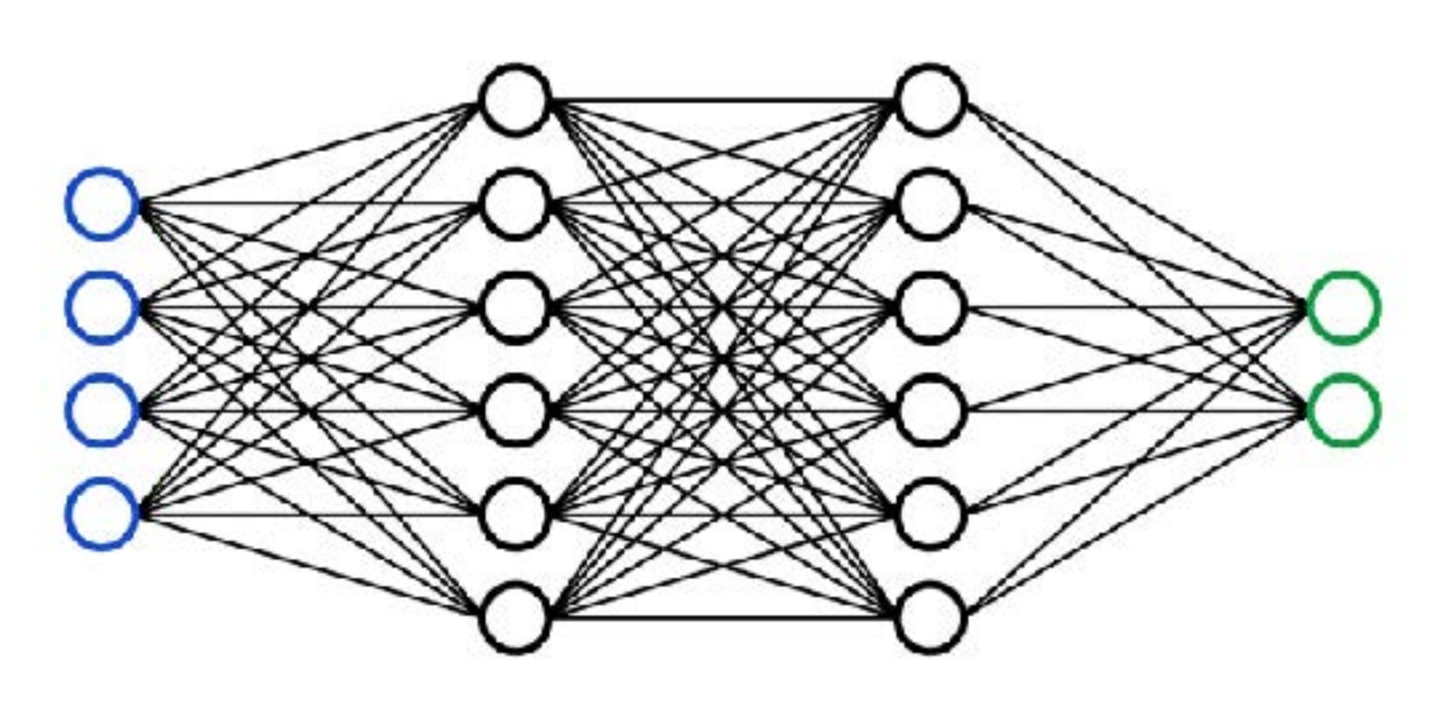
Trends in machine learning

Machine learning

Classical methods



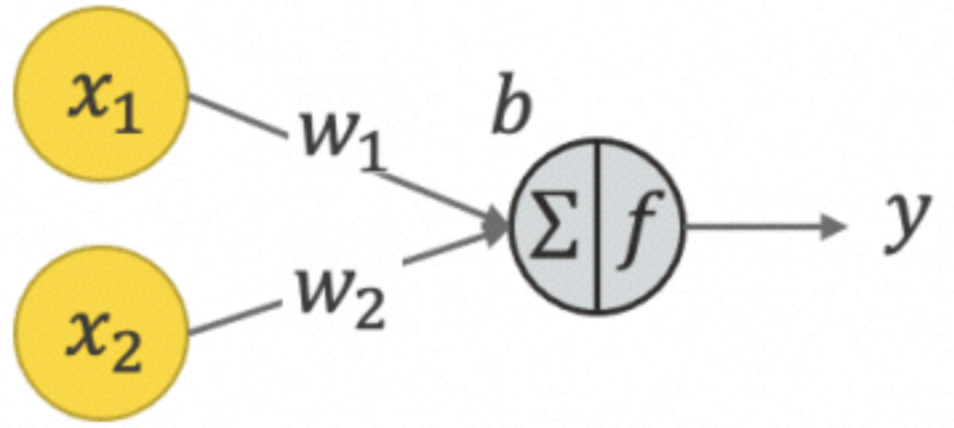
Neural network



Artificial Neural Network
Convolutional Neural Network
Auto-encoder
Generative Adversarial Network (GAN*)

*** Masato Shirasaki is an expert of GAN**

Neural network

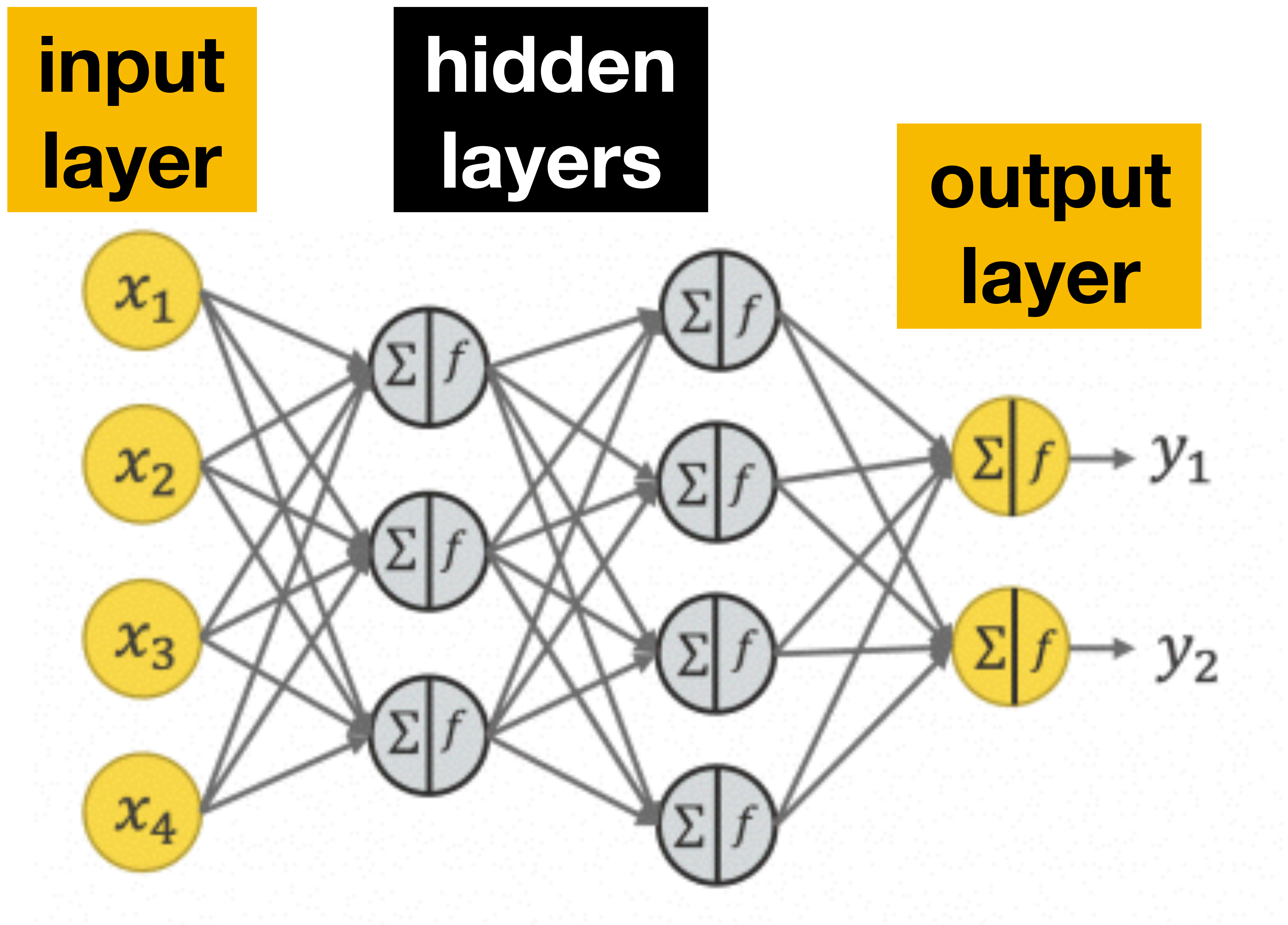


weighted sum of input nodes

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

f: activation function

$\mathbb{R}^N$



$\mathbb{R}^M$

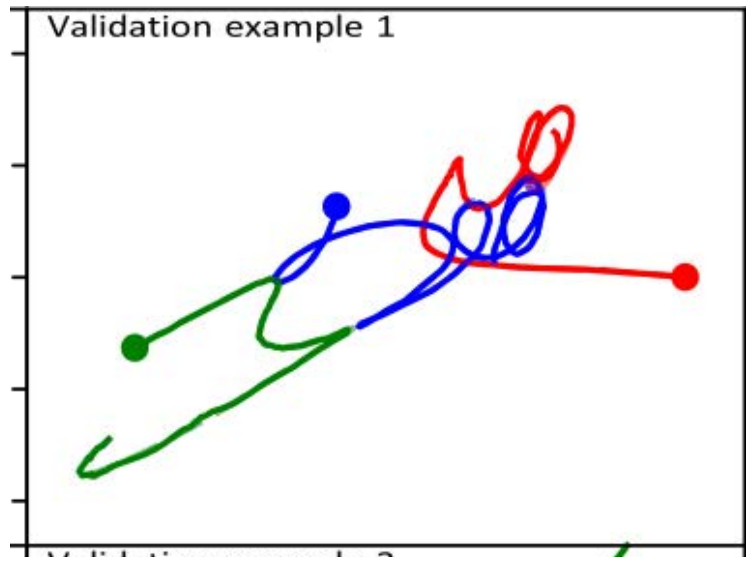
hidden layers = $g : \mathbb{R}^N \to \mathbb{R}^M$

Many parameters (weights) allow large flexibility.
We find the optimal weights by training.

If N>M, this can be seen as a dimensionality reduction.

Figure credit: https://www.knime.com/blog/a-friendly-introduction-to-deep-neural-networks

Model generation



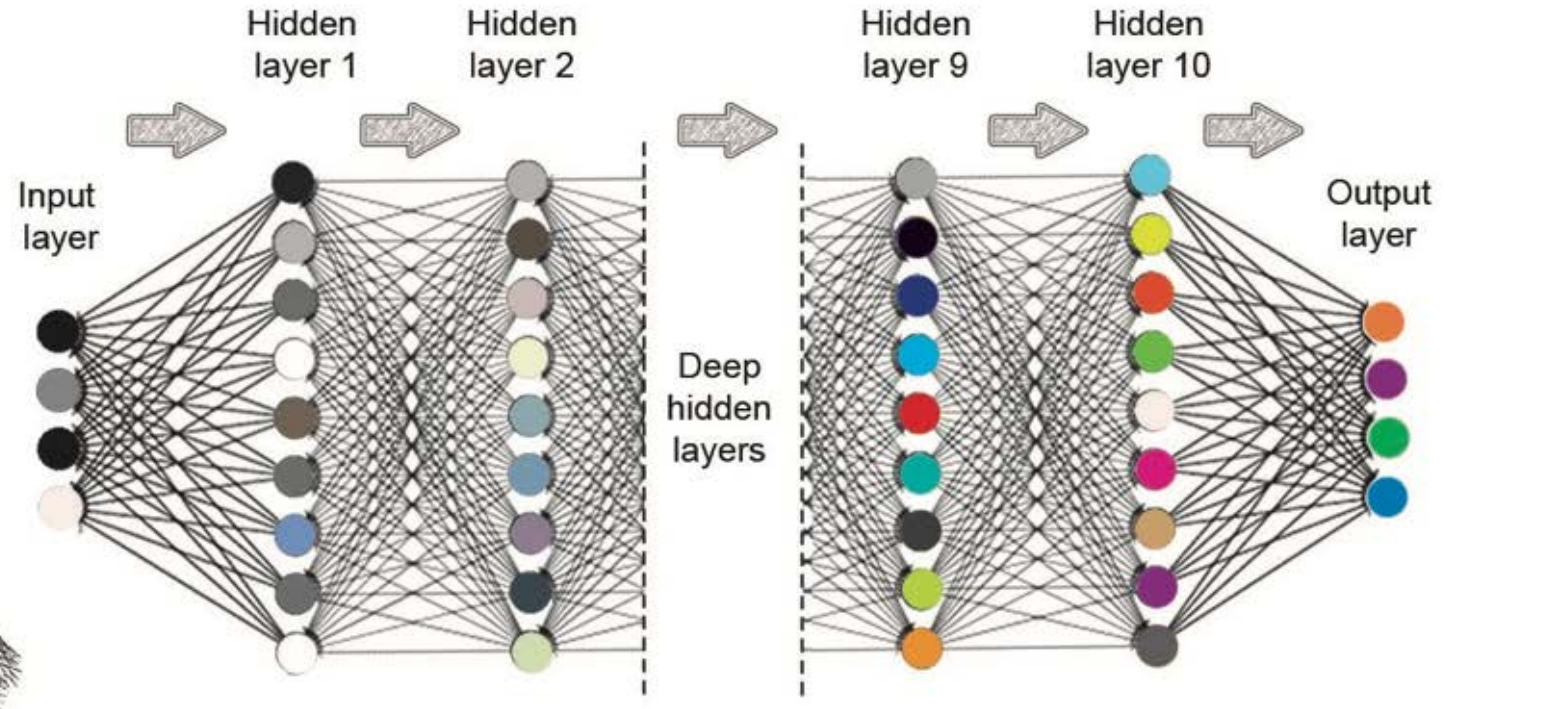
3-body problem in math.

$$\begin{bmatrix} x_1(T) \\ x_2(T) \\ x_3(T) \\ v_1(T) \\ v_2(T) \\ v_3(T) \end{bmatrix} = f \left(\begin{bmatrix} x_1(0) \\ x_2(0) \\ x_3(0) \\ v_1(0) \\ v_2(0) \\ v_3(0) \\ T \end{bmatrix} \right)$$

Operator of
"solving equation of motion"

Newton vs the machine: solving the chaotic three-body problem using deep neural networks

Philip G. Breen[1⋆†], Christopher N. Foley[2 ⋆‡], Tjarda Boekholt[3]
and Simon Portegies Zwart[4]

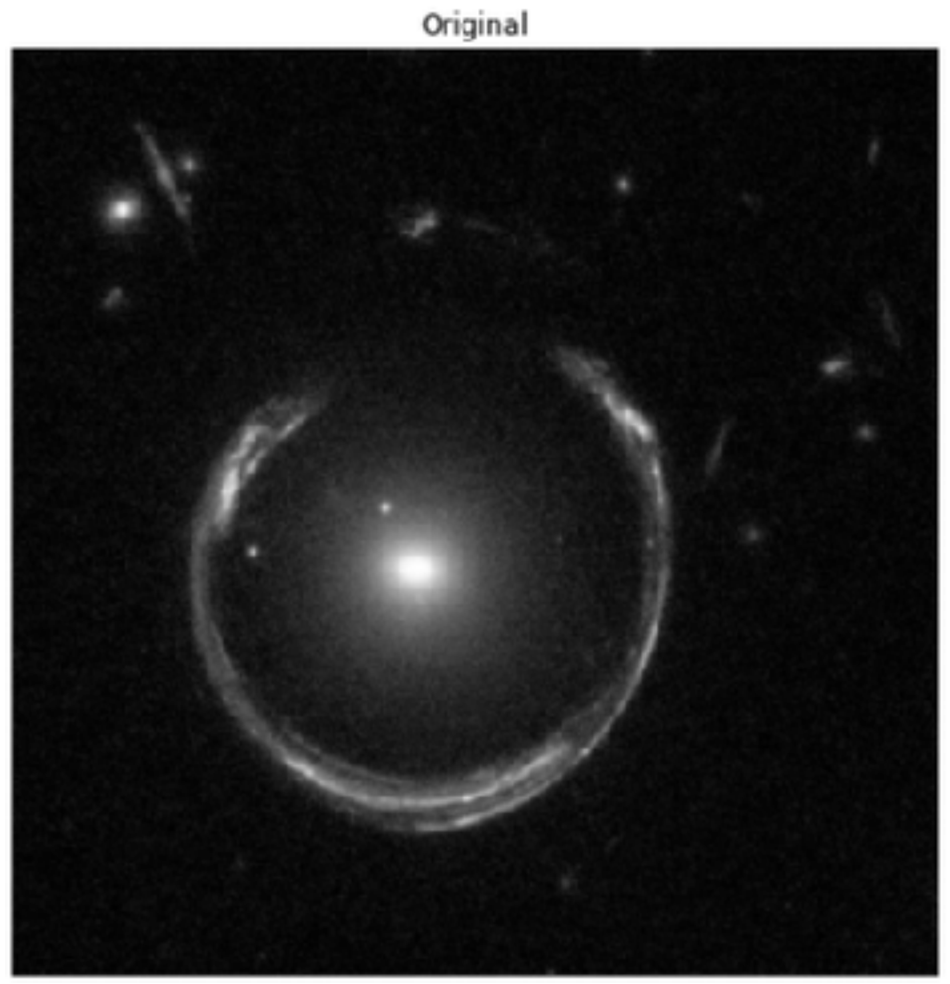


Breen et al. (2020)

(step 1) Solve 3-body problem
with various initial conditions.
(step 2) Train the NN.
(step 3) NN solves 3-body problem
with any initial condition quickly.

CNN — Convolutional Neural Network

raw image



Convolution

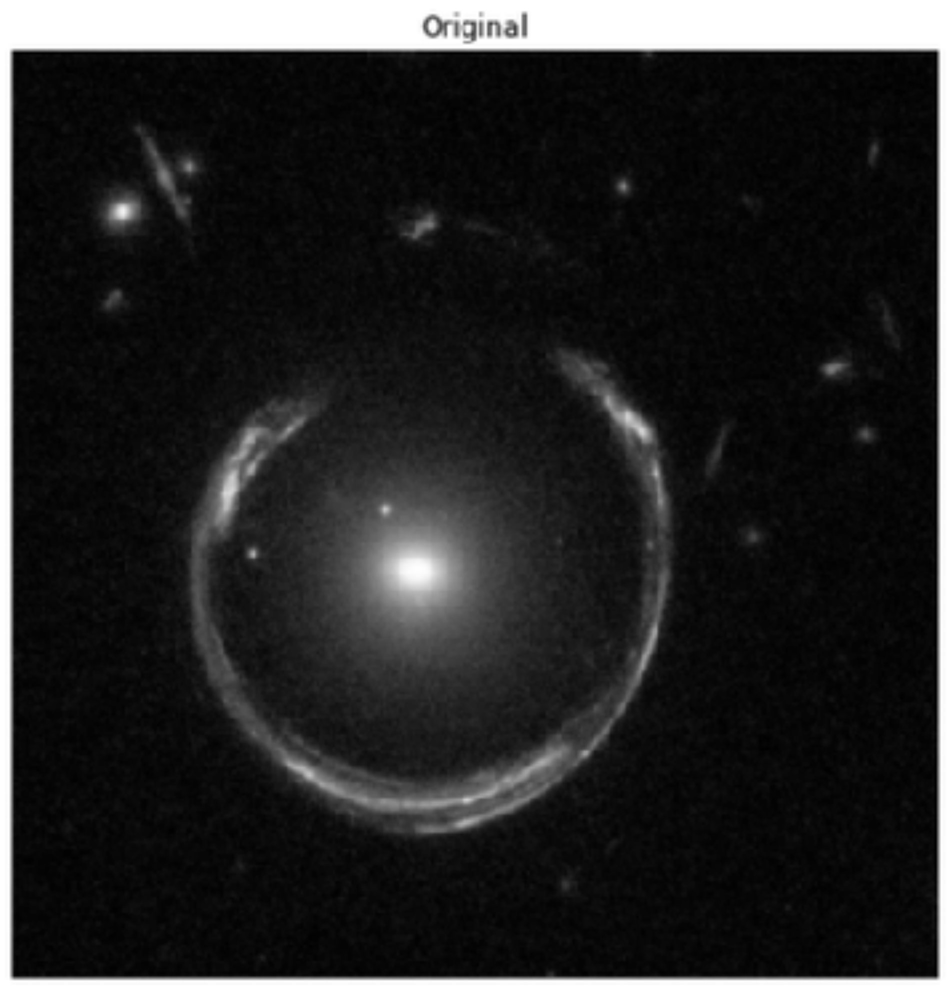
*

edge-detecting filter

-1	-1	-1	-1	-1
-1	1.78	1.78	1.78	-1
-1	1.78	1.78	1.78	-1
-1	1.78	1.78	1.78	-1
-1	-1	-1	-1	-1

CNN — Convolutional Neural Network

raw image



Convolution

$*$

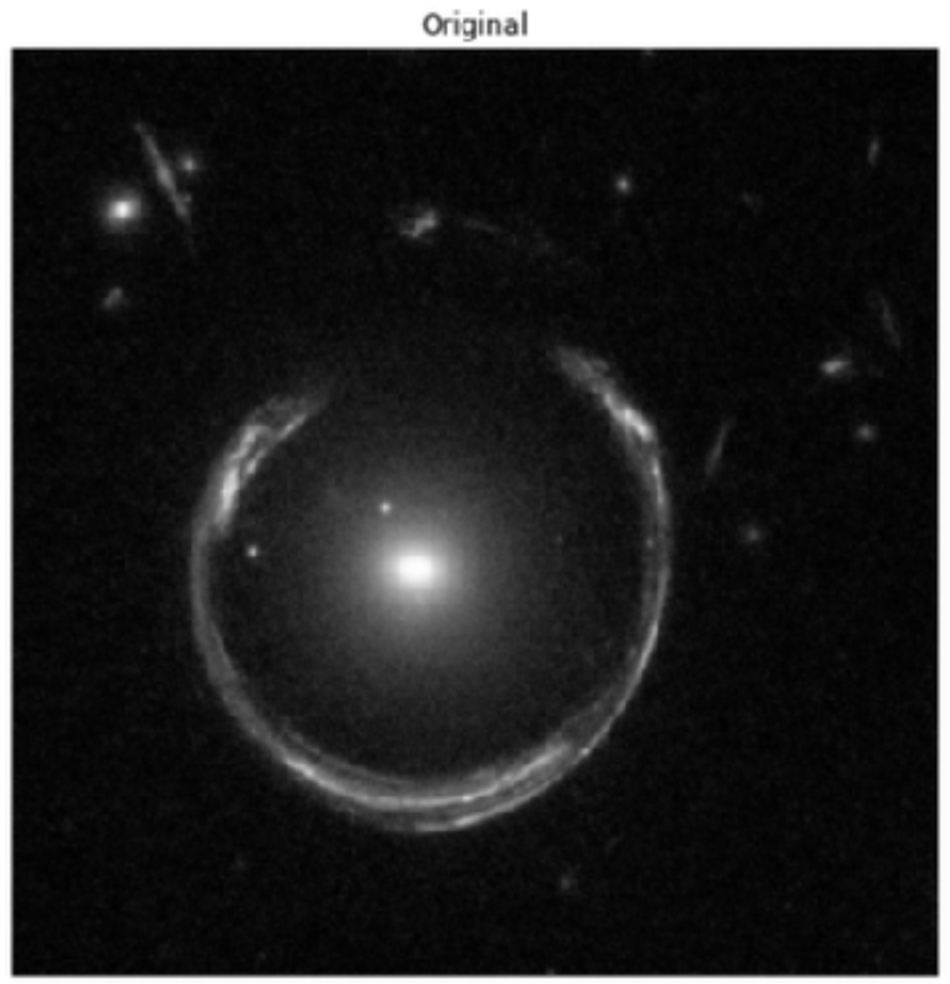
edge-detecting filter

-1	-1	-1	-1	-1
-1	1.78	1.78	1.78	-1
-1	1.78	1.78	1.78	-1
-1	1.78	1.78	1.78	-1
-1	-1	-1	-1	-1

Smooth components disappear.

CNN — Convolutional Neural Network

raw image



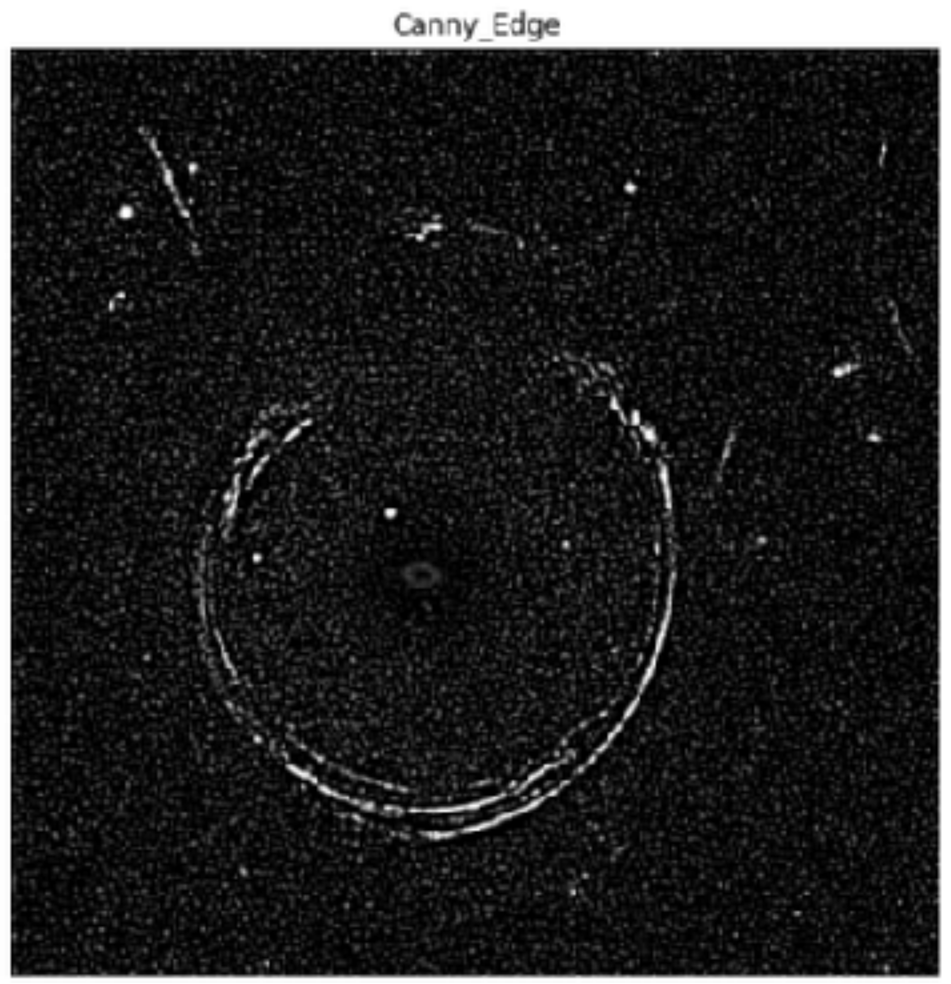
Convolution

$*$

edge-detecting filter

-1	-1	-1	-1	-1
-1	1.78	1.78	1.78	-1
-1	1.78	1.78	1.78	-1
-1	1.78	1.78	1.78	-1
-1	-1	-1	-1	-1

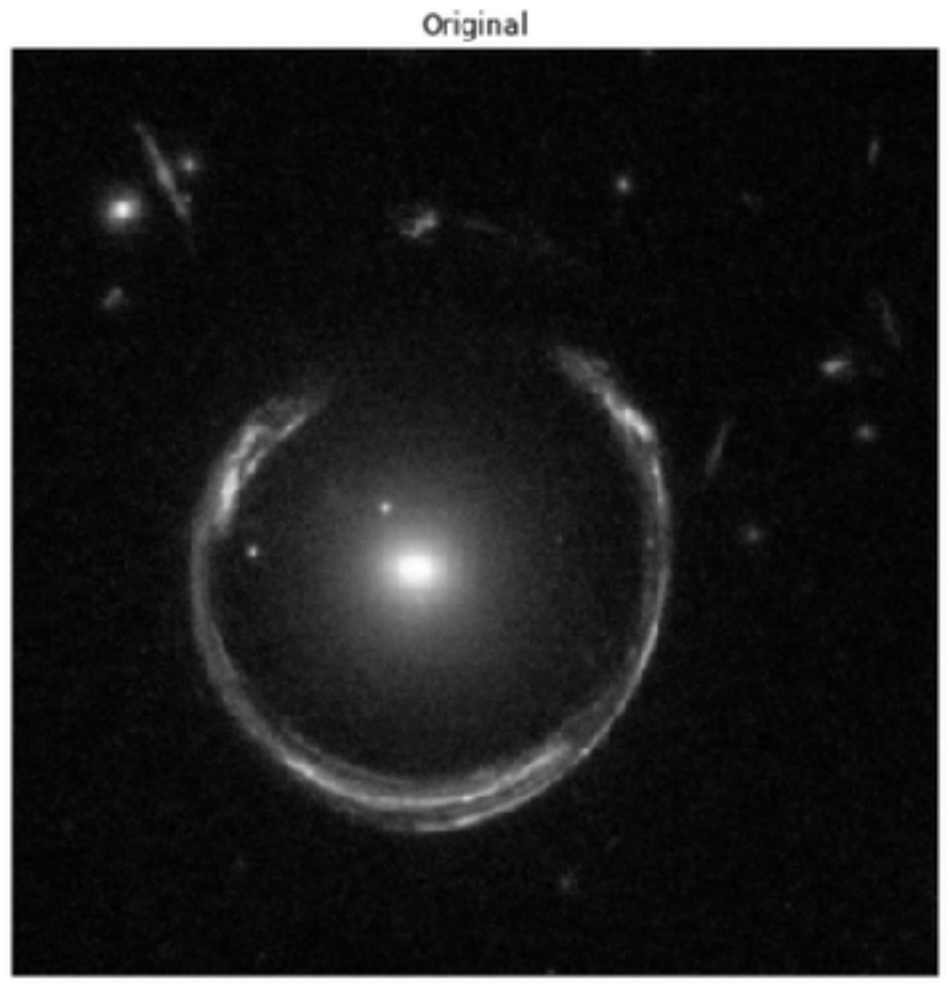
Smooth components disappear.



Convolutions extract the local pattern (e.g., “edge”) in the data.

CNN — Convolutional Neural Network

raw image



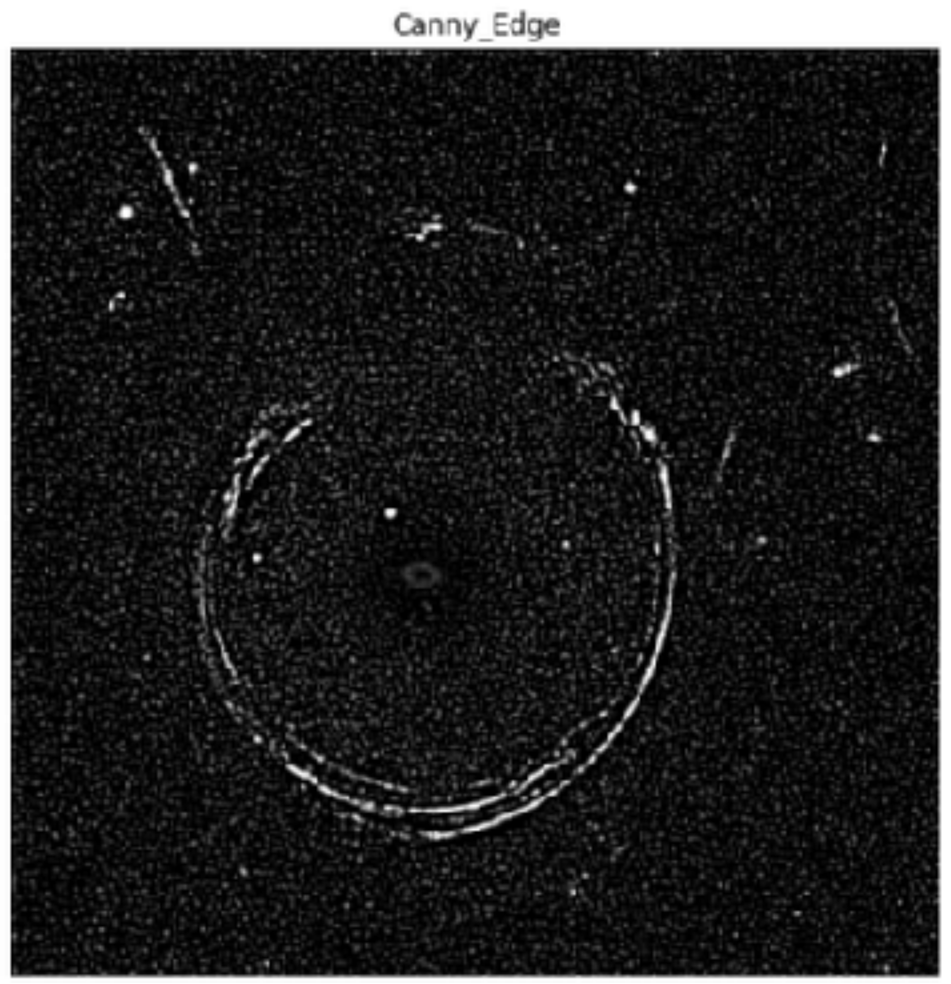
Convolution

*

edge-detecting filter

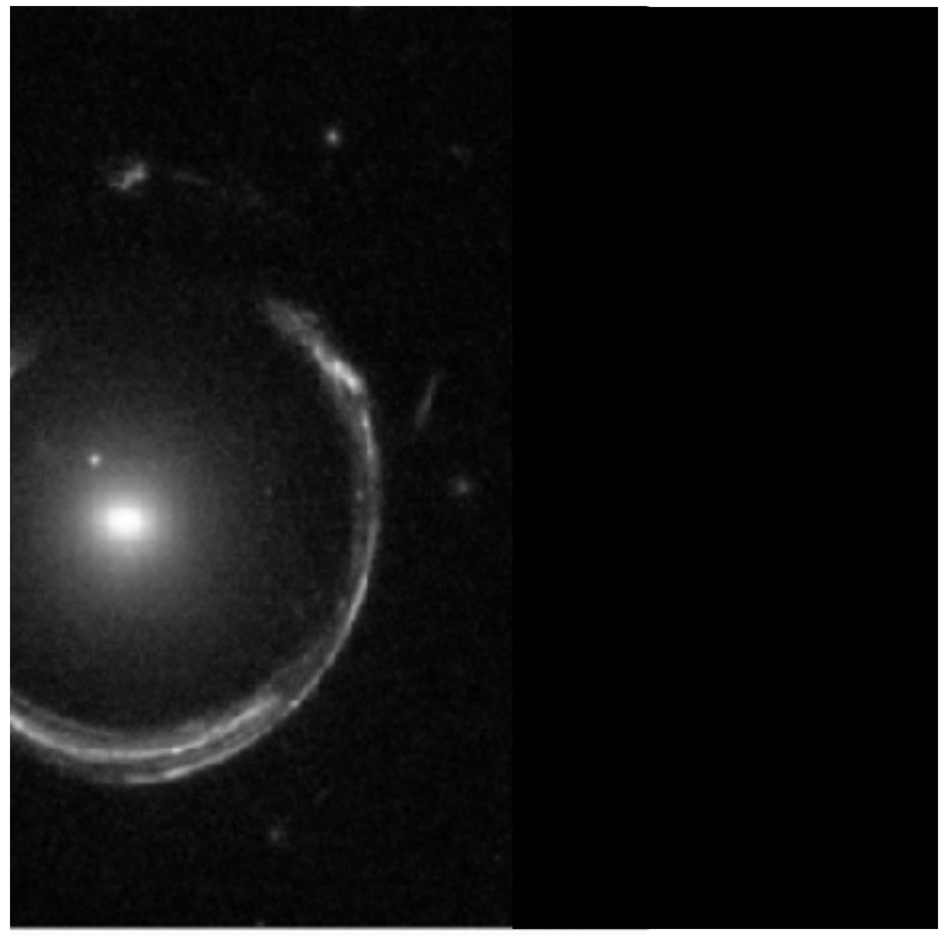
-1	-1	-1	-1	-1
-1	1.78	1.78	1.78	-1
-1	1.78	1.78	1.78	-1
-1	1.78	1.78	1.78	-1
-1	-1	-1	-1	-1

Smooth components disappear.



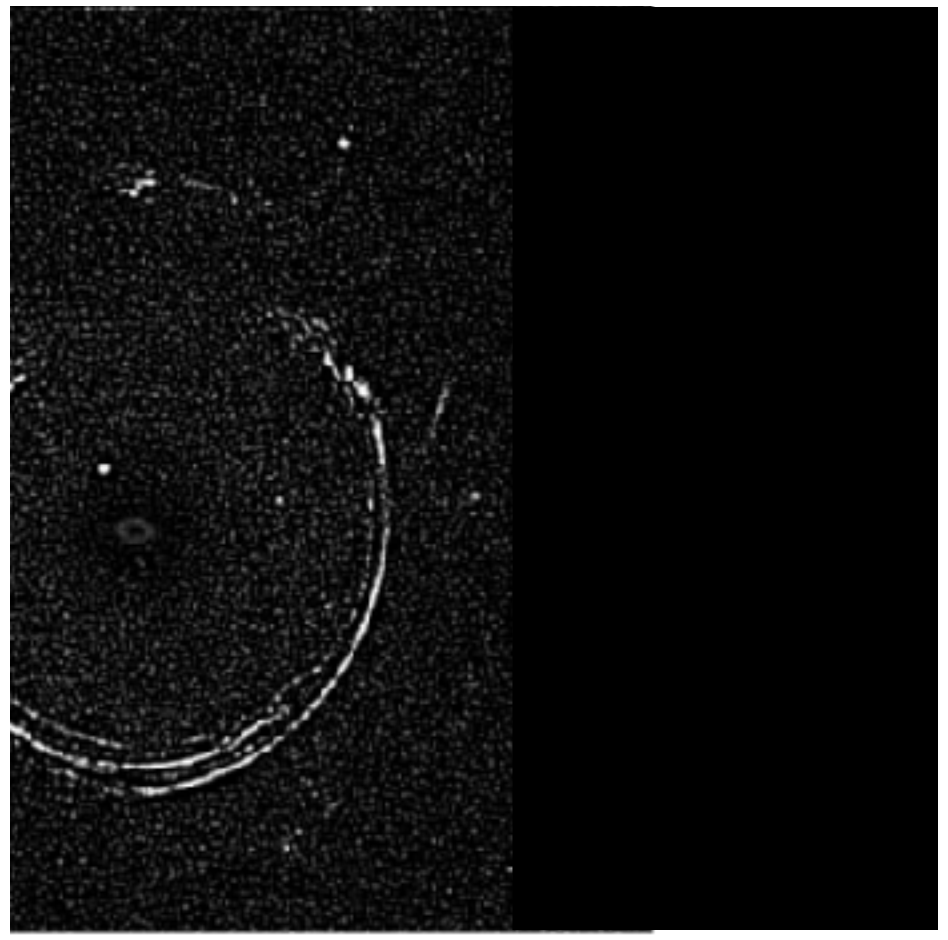
Convolutions extract the local pattern (e.g., "edge") in the data.

The extracted information is **translation invariant.**



*

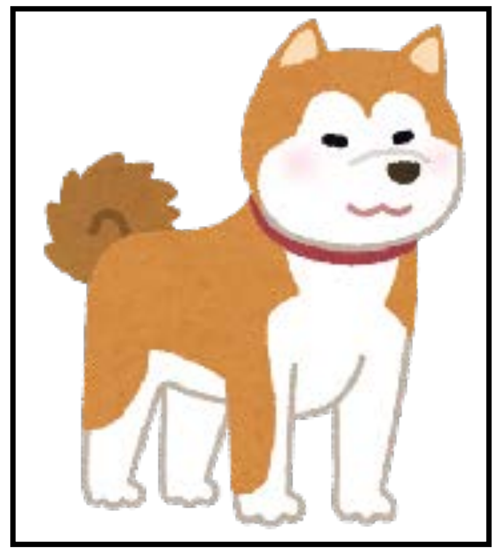
-1	-1	-1	-1	-1
-1	1.78	1.78	1.78	-1
-1	1.78	1.78	1.78	-1
-1	1.78	1.78	1.78	-1
-1	-1	-1	-1	-1



CNN — Convolutional Neural Network

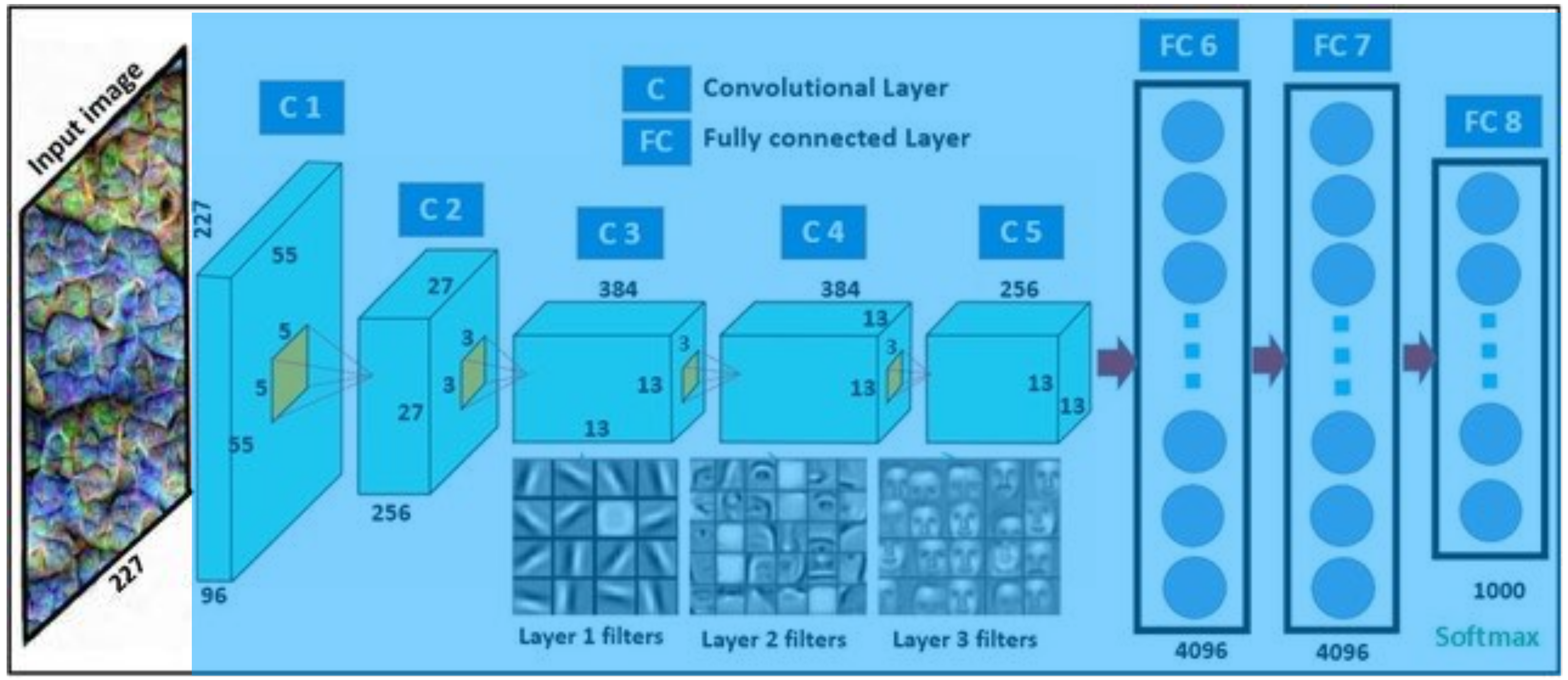
AlexNet 2012

Krizhevsky et. al. (2012)
Hassan & Gousev (2019)



(c) いらすとや

input image



output array

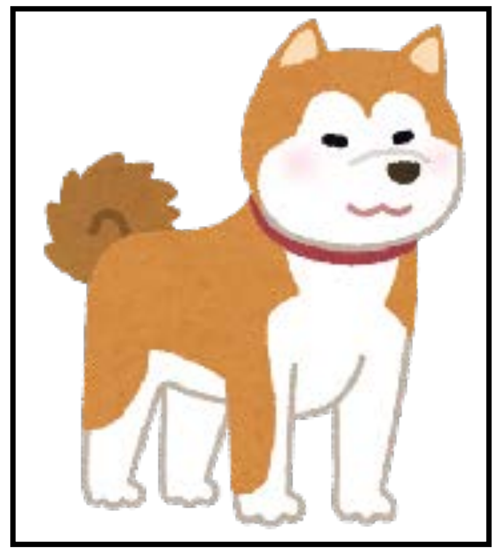
$$\begin{bmatrix} 0.01 \\ 0.00 \\ \vdots \\ 0.94 \\ 0.02 \end{bmatrix}$$

P(dog) = 0.94

CNN — Convolutional Neural Network

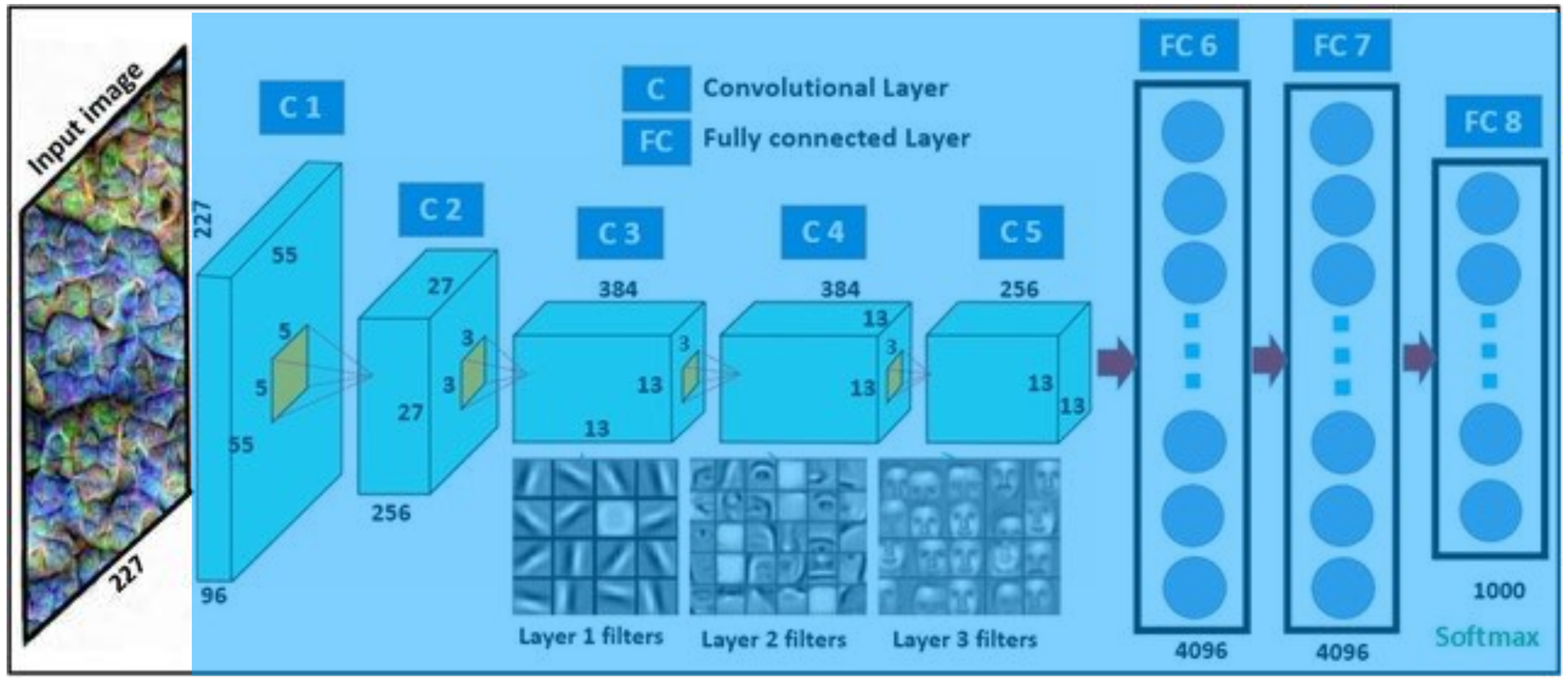
AlexNet 2012

Krizhevsky et. al. (2012)
Hassan & Gousev (2019)



(c) いらすとや

input image



output array

$$\begin{bmatrix}0.01\\0.00\\ \vdots \\0.94\\0.02\end{bmatrix}$$

P(dog) = 0.94

Convolution = Extract "local patterns" that are translation invariant.
Various filters are used = Various patterns are extracted.

-1	-1	-1
-1	8	-1
-1	-1	-1

-1	0	1
-1	0	1
-1	0	1

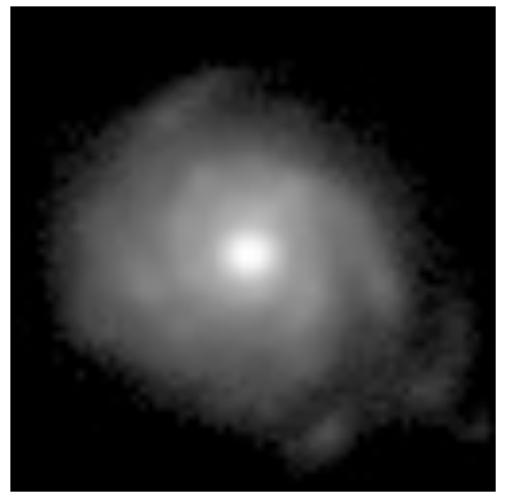
-1	-1	-1
0	0	0
1	1	1

(original work) input = 2D image … (N x N x 3) array for RGB color image

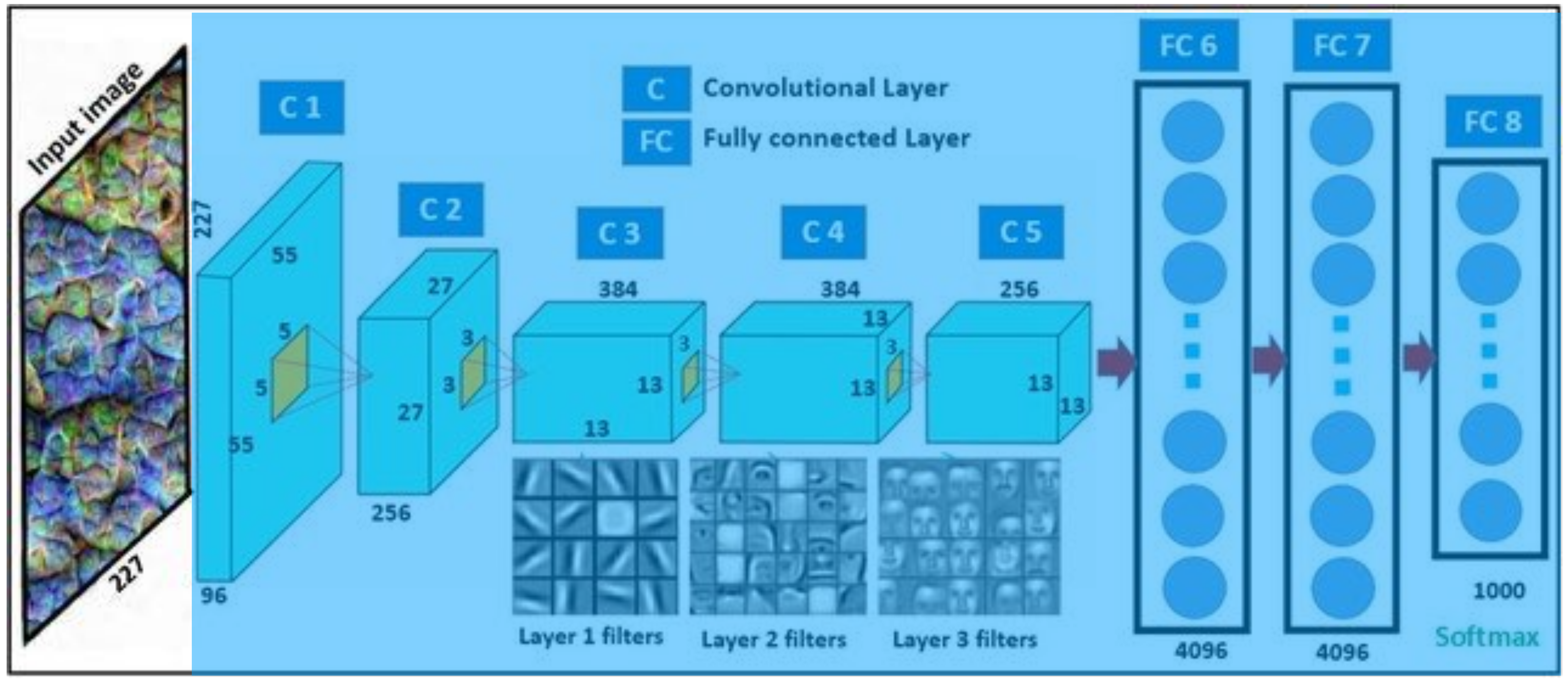
CNN — Convolutional Neural Network

AlexNet 2012

Krizhevsky et. al. (2012)
Hassan & Gousev (2019)



input image



output array

$$\begin{bmatrix} 0.01 \\ 0.00 \\ \vdots \\ 0.94 \\ 0.02 \end{bmatrix}$$

P(spiral) = 0.94

Convolution = Extract "local patterns" that are translation invariant.
Various filters are used = Various patterns are extracted.

-1	-1	-1
-1	8	-1
-1	-1	-1

-1	0	1
-1	0	1
-1	0	1

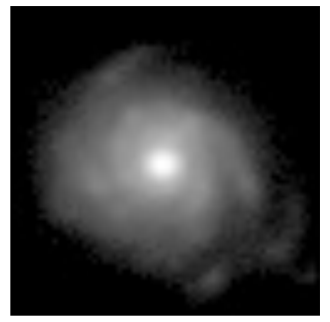
-1	-1	-1
0	0	0
1	1	1

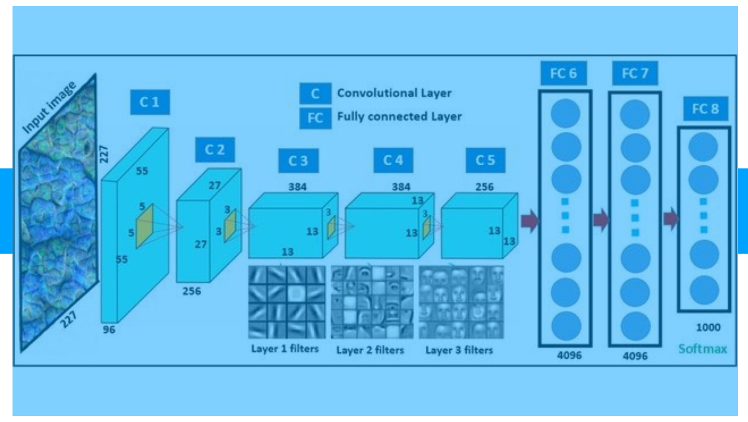
(original work) input = 2D image … (N x N x 3) array for RGB color image
(astro works) input = 2D image … (N x N x 1) array for gray-scale image
(astro works) input = 1D vector … 1D spectra

CNN — Convolutional Neural Network

CNN

2D image (galaxy)



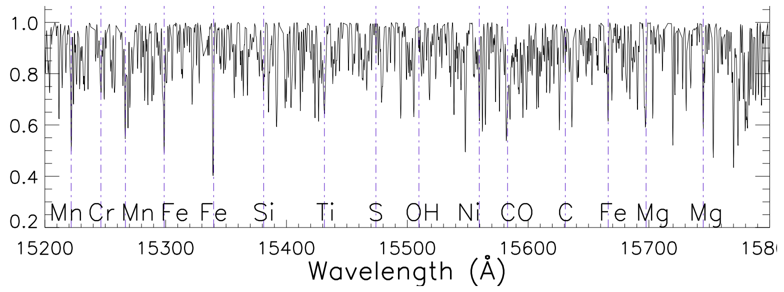


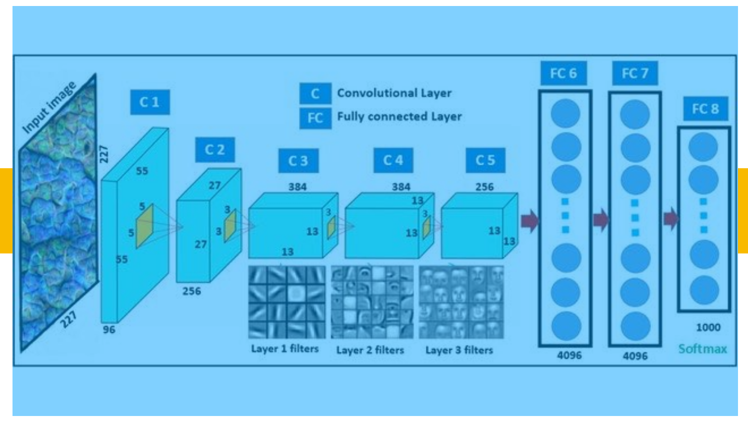
Tadaki et al. (2020)

classification

galaxy morphology

1D stellar spectrum (stellar physics)



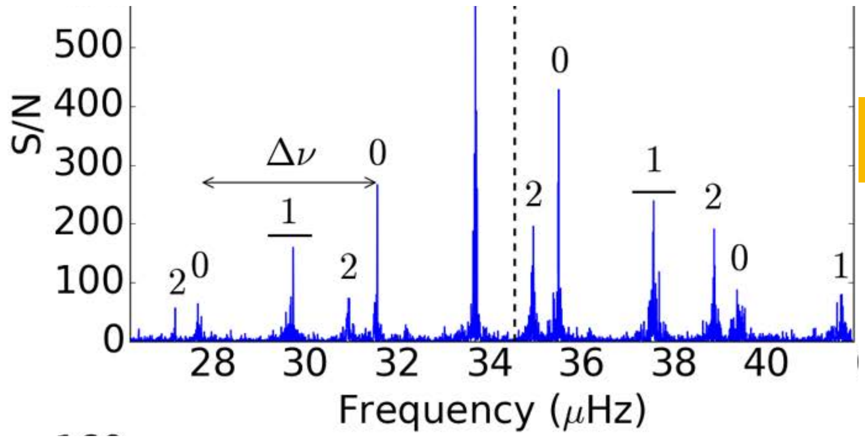


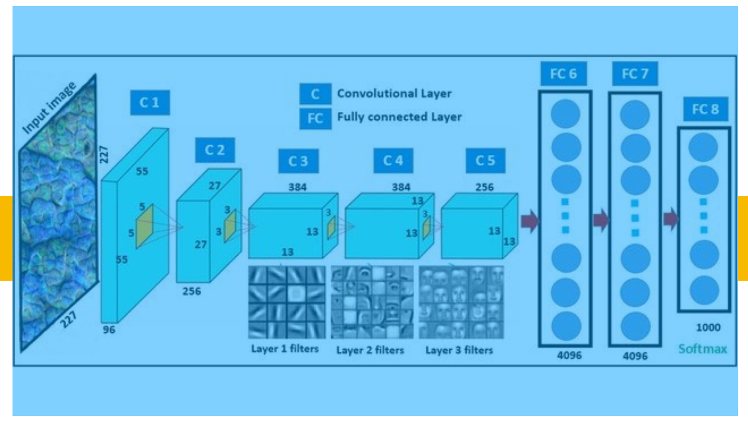
Leung & Bovy (2019)

inference

chemical abundances

1D power spectrum (asteroseismology)



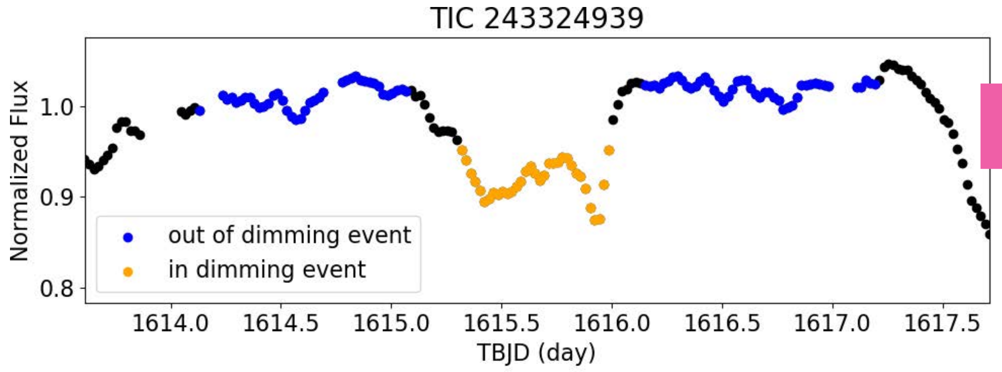


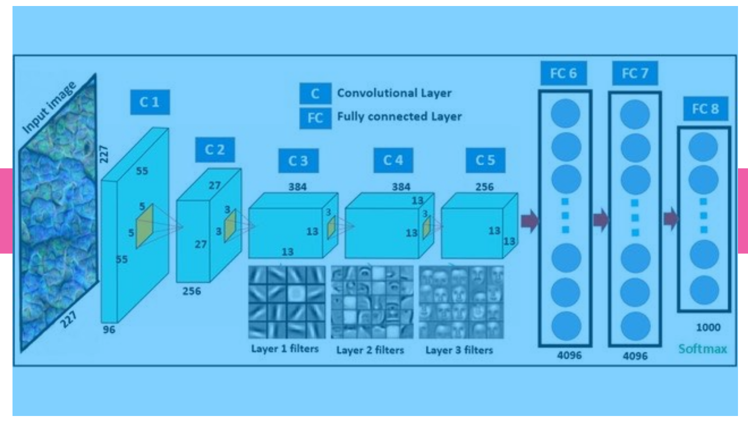
Hon et al. (2017)

inference

stellar parameters

1D light curve (dimming stars)





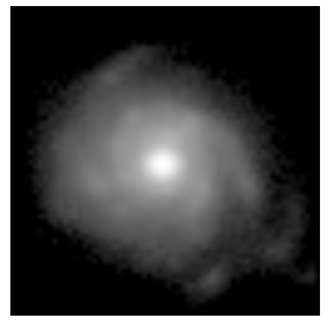
Tajiri et al. (2020)

identification

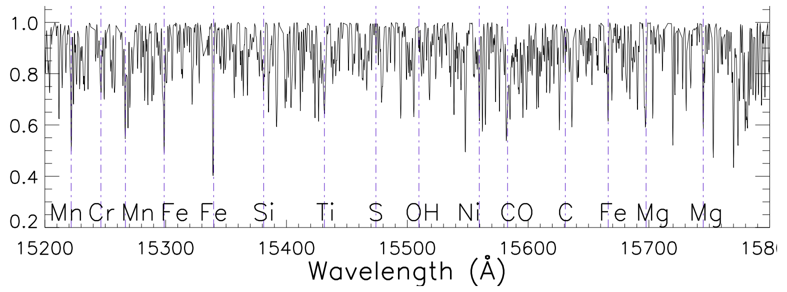
Find a specific type of stars

CNN — Convolutional Neural Network

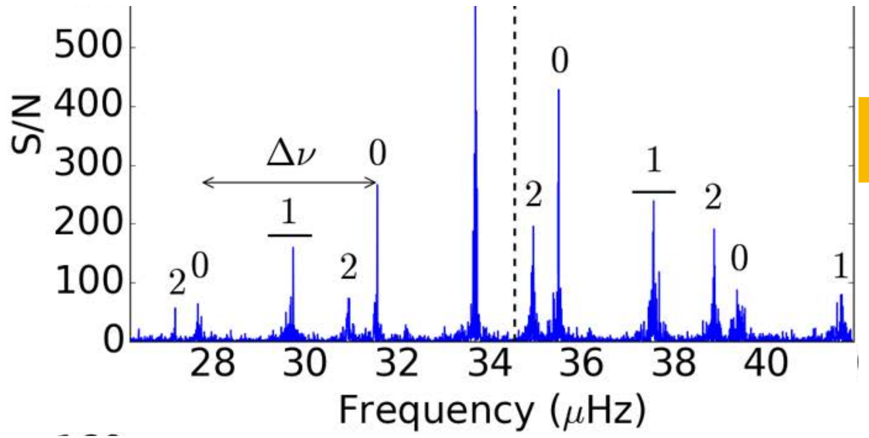
2D image (galaxy)



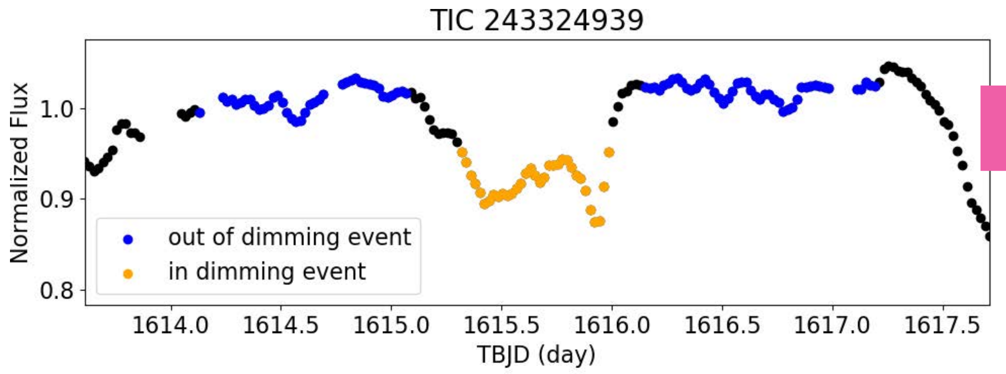
1D stellar spectrum (stellar physics)



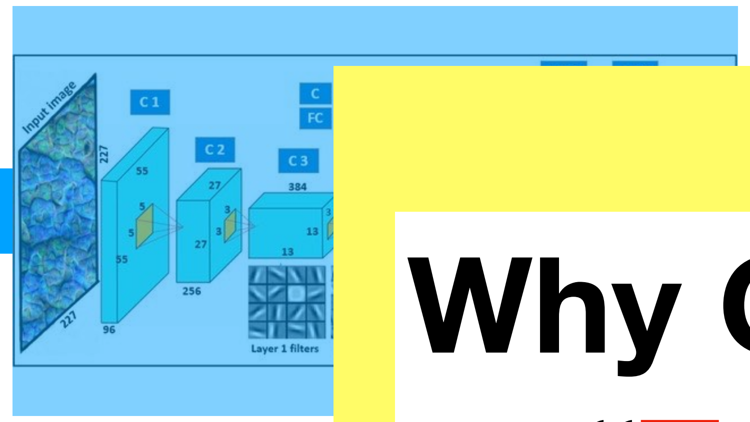
1D power spectrum (asteroseismology)

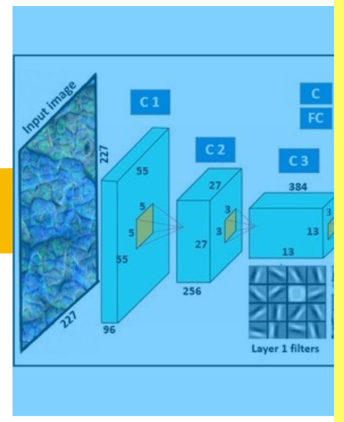


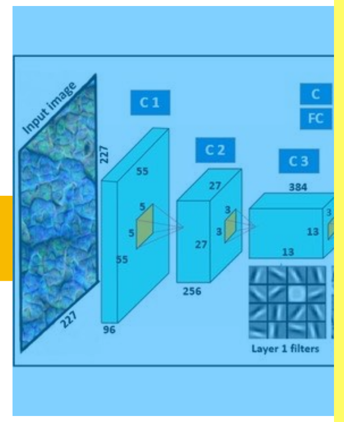
1D light curve (dimming stars)

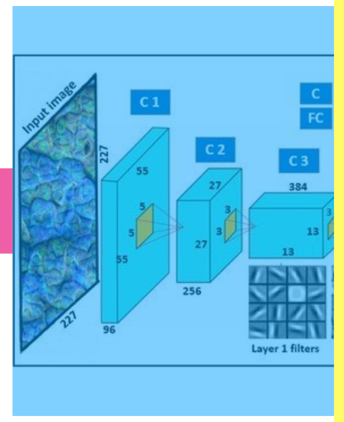


CNN









Why CNN works for various data?
— "Translational invariance"

2D image: Freedom to choose the origin

1D stellar spectrum: Doppler shift

1D power spectrum: Δν

1D light curve: Freedom to choose the origin

CNN is suited for natural science.
— Reason for
"Inflation of ML universe"

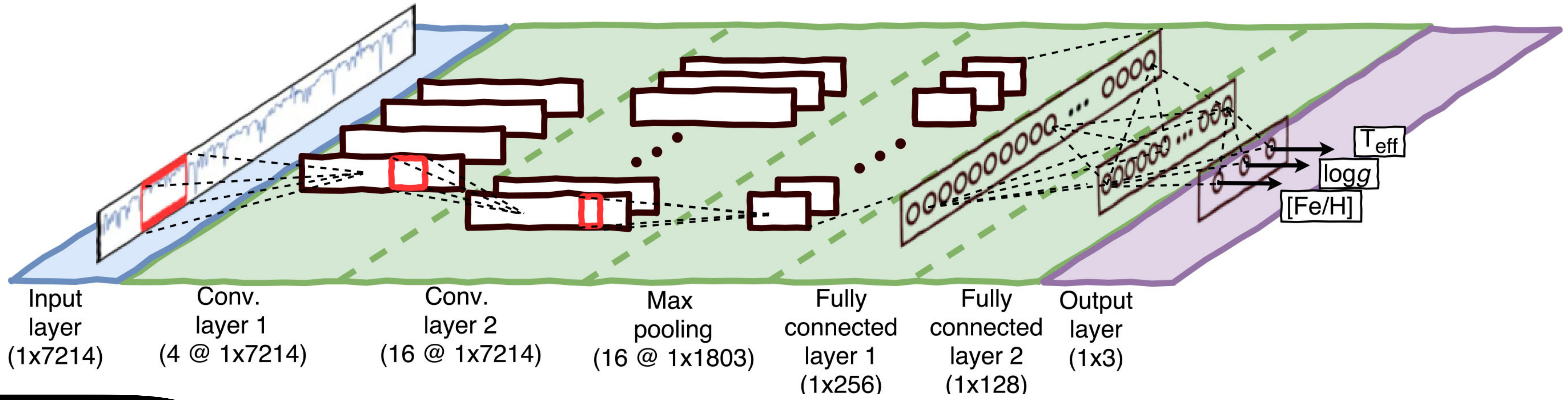
Some concerns about neural network

NN is sometimes described as a “black box.”

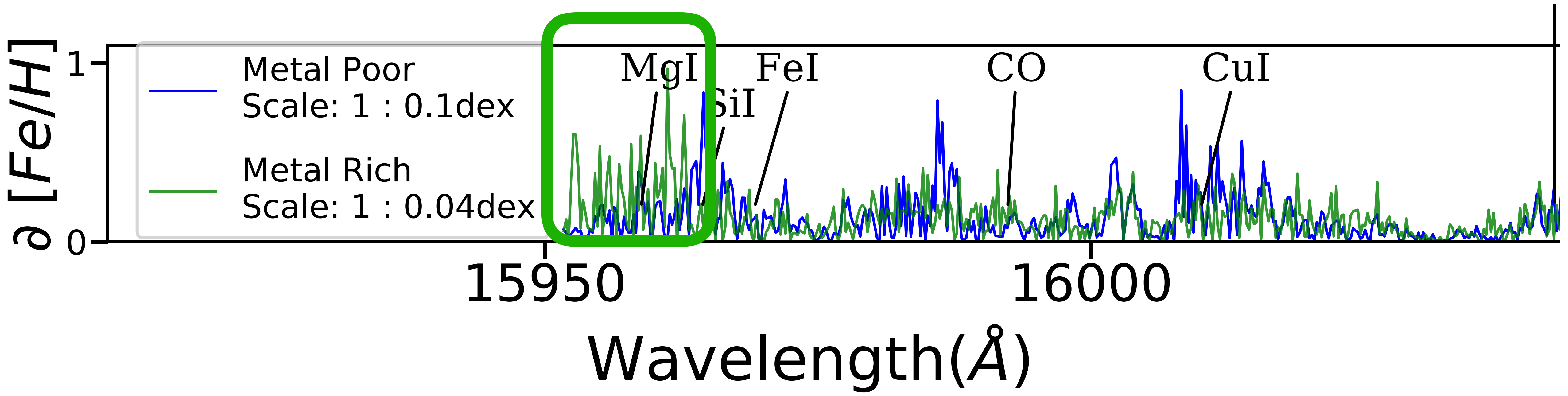
- **Bias in the test data**
- **Interpretability of NN — “Explainable AI” (XAI)**
- Uncertainty quantification — Bayesian NN
- High degrees of freedom — No unique solution.

Bias in the test data

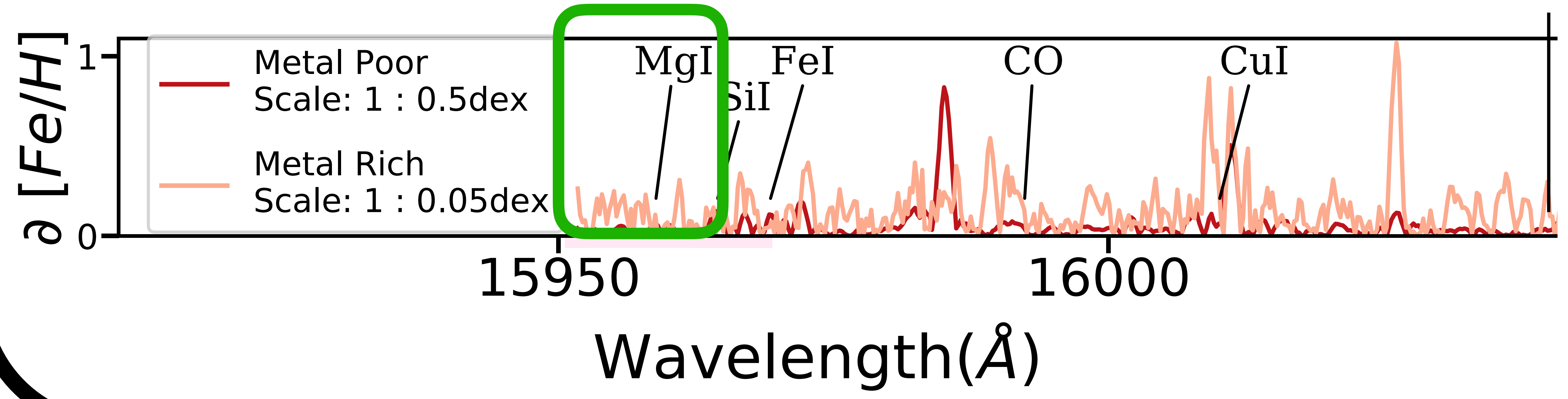
Example Fabbro et al. (2018)



(1) Train a NN to learn the relationship between [X/H] and spectra using APOGEE spectra



(2) Train the same NN to learn the relationship between [X/H] and spectra using synthetic spectra



Result:

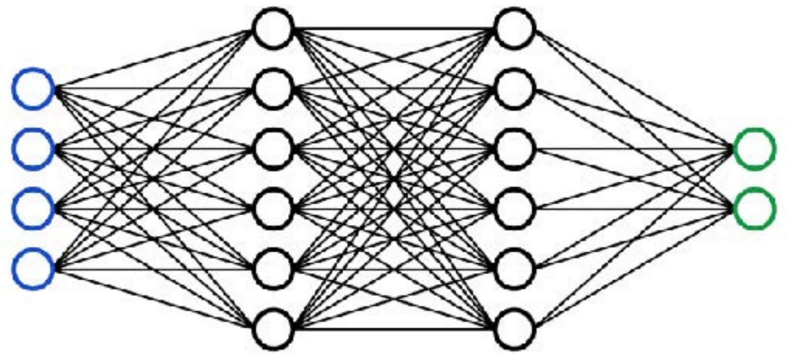
NNs prioritize different lines to infer [X/H], because synthetic spectra ≠ APOGEE

A bias in test data will affect performance.

When designing stellar spectra pipeline, this bias should be kept in mind.

Modified in the public version

Interpretability of NN



A funny example

Train NN to learn the relationship
between chemistry ([Fe/H]) and observation data using
(1) observation fits file (spectra, date, observer's name etc..)
(2) [X/H] from another catalog
… Supervised learning.

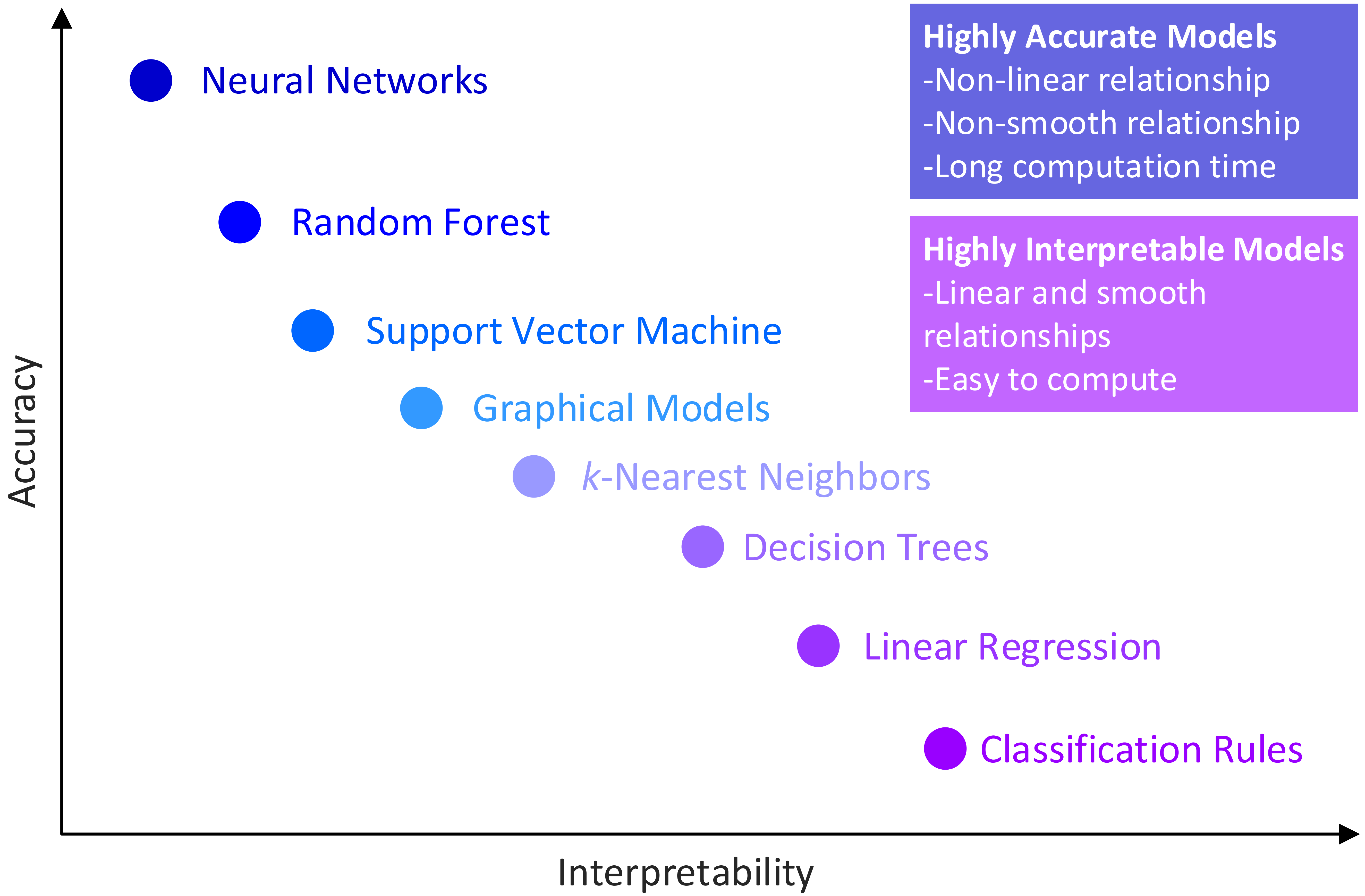
Result:
NN learned how to measure [Fe/H] from fits file.

This result *might be* wrong!
NN might learn the correlation between observer's name and chemistry.)
("Dr. AAA only observes low-[Fe/H] stars," etc.)

Do not use NN as a black-box.

Trade-off: interpretability vs accuracy

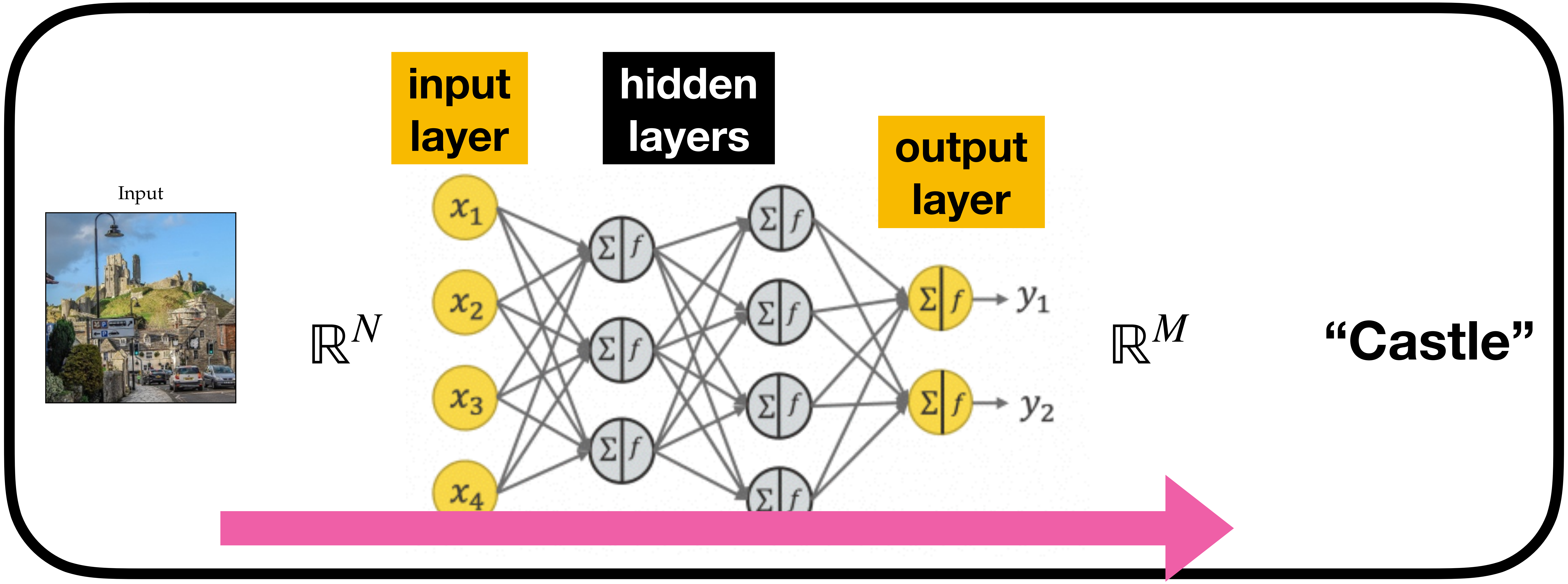
accuracy

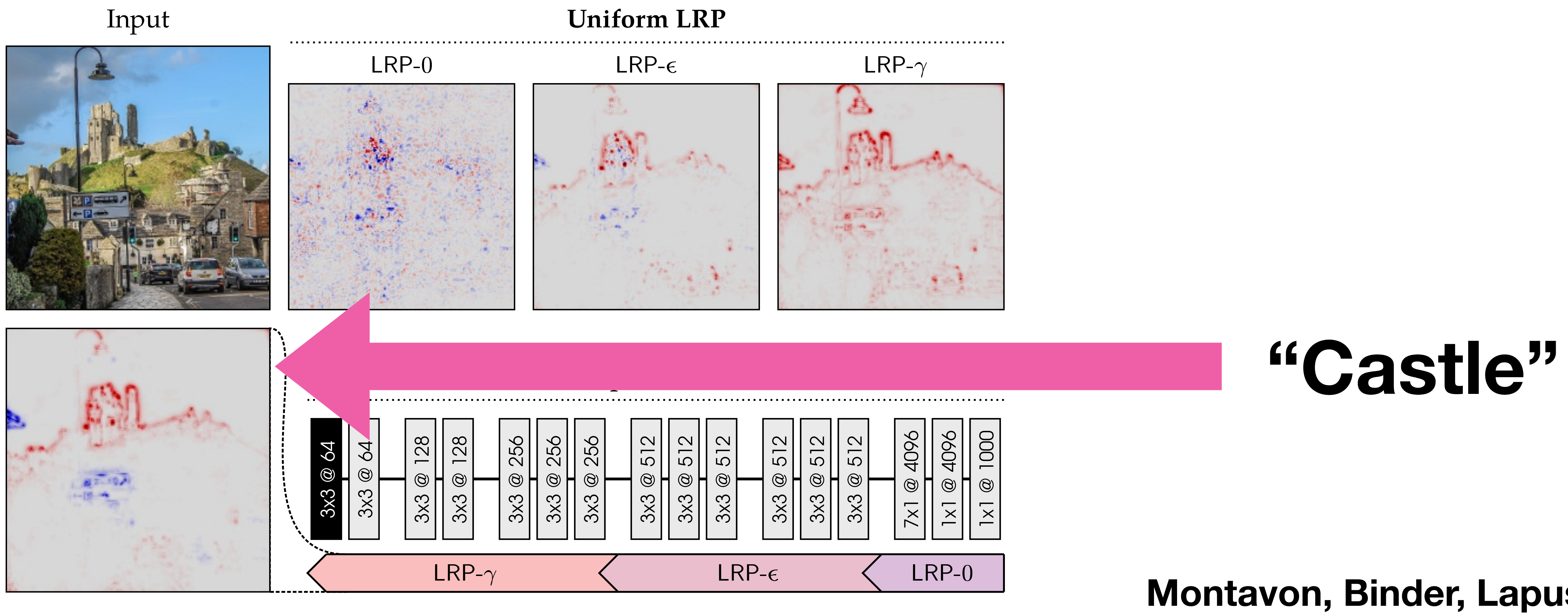


interpretability

From “Machine Learning for 5G/B5G Mobile and Wireless Communications: Potential, Limitations, and Future Directions“

Toward explainable AI: Layer-wise Relevance Propagation





"Red part of the image was useful for classification"

Montavon, Binder, Lapuschkin et al. (2019)

- (1) Era of big data
- (2) Dimensionality reduction
- (3) Sparsity
- (4) Bayesian analysis
- (5) Machine learning
- (6) Deep learning
- (7) Data challenge

Robustness / reliability of new methods

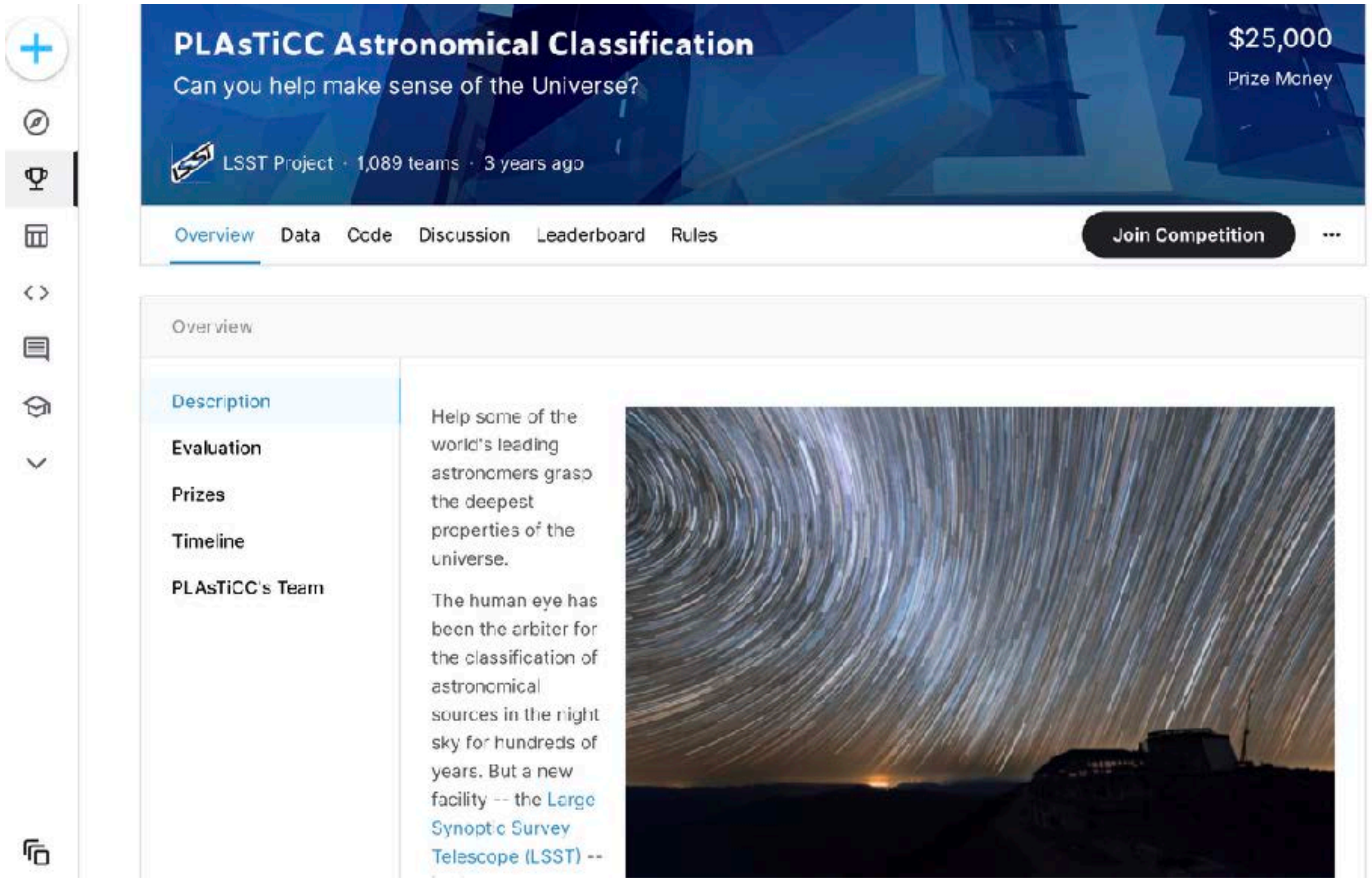
Data Challenge (Test to validate methods)

- Photometric LSST Astronomical Time-Series Classification Challenge (https://www.kaggle.com/c/PLAsTiCC-2018)
- **Exoplanet**: ARIEL Mission Data Challenges (https://www.ariel-datachallenge.space/ML/documentation/description)
- **Radio astronomy**: SKA Data Challenge Competition #1 (https://astronomers.skatelescope.org/ska-science-data-challenge-1/)
- **Microlensing** Data Challenge (https://microlensing-source.org/data-challenge/)
- Galaxy Zoo: **galaxy morphology** classification challenge (https://www.kaggle.com/c/galaxy-zoo-the-galaxy-challenge)
- **Mapping dark matter** competition (https://www.kaggle.com/c/mdm/overview)
- **Strong lensing** data challenge (https://bolognalensfactory.wordpress.com/home-2/blfkids-lens-finding-challenge/)
- **Gaia challenge** (http://astrowiki.ph.surrey.ac.uk/dokuwiki/doku.php)

Photometric LSST Astronomical Time-Series Classification Challenge

***Kaggle*-based competition of classifying mock LSST data**
with generous prize money :)

1st prize: Kyle Boone (Astro PhD student)
2nd prize: Mike & Silogram (Non-astro group)
3rd prize: Major Tom, mamas & nyanp (Non-astro group)



Inflation of ML universe.

Is our community open to data scientists, given the need for ML talents?

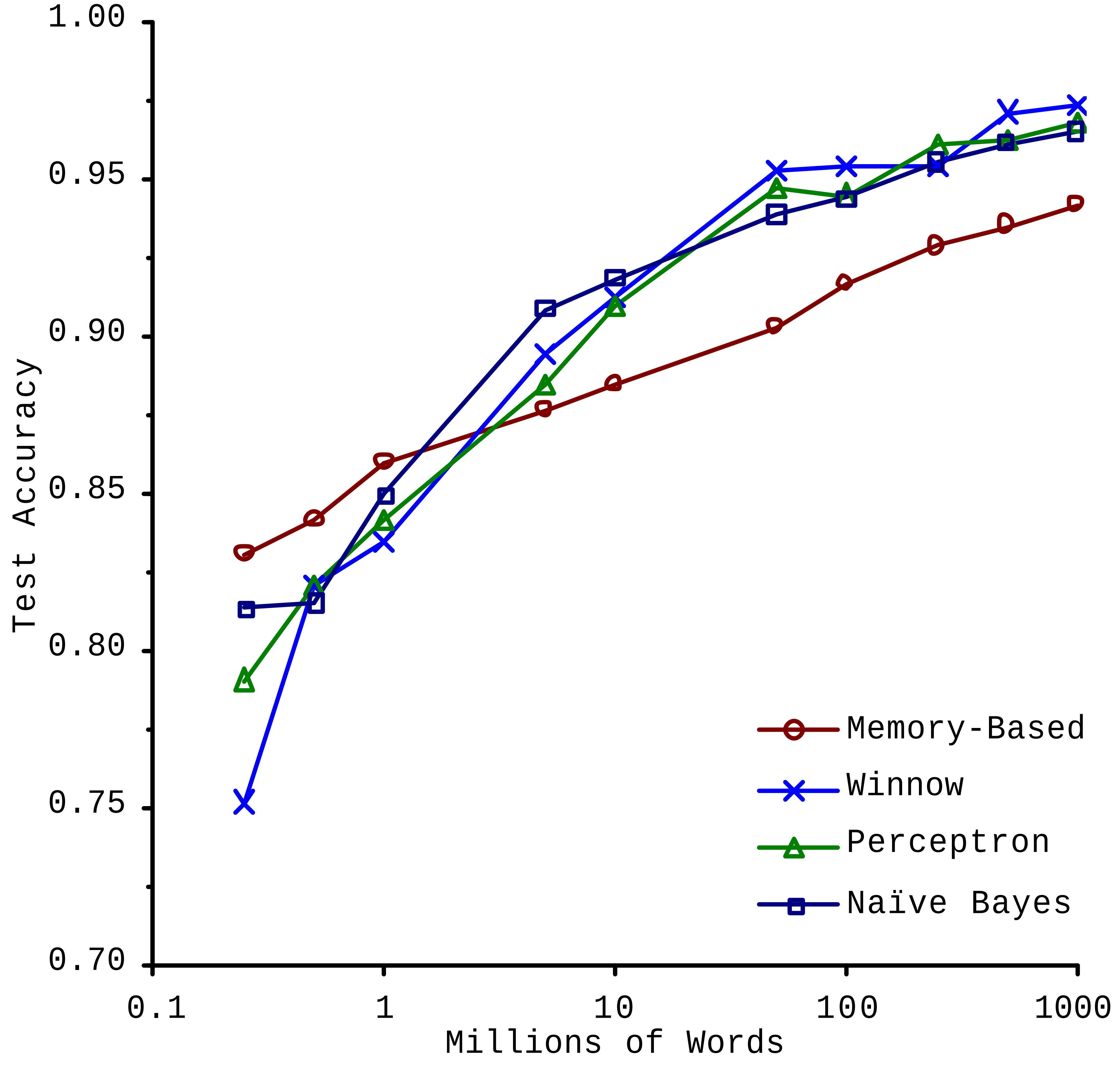
Data challenges for
- **mock ULTIMATE-Subaru data?**
- **mock JASMINE data?**
- **mock [your favorite project] data?**

… These will galvanize young/enthusiastic members of astro/non-astro community!!

Concluding remarks: Era of Big Data

Natural language tasks
[自然言語処理=機械英文チェックのテスト]

Score (Accuracy)

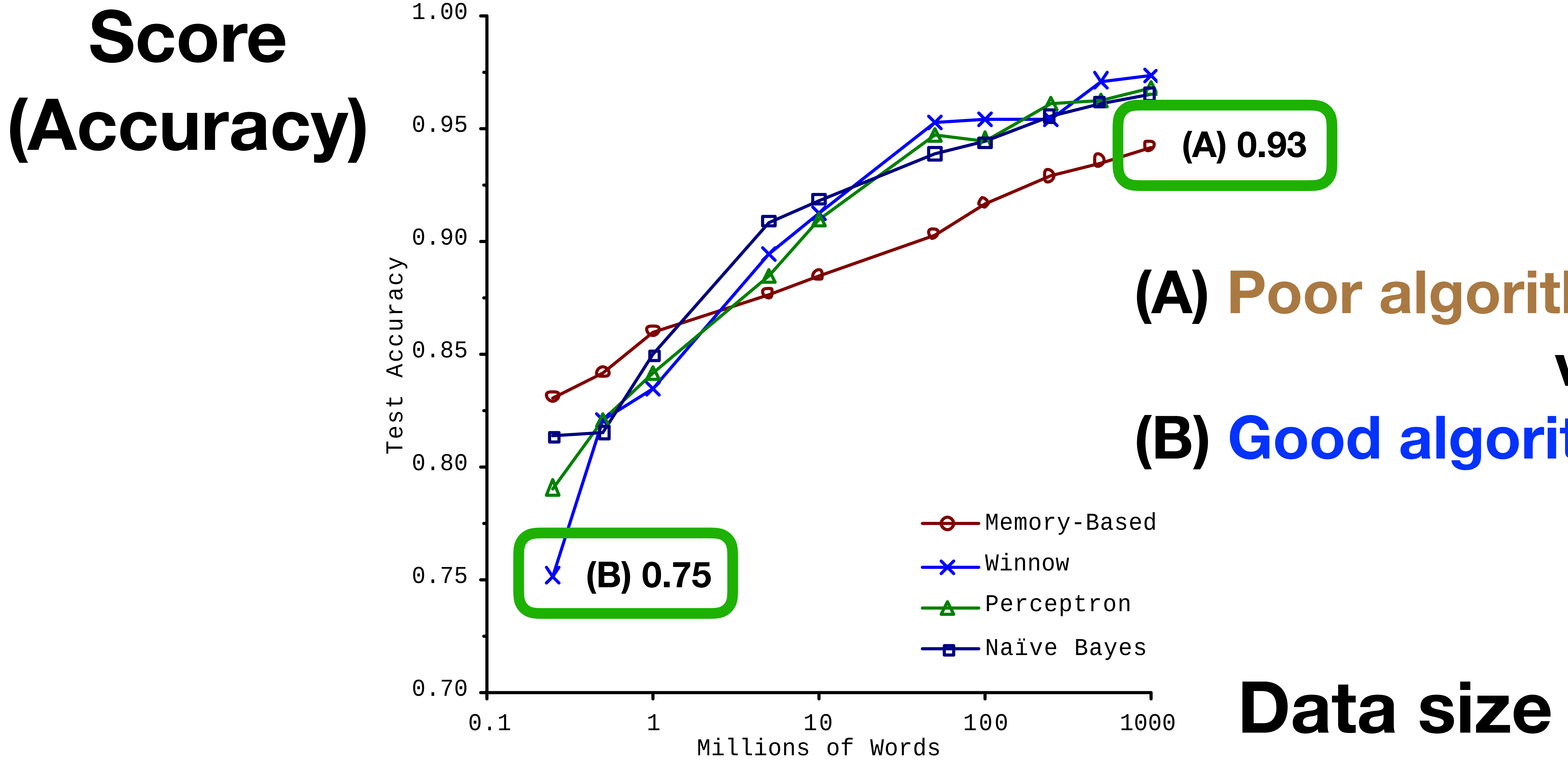


Data size

Banko & Brill (2001) “Scaling to Very Very Large Corpora for Natural Language Disambiguation”

Concluding remarks: Era of Big Data

It's NOT who has the best algorithms that wins.
It's who has the most data.



(A) Poor algorithm + Big data

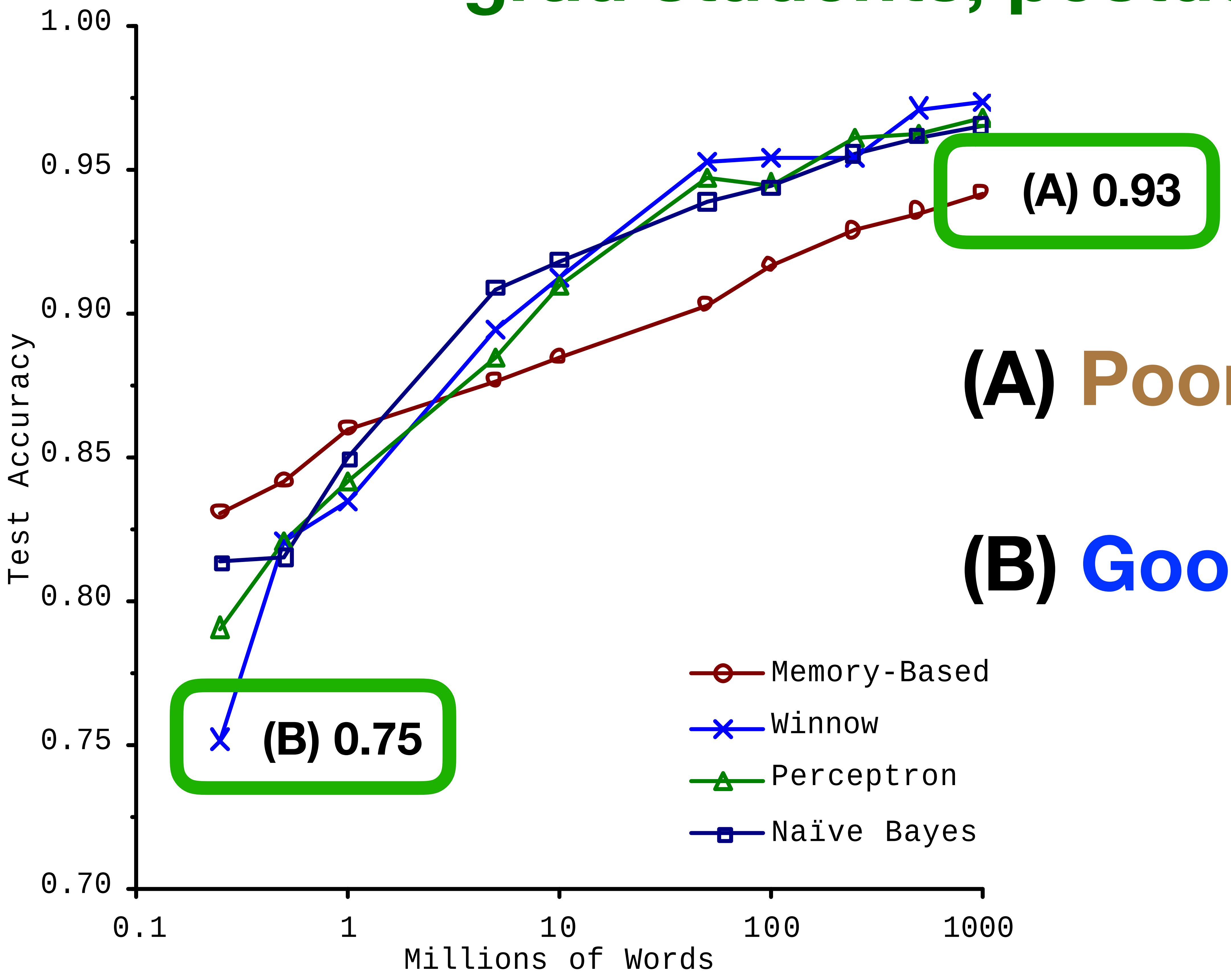
vs

(B) Good algorithm + Small data

Banko & Brill (2001) "Scaling to Very Very Large Corpora for Natural Language Disambiguation"

Concluding remarks: Era of Big Data

We have the biggest data.
We need a scope for fostering data scientists,
because the future of NAOJ is on the shoulders of grad students, postdocs, and young researchers.



(A) Poor algorithm + Big data
vs
(B) Good algorithm + Small data

Banko & Brill (2001) “Scaling to Very Very Large Corpora for Natural Language Disambiguation”

End