Machine learning in cosmology and astrophysics

Ben Hoyle

Max Plank Institute for Extraterrestrial Physics
University Sternwarte Munich

String Theory, Munich, March 2018

Supervised Machine Learning Framework

Training data

Inputs: Easily measured or derived features, X

Targets: y
The quantity you want to learn.

$$y_{train} \approx \hat{y}_{train} = f(X_{train})$$

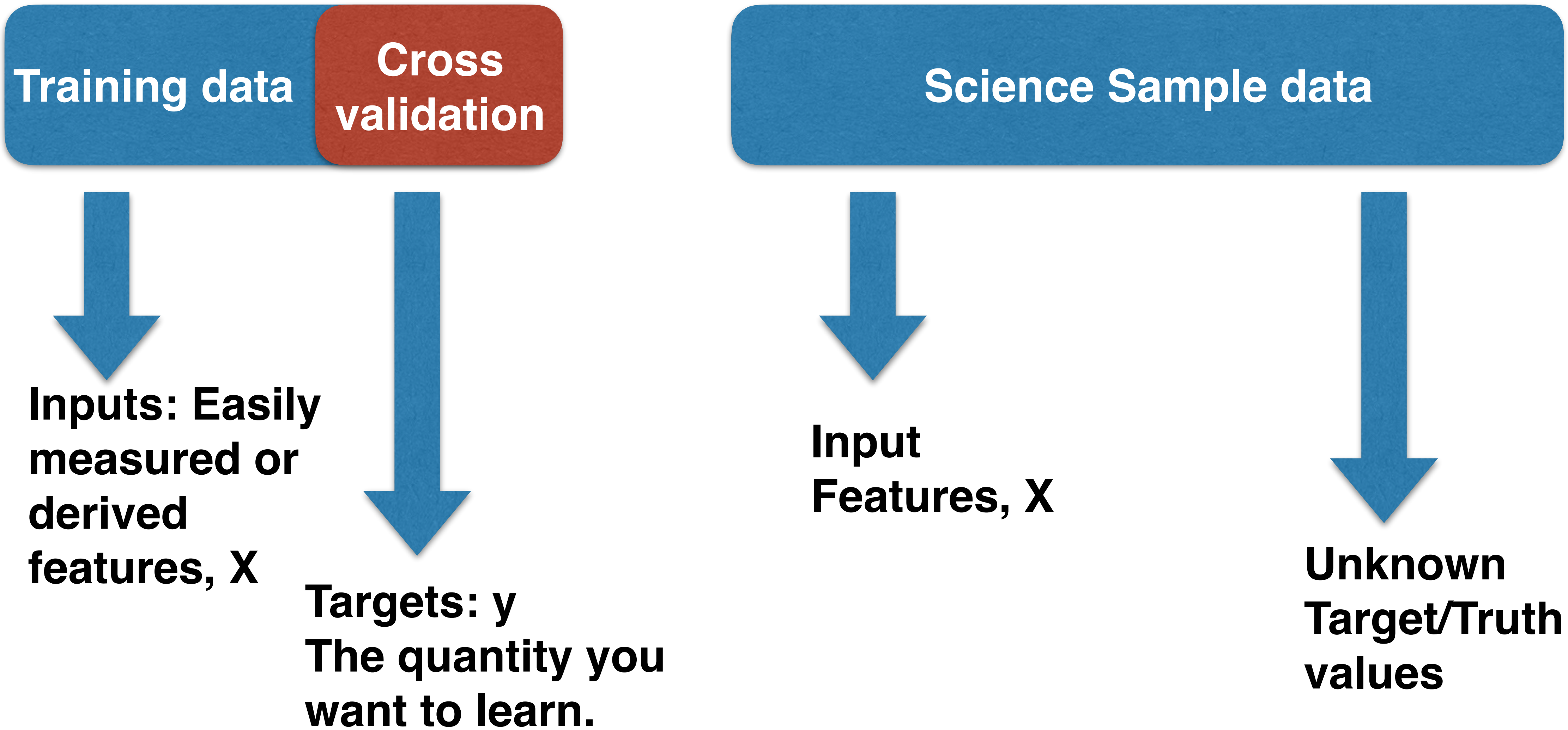
Science Sample data

Input Features, X

Unknown Target/Truth values

$$\hat{y}_{SS} = f(X_{SS})$$

Supervised Machine Learning Framework

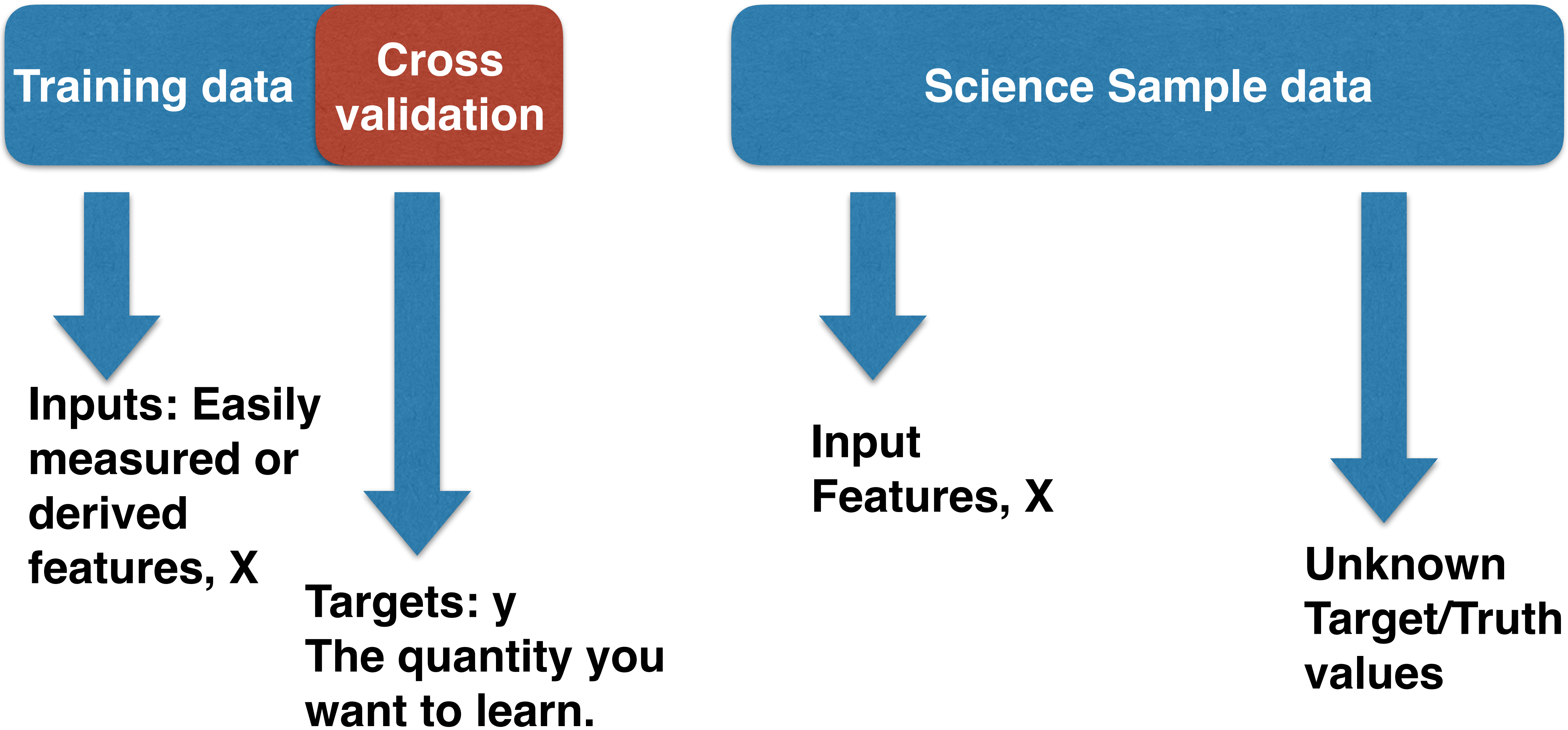


$$y_{train} \approx \hat{y}_{train} = f(X_{train}) \qquad \hat{y}_{SS} = f(X_{SS})$$

Expected Error on prediction

$$\Delta = \hat{y}_{x-val} - y_{x-val}$$

Supervised Machine Learning Framework



$$y_{train} \approx \hat{y}_{train} = f(X_{train})$$

$$\hat{y}_{SS} = f(X_{SS})$$

Expected Error on prediction

$$\Delta = \hat{y}_{x-val} - y_{x-val}$$

If the validation data is not representative of the SS data, you can't use machine learning (or any similar analysis!) to quantify how good the predictions will be on the science sample.

Overview

The supervised Machine Learning (ML) framework

An introduction to photometric redshifts

My typical ML workflow

A common ML application:

- **Photometric redshifts**

The biggest problem for ML in cosmology:

- **Unrepresentative labelled data**

Dealing with unrepresentative labelled data

Other common applications of ML

Recent, novel applications of ML

Conclusions

Overview

The supervised Machine Learning (ML) framework

An introduction to photometric redshifts

My typical ML workflow

A common ML application:

Photometric redshifts

The biggest problem for ML in cosmology:

Unrepresentative labelled data

Dealing with unrepresentative labelled data

Other common applications of ML

Recent, novel applications of ML

Conclusions

Photometric redshifts for galaxies

Photometric redshifts have been the test bed of machine learning applications in cosmology since the 1990's, recently the latest ideas from ML been ported into cosmology (see BH on the arXiv).

An galaxy imaged with a large camera and a few broad bands. We have images like this for 500 Milllion objects. It is relatively "cheap" to obtain. We can measure the properties of these galaxies from these types of images.

To use these galaxies for science, we need accurate distance information.

Photometric redshifts for galaxies

Observationally, we see that galaxies further away from us are receding faster. This is due to the expansion of the Universe. The cosmological "redshifting" of the galaxy light, is similar to a doppler shift in frequency, and allows us to estimate the distance to the galaxy.

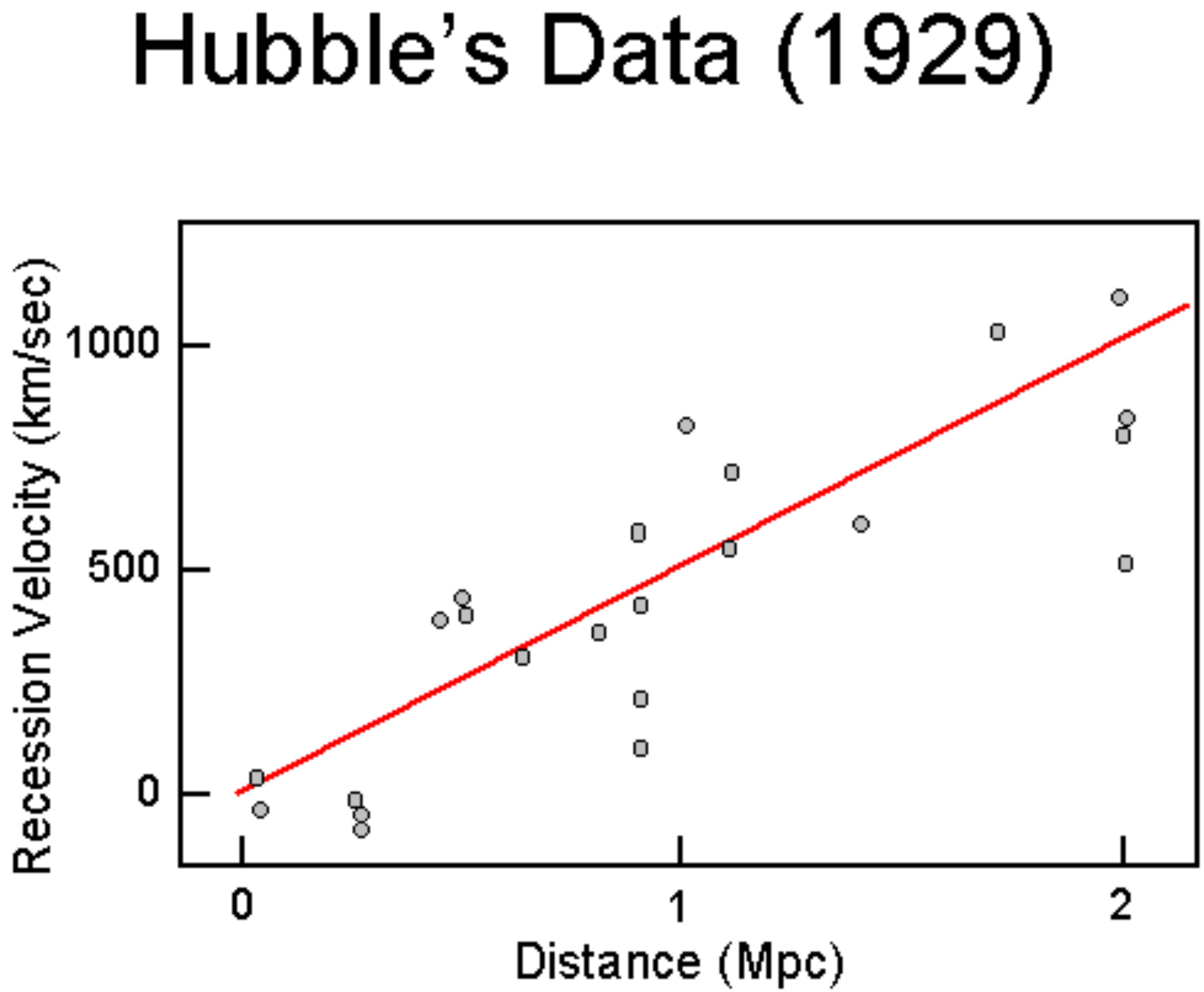


image from
http://www.astronomy.ohio-state.edu/~pogge/Ast162/Unit5/expand.html

Photometric redshifts for galaxies

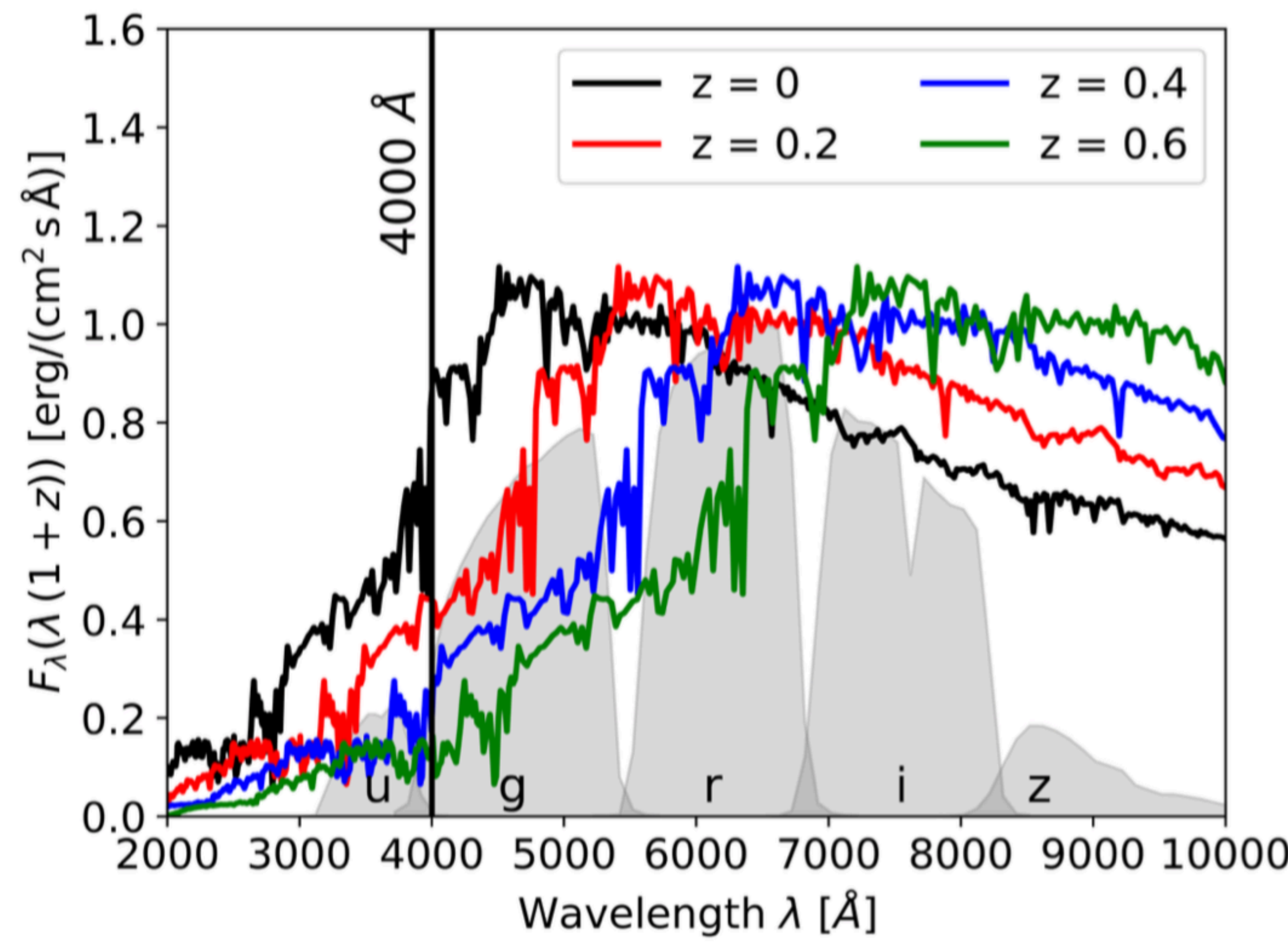
Photometric redshifts have been the test bed of machine learning applications in cosmology since the 1990’s, recently the latest ideas from ML been ported into cosmology (see BH on the arXiv).

Photometric redshifts for galaxies

Photometric redshifts have been the test bed of machine learning applications in cosmology since the 1990's, recently the latest ideas from ML been ported into cosmology (see BH on the arXiv).

Photometric redshifts for galaxies

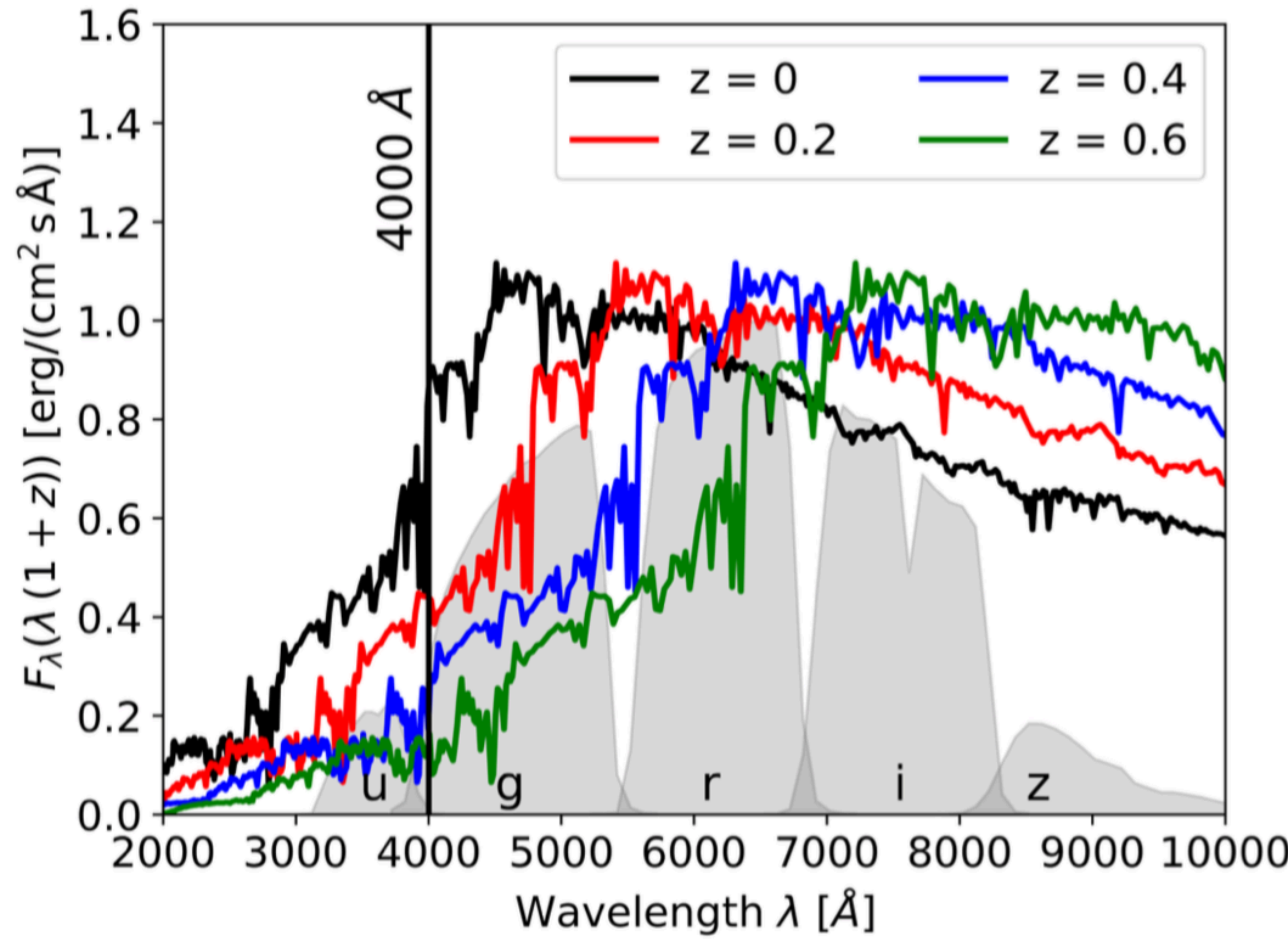
Photometric redshifts have been the test bed of machine learning applications in cosmology since the 1990's, recently the latest ideas from ML been ported into cosmology (see BH on the arXiv).



Photometric redshifts for galaxies

Photometric redshifts have been the test bed of machine learning applications in cosmology since the 1990's, recently the latest ideas from ML been ported into cosmology (see BH on the arXiv).

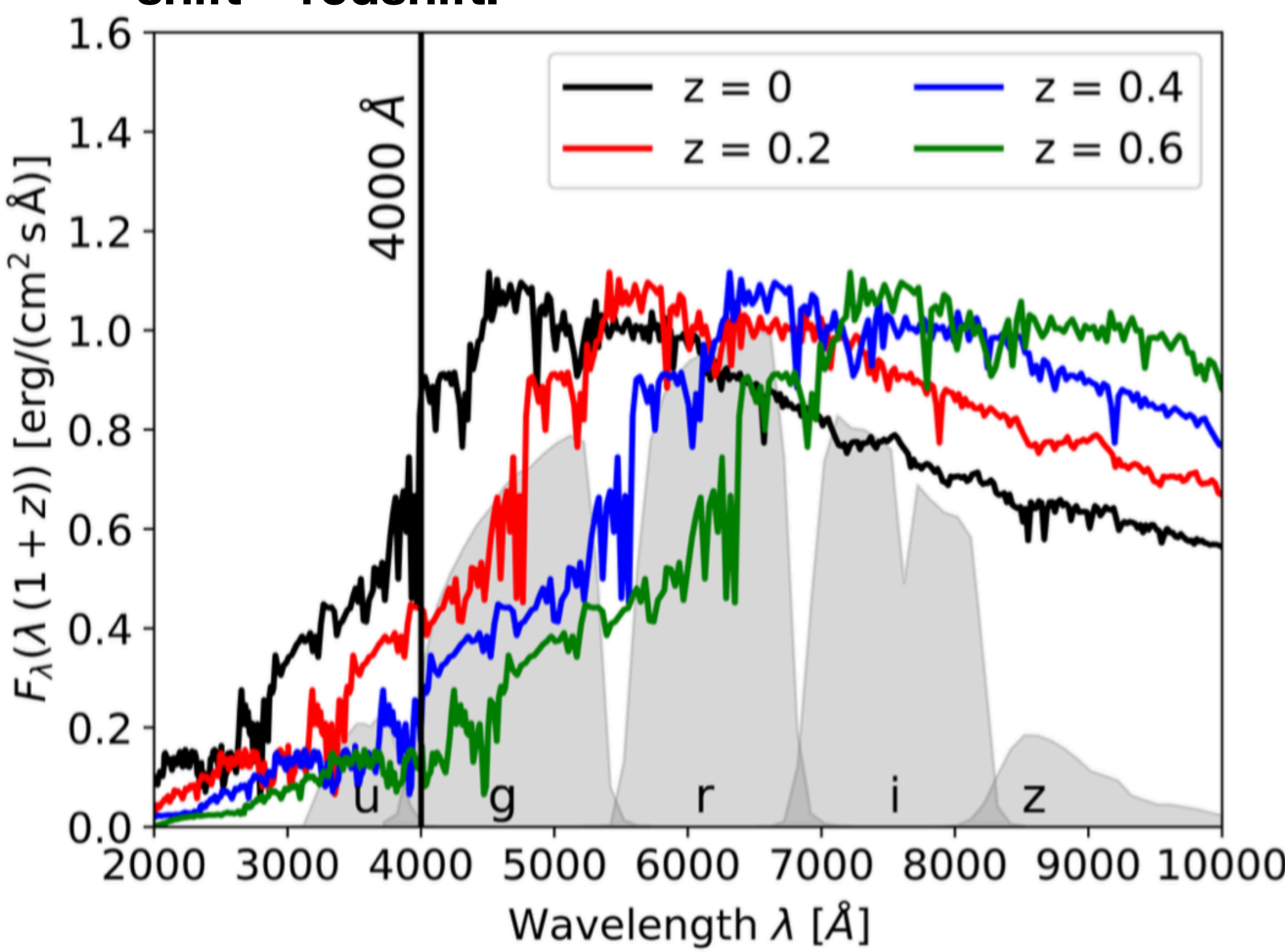
A spectrograph has a high wavelength resolution, allowing the ID of absorption/ emission lines, each with a "fingerprint". Compare to the wavelength of these fingerprints measured in the lab, and lambda shift = redshift.



Photometric redshifts for galaxies

Photometric redshifts have been the test bed of machine learning applications in cosmology since the 1990's, recently the latest ideas from ML been ported into cosmology (see BH on the arXiv).

A spectrograph has a high wavelength resolution, allowing the ID of absorption/ emission lines, each with a "fingerprint". Compare to the wavelength of these fingerprints measured in the lab, and lambda shift = redshift.



Spec-z is very accurate but expensive and is only be obtained for a small subset of data.

This corresponds to the truth values, redshift (distance) and object type "Star/Galaxy"

Photometric redshifts for galaxies

Photometric redshifts have been the test bed of machine learning applications in cosmology since the 1990's, recently the latest ideas from ML been ported into cosmology (see BH on the arXiv).

We can measure the properties of billions of galaxies from these types of images.

Use machine learning to approximate the mapping "f":

redshift = f(photometric properties of training sample)

f(photometric properties of 3 billion galaxies) => photometric redshift

Photometric redshifts for galaxies

Photometric redshifts have been the test bed of machine learning applications in cosmology since the 1990's, recently the latest ideas from ML been ported into cosmology (see BH on the arXiv).

We can measure the properties of billions of galaxies from these types of images.

Use machine learning to approximate the mapping "f":

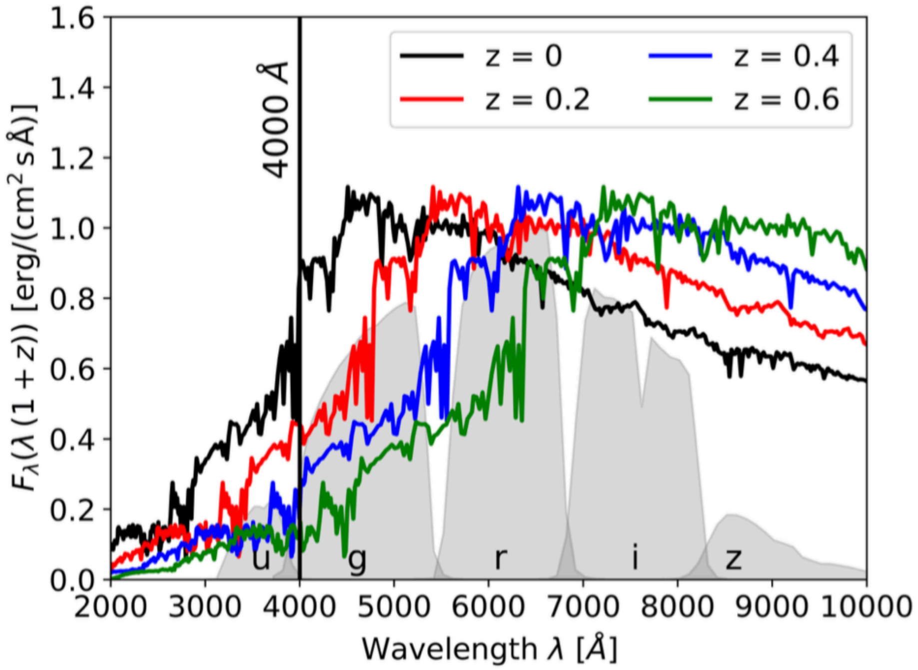
redshift = f(photometric properties of training sample)

f(photometric properties of 3 billion galaxies) => photometric redshift

Photometric redshifts for galaxies

Photometric redshifts have been the test bed of machine learning applications in cosmology since the 1990's, recently the latest ideas from ML been ported into cosmology (see BH on the arXiv).

We can measure the properties of billions of galaxies from these types of images.



Use machine learning to approximate the mapping "f":

redshift = f(photometric properties of training sample)

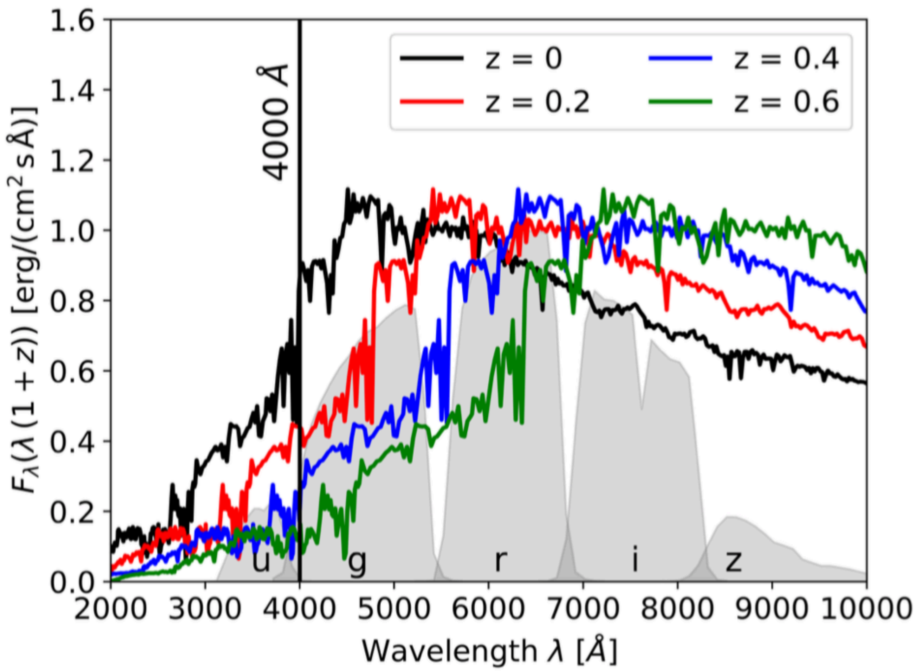
f(photometric properties of 3 billion galaxies) => photometric redshift

Photometric redshifts for galaxies

Photometric redshifts have been the test bed of machine learning applications in cosmology since the 1990's, recently the latest ideas from ML been ported into cosmology (see BH on the arXiv).

We can measure the properties of billions of galaxies from these types of images.

We can use a spectrograph to determine the redshift "truth values" of a subsample,.



Use machine learning to approximate the mapping "f":

redshift = f(photometric properties of training sample)

f(photometric properties of 3 billion galaxies) => photometric redshift

Overview

The supervised ML framework

An introduction to photometric redshifts

My typical ML workflow

A common ML application:

Photometric redshifts

The biggest problem for ML in cosmology:

Unrepresentative labelled data

Dealing with unrepresentative labelled data

Other common applications of ML

Recent, novel applications of ML

Conclusions

My Supervised Machine learning workflow

Examine the training / test / science sample data.
Is the test data representative of the science sample data?

Feature generation.
What has been used before, can we include it?

Feature pre-selection / feature importance
Random Forests / M.I.N.T. (see He et al 2013)

Training
Use heaps of algorithms & randomly explore hyper-parameter space.
Don’t have a favourite algorithm (mine is AdaBoost!). Use as benchmark.

Calibrate predictions
Do ML “pdfs” have the statistical properties of pdfs.

Application:
Apply calibrated predictions to “test” sample to estimate predictive power.
Apply calibrated predictions to “science sample” use this for science analysis.

Overview

The supervised ML framework

An introduction to photometric redshifts

My typical ML workflow

A common ML application:

Photometric redshifts

The biggest problem for ML in cosmology:

Unrepresentative labelled data

Dealing with unrepresentative labelled data

Other common applications of ML

Recent, novel applications of ML

Conclusions

Feature Extraction: Astrophysics example

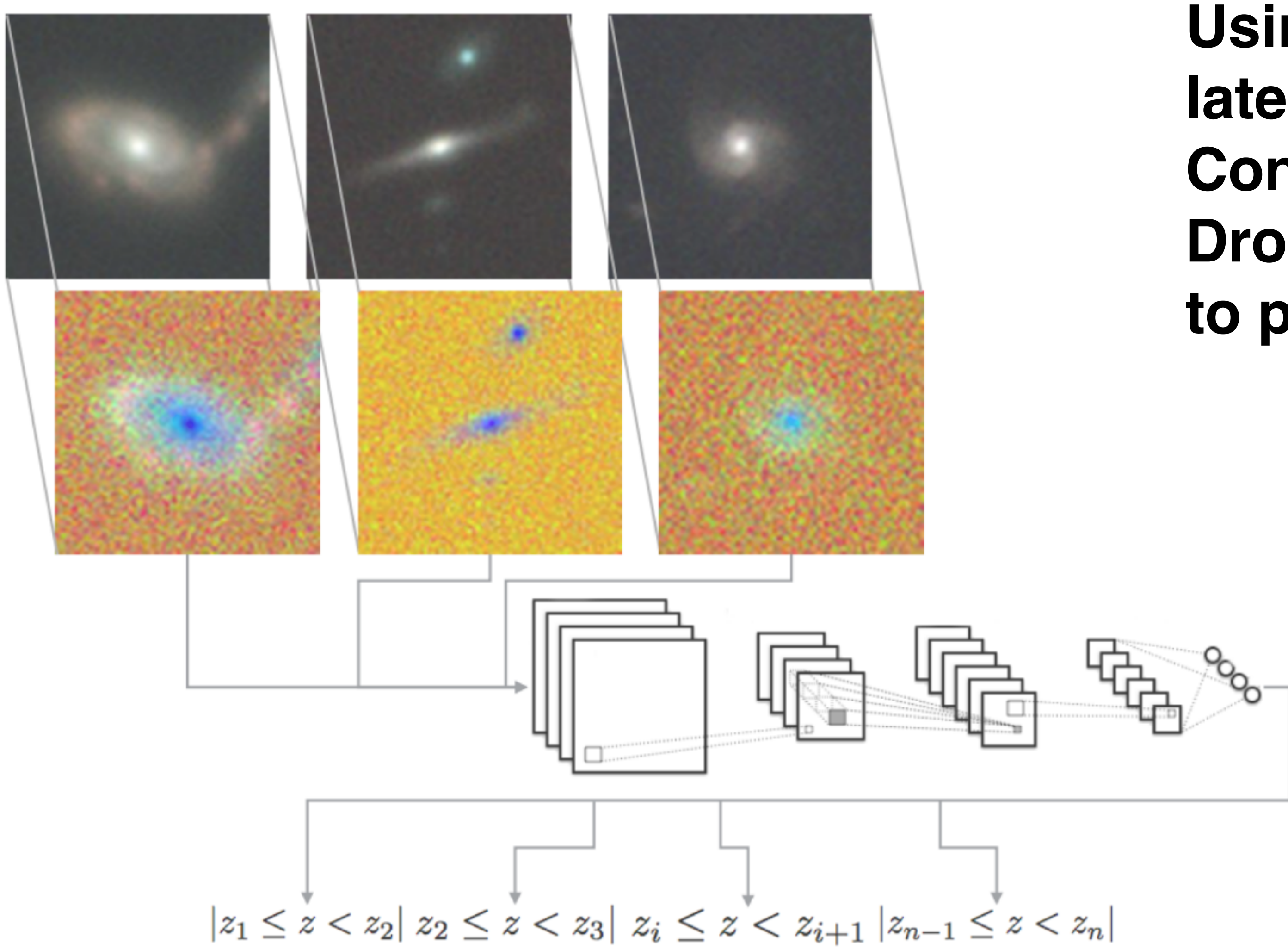
Broad band photometric images are easy and cheap, but accurate distances are required to use these objects for science.

Feature "extraction", catalogue level data: => standard machine learning

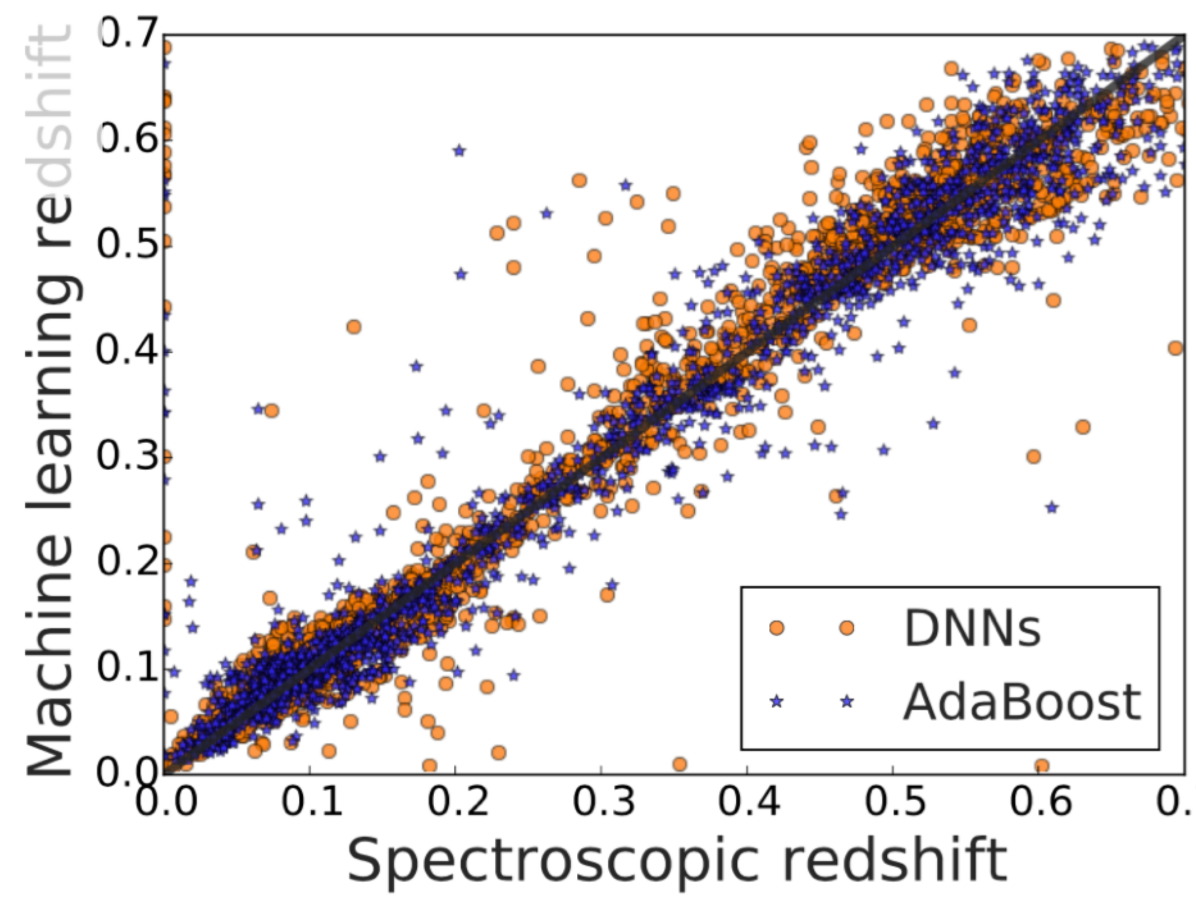
Description	Feature name
Magnitudes	dered_u dered_g dered_r dered_i dered_z psfMag_u psfMag_g psfMag_r psfMag_i psfMag_z fiberMag_u fiberMag_g fiberMag_r fiberMag_i fiberMag_z
Radii	petroRad_u petroRad_g petroRad_r petroRad_i petroRad_z expRad_u expRad_g expRad_r expRad_i expRad_z deVRad_u deVRad_g deVRad_r deVRad_i deVRad_z

Profile	fracDeV_u fracDeV_g fracDeV_r fracDeV_i fracDeV_z
Ellipticity	expAB_u expAB_g expAB_r expAB_i expAB_z deVAB_u deVAB_g deVAB_r deVAB_i deVAB_z
Means Stokes	q_u u_u q_g u_g q_r u_r q_i u_i q_z u_z

Deep learning using the galaxy image

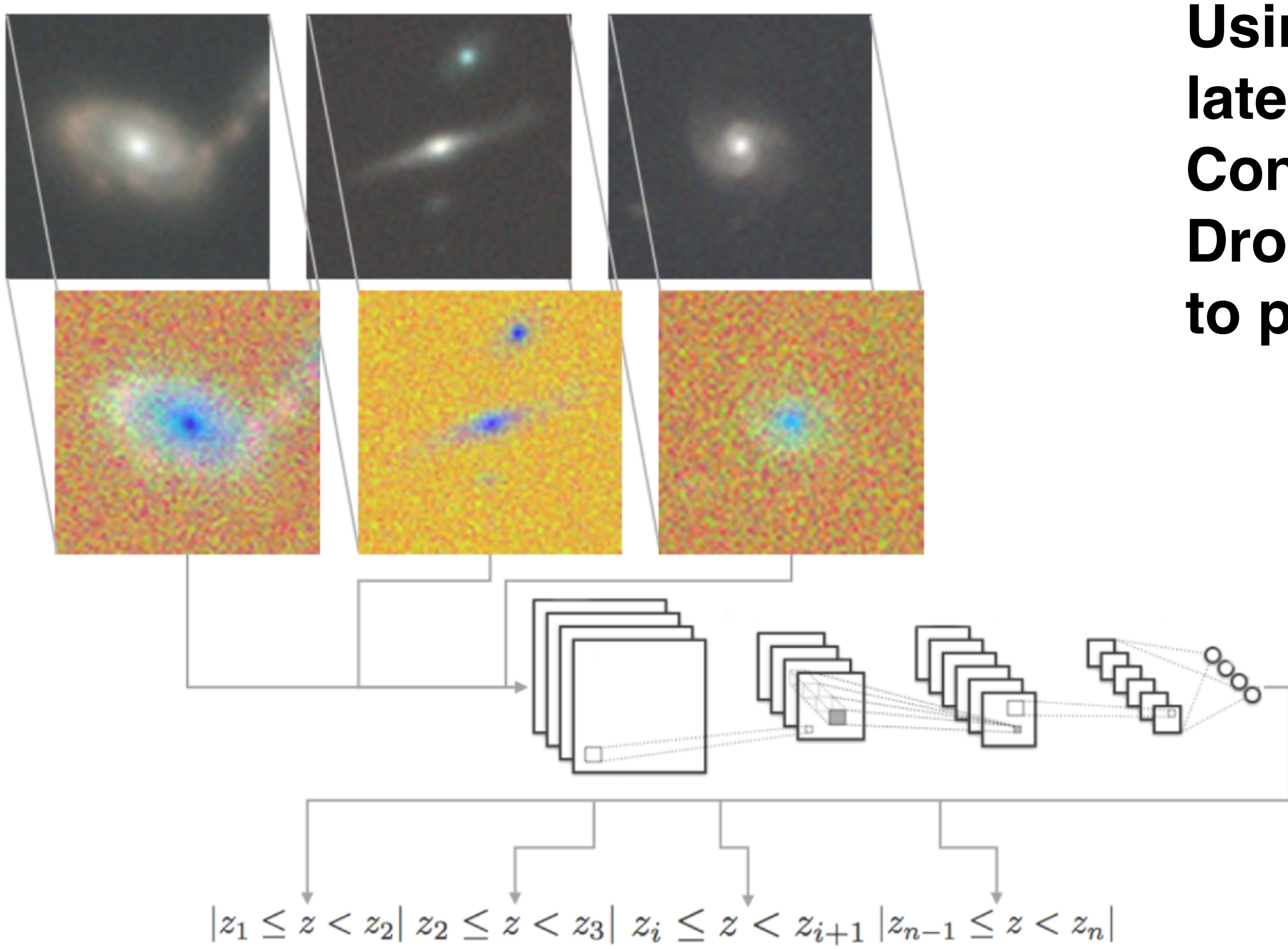


Using Deep Neural Networks (DNNs) with latest tricks from Computer science, e.g., Convolutional Neural Networks (ImageNet), Dropout, Maxout, applied to galaxy images to predict redshifts.

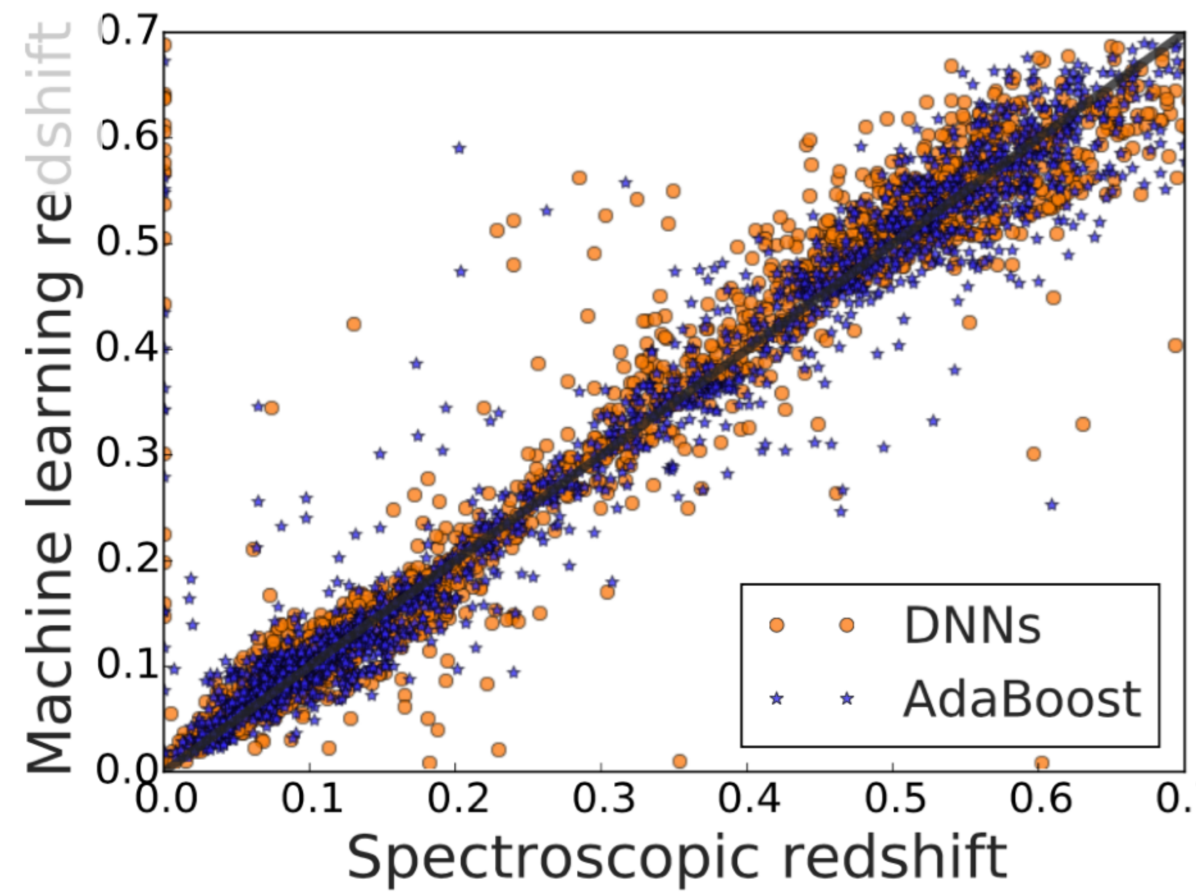


Hoyle 2015

Deep learning using the galaxy image



Using Deep Neural Networks (DNNs) with latest tricks from Computer science, e.g., Convolutional Neural Networks (ImageNet), Dropout, Maxout, applied to galaxy images to predict redshifts.



Use the full image, and "learn the features on the fly". Beware that the features become difficult to interpret.

Hoyle 2015

My Supervised Machine learning workflow

Examine the training / test / science sample data.
Is the test data representative of the science sample data?

Feature generation.
What has been used before, can we include it?

Feature pre-selection / feature importance
Random Forests / M.I.N.T (see He et al 2013)

Training
Use heaps of algorithms & randomly explore hyper-parameter space.
Don't have a favourite algorithm (mine is AdaBoost!). Use as benchmark.

Calibrate predictions
Do ML "pdfs" have the statistical properties of pdfs.

Application:
Apply calibrated predictions to "test" sample to estimate predictive power.
Apply calibrated predictions to "science sample" use this for science analysis.

Feature generation

However many more input ‘features’ are available. If we were really data driven, we’d go (a bit) crazy.

Unintelligent Feature generation

Use all features we can imagine! No pre selection (other than physically motivated pruning).

+ Linear combinations of features.
  This is what a standard Feed Forward NN could learn. However NN’s don’t like uninformative features, or too many features.

+ PCA, kernel PCA
  This maximises the “information content” of the system. Add these to the feature list.

Intelligent feature generation

+ Any domain specific knowledge
— what is currently being used in the field, and can it be included as an additional input feature? This means (depending on the algorithm) that your predictions will now be at least as good as the current standard.

My Supervised Machine learning workflow

Examine the training / test / science sample data.
Is the test data representative of the science sample data?

Feature generation.
What has been used before, can we include it?

Feature pre-selection / feature importance
Random Forests / M.I.N.T.

Training
Use heaps of algorithms & randomly explore hyper-parameter space.
Don't have a favourite algorithm (mine is AdaBoost!). Use as benchmark.

Calibrate predictions
Do ML "pdfs" have the statistical properties of pdfs.

Application:
Apply calibrated predictions to "test" sample to estimate predictive power.
Apply calibrated predictions to "science sample" use this for science analysis.

Feature pre-selection

We are now swimming in ‘M’ input features. Most algorithms don’t work well with many 10’s or 100’s of input features. Which ’S’ of those ‘M’ features should we use?

Feature importance with Decision Trees/ Random Forests

Machine learning techniques determine which input features provide the most predictive power when estimating ‘targets’. These more-important features can then be used in the algorithms of choice.

Feature pre-selection

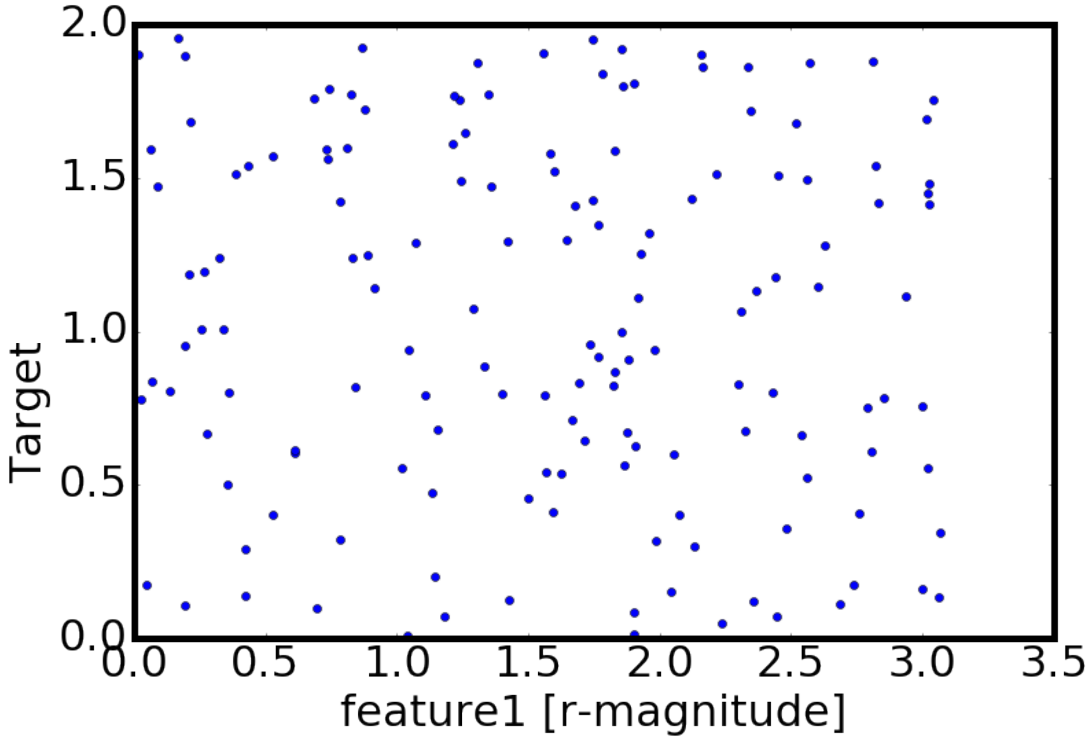
We are now swimming in ‘M’ input features. Most algorithms don’t work well with many 10’s or 100’s of input features. Which ’S’ of those ‘M’ features should we use?

Feature importance with Decision Trees/ Random Forests

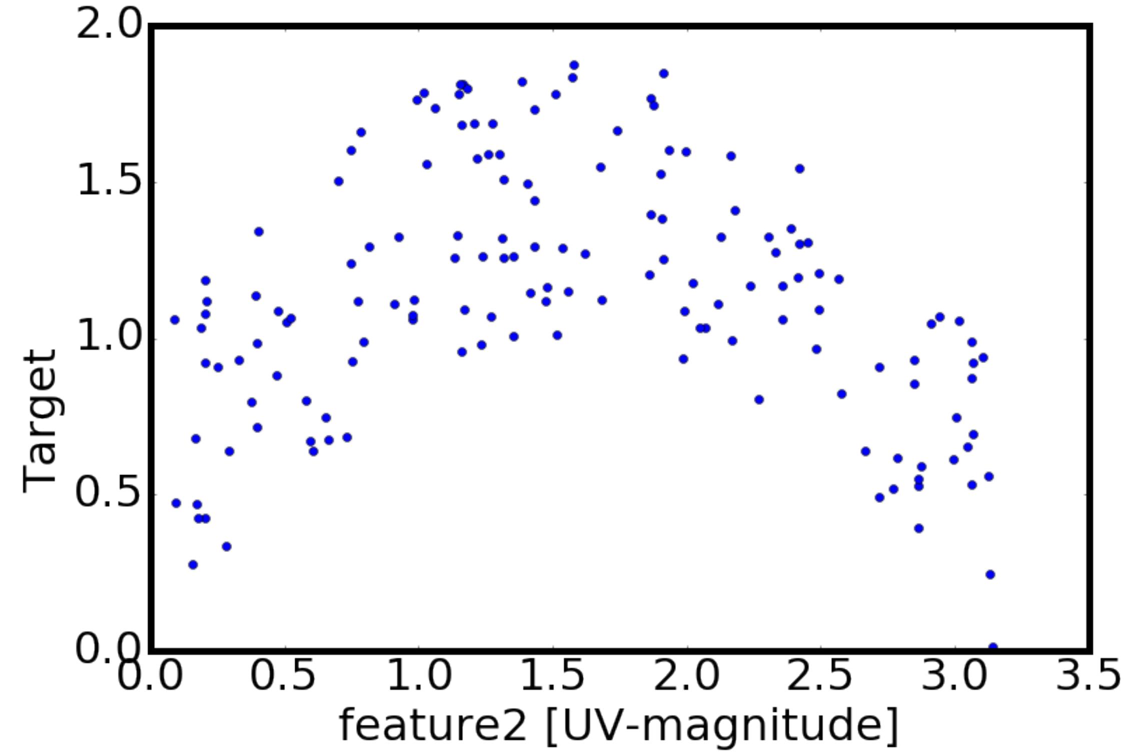
Machine learning techniques determine which input features provide the most predictive power when estimating ‘targets’. These more-important features can then be used in the algorithms of choice.

How can feature importance be calculated using the training data?

$$\sigma(target)$$



$$\sigma(target)$$



Feature pre-selection

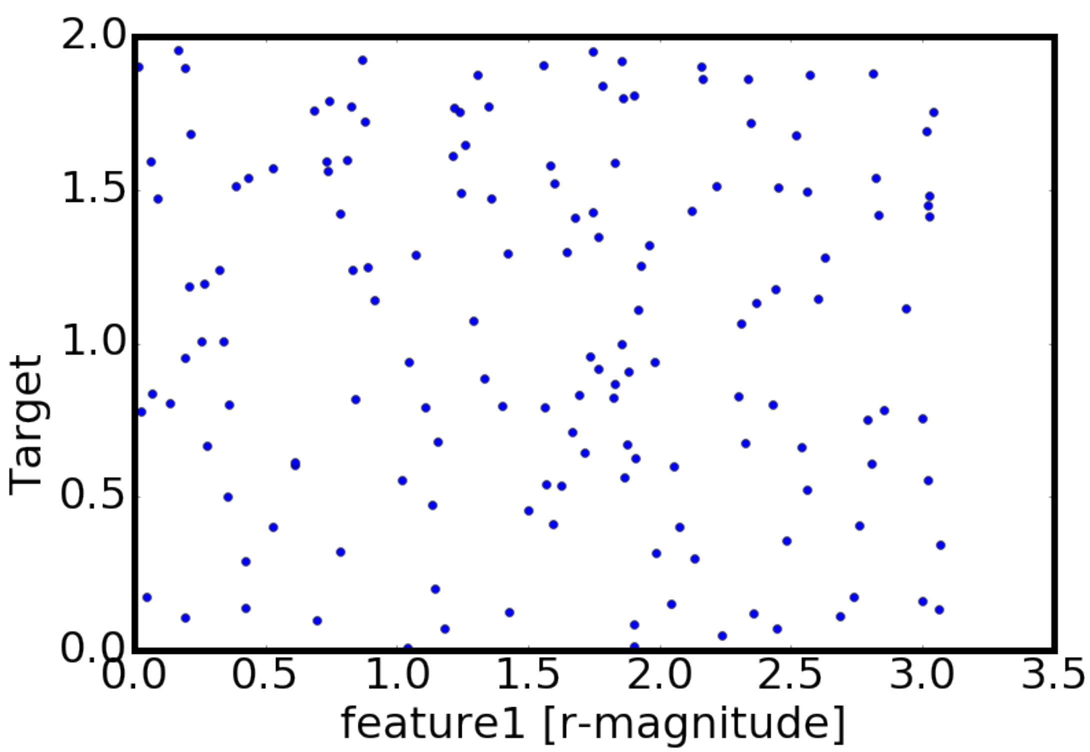
We are now swimming in ‘M’ input features. Most algorithms don’t work well with many 10’s or 100’s of input features. Which ’S’ of those ‘M’ features should we use?

Feature importance with Decision Trees/ Random Forests

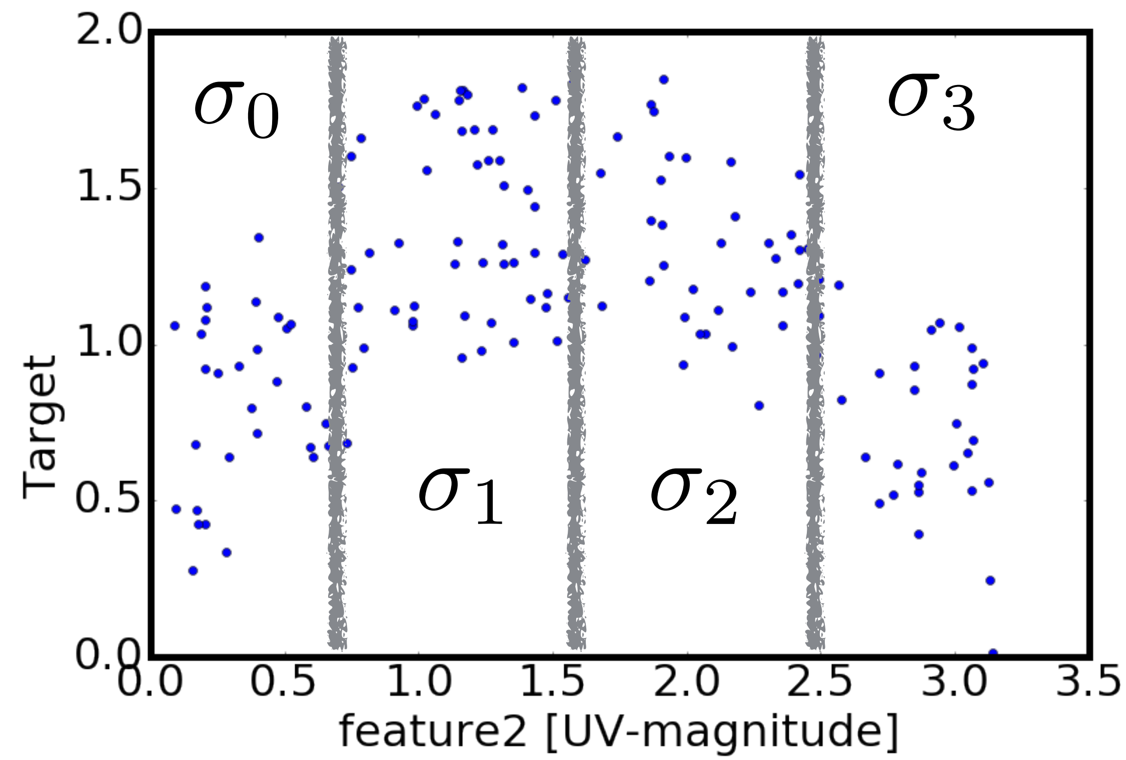
Machine learning techniques determine which input features provide the most predictive power when estimating ‘targets’. These more-important features can then be used in the algorithms of choice.

How can feature importance be calculated using the training data?

$\sigma(target)$



$\sigma_N << \sigma(target)$



Gini criteria: Reduce the scatter of the target feature by sub dividing the input feature space. The more slices in a feature dimension, which each time reduce the Gini index, the more that feature dimension has predictive power. In practice this happens in a high dimensional hyper-cube.

Feature pre-selection

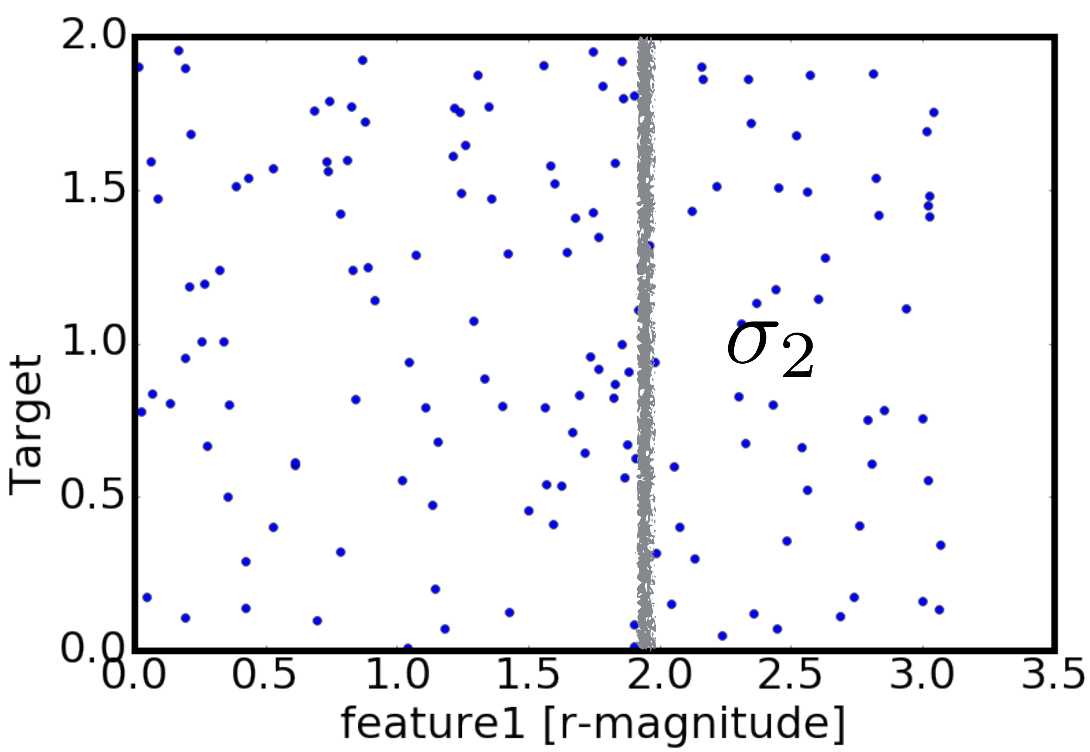
We are now swimming in 'M' input features. Most algorithms don't work well with many 10's or 100's of input features. Which 'S' of those 'M' features should we use?

Feature importance with Decision Trees/ Random Forests

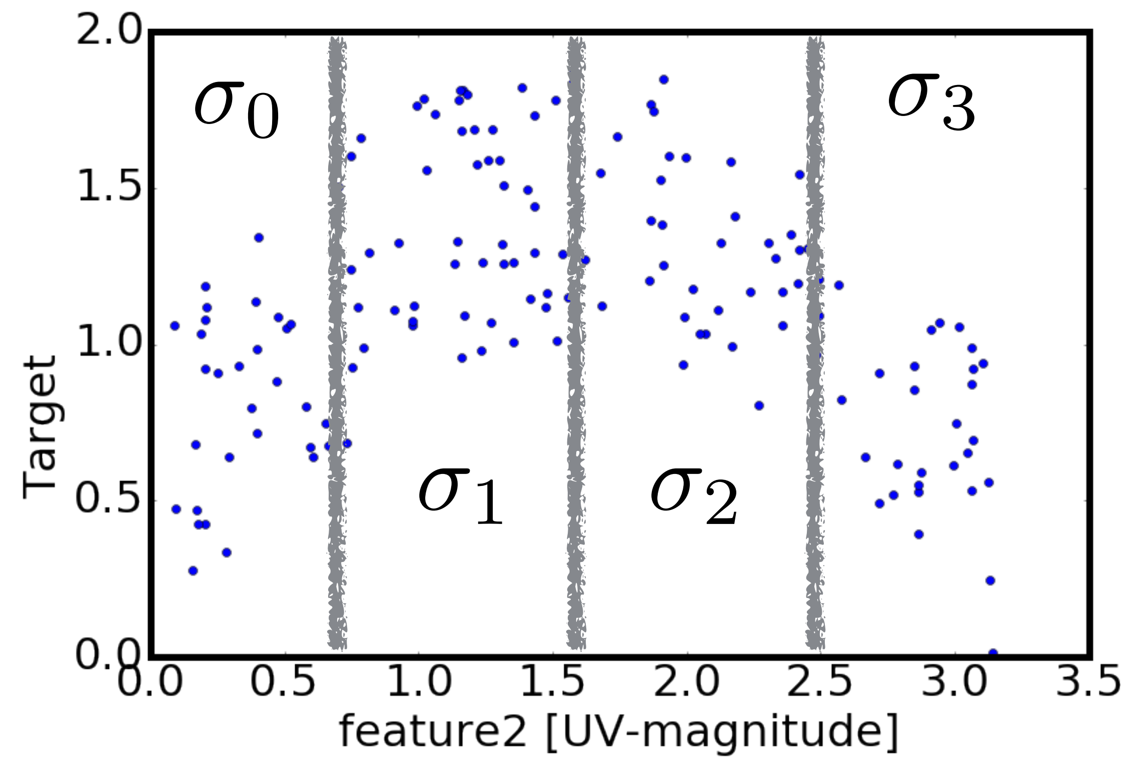
Machine learning techniques determine which input features provide the most predictive power when estimating 'targets'. These more-important features can then be used in the algorithms of choice.

How can feature importance be calculated using the training data?

$$\sigma_N \sim \sigma(target)$$



$$\sigma_N << \sigma(target)$$



Gini criteria: Reduce the scatter of the target feature by sub dividing the input feature space. The more slices in a feature dimension, which each time reduce the Gini index, the more that feature dimension has predictive power. In practice this happens in a high dimensional hyper-cube.

Feature pre-selection

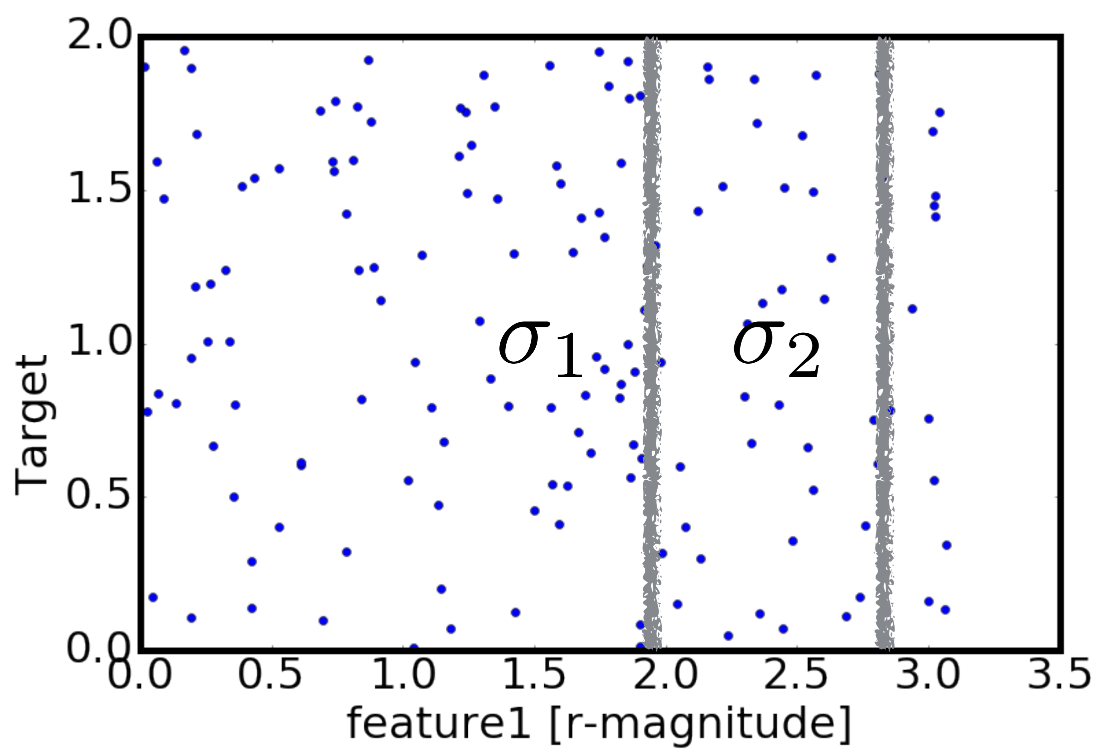
We are now swimming in 'M' input features. Most algorithms don't work well with many 10's or 100's of input features. Which 'S' of those 'M' features should we use?

Feature importance with Decision Trees/ Random Forests

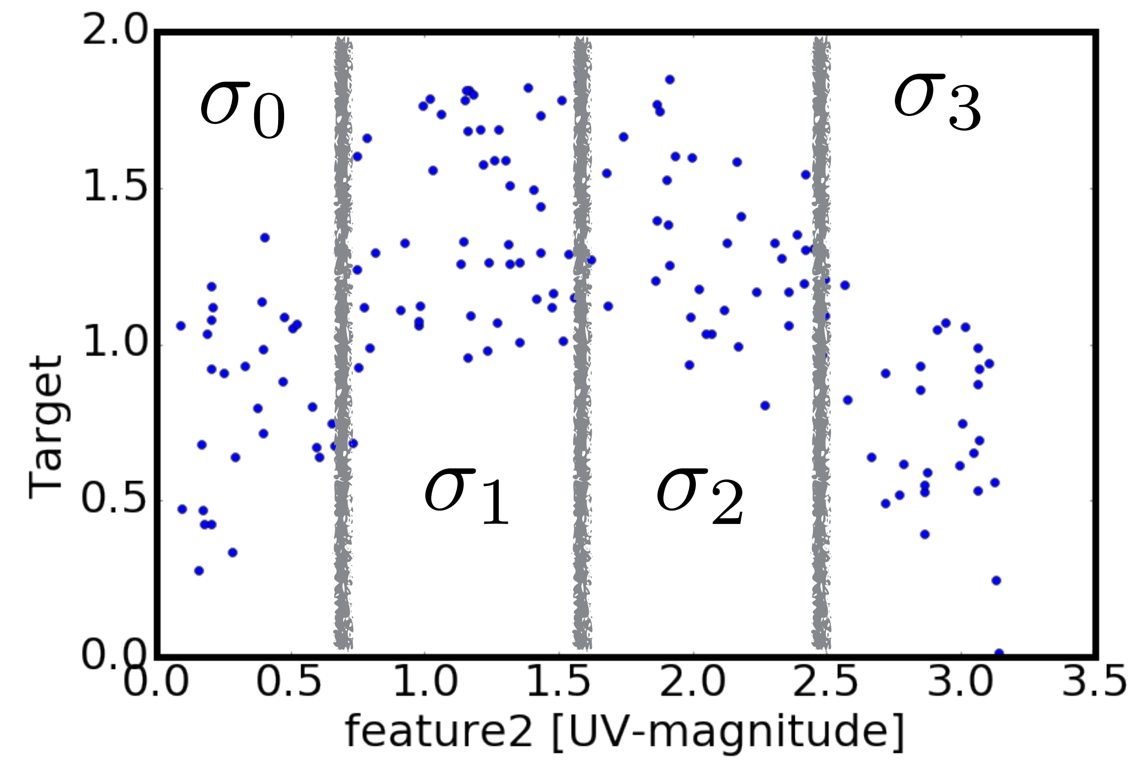
Machine learning techniques determine which input features provide the most predictive power when estimating 'targets'. These more-important features can then be used in the algorithms of choice.

How can feature importance be calculated using the training data?

$$\sigma_N \sim \sigma(target)$$



$$\sigma_N << \sigma(target)$$



Gini criteria: Reduce the scatter of the target feature by sub dividing the input feature space. The more slices in a feature dimension, which each time reduce the Gini index, the more that feature dimension has predictive power. In practice this happens in a high dimensional hyper-cube.

Feature pre-selection

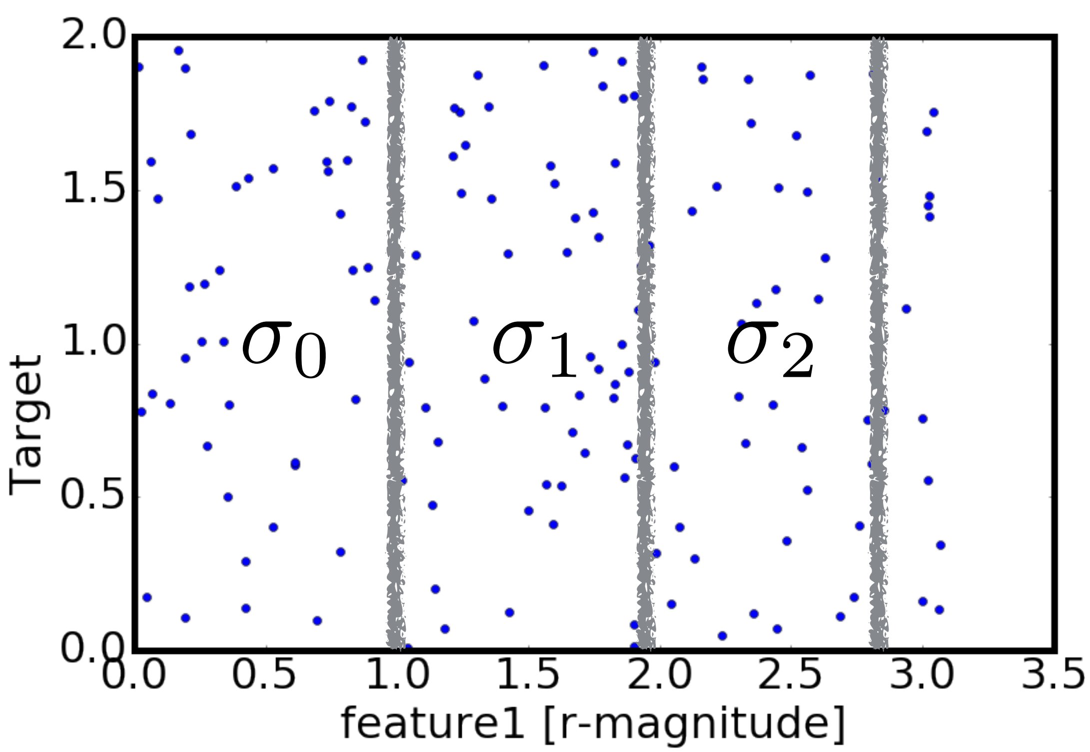
We are now swimming in 'M' input features. Most algorithms don't work well with many 10's or 100's of input features. Which 'S' of those 'M' features should we use?

Feature importance with Decision Trees/ Random Forests

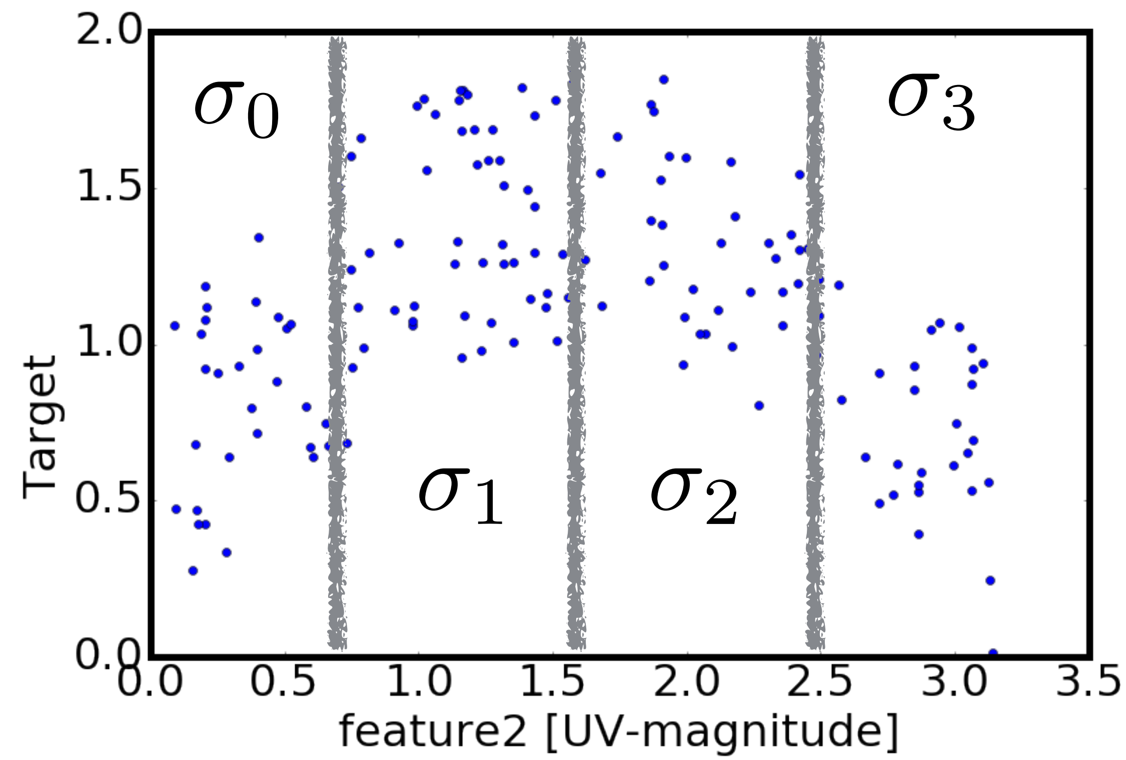
Machine learning techniques determine which input features provide the most predictive power when estimating 'targets'. These more-important features can then be used in the algorithms of choice.

How can feature importance be calculated using the training data?

$$\sigma_N \sim \sigma(target)$$



$$\sigma_N << \sigma(target)$$



Gini criteria: Reduce the scatter of the target feature by sub dividing the input feature space. The more slices in a feature dimension, which each time reduce the Gini index, the more that feature dimension has predictive power. In practice this happens in a high dimensional hyper-cube.

Feature importance applied to photo-z

Description	Feature name
Magnitudes	dered_u dered_g dered_r dered_i dered_z psfMag_u psfMag_g psfMag_r psfMag_i psfMag_z fiberMag_u fiberMag_g fiberMag_r fiberMag_i fiberMag_z
Radii	petroRad_u petroRad_g petroRad_r petroRad_i petroRad_z expRad_u expRad_g expRad_r expRad_i expRad_z deVRad_u deVRad_g deVRad_r deVRad_i deVRad_z
Colors	dered_z-dered_i dered_z-dered_r dered_z-dered_g dered_z-dered_u dered_i-dered_r dered_i-dered_g dered_i-dered_u dered_r-dered_g dered_r-dered_u dered_g-dered_u fiberMag_z-fiberMag_i fiberMag_z-fiberMag_r fiberMag_z-fiberMag_g fiberMag_z-fiberMag_u fiberMag_i-fiberMag_r fiberMag_i-fiberMag_g fiberMag_i-fiberMag_u fiberMag_r-fiberMag_g fiberMag_r-fiberMag_u fiberMag_g-fiberMag_u psfMag_z-psfMag_i psfMag_z-psfMag_r psfMag_z-psfMag_g psfMag_z-psfMag_u psfMag_i-psfMag_r psfMag_i-psfMag_g psfMag_i-psfMag_u psfMag_r-psfMag_g psfMag_r-psfMag_u psfMag_g-psfMag_u
Profile	fracDeV_u fracDeV_g fracDeV_r fracDeV_i fracDeV_z
Ellipticity	expAB_u expAB_g expAB_r expAB_i expAB_z deVAB_u deVAB_g deVAB_r deVAB_i deVAB_z
Means Stokes	q_u u_u q_g u_g q_r u_r q_i u_i q_z u_z

arXiv:1410.4696 [pdf, other]

Feature importance for machine learning redshifts applied to SDSS galaxies

Ben Hoyle, Markus Michael Rau, Roman Zitlau, Stella Seitz, Jochen Weller

Comments: 10 pages, 4 figures, updated to match version accepted in MNRAS

Subjects: **Instrumentation and Methods for Astrophysics (astro-ph.IM)**; Cosmology and Nongalactic Astrophysics (astro-ph.CO)

Feature importance applied to photo-z

Description	Feature name
Magnitudes	dered_u dered_g dered_r dered_i dered_z psfMag_u psfMag_g psfMag_r psfMag_i psfMag_z fiberMag_u fiberMag_g fiberMag_r fiberMag_i fiberMag_z
Radii	petroRad_u petroRad_g petroRad_r petroRad_i petroRad_z expRad_u expRad_g expRad_r expRad_i expRad_z deVRad_u deVRad_g deVRad_r deVRad_i deVRad_z
Colors	dered_z-dered_i dered_z-dered_r dered_z-dered_g dered_z-dered_u dered_i-dered_r dered_i-dered_g dered_i-dered_u dered_r-dered_g dered_r-dered_u dered_g-dered_u fiberMag_z-fiberMag_i fiberMag_z-fiberMag_r fiberMag_z-fiberMag_g fiberMag_z-fiberMag_u fiberMag_i-fiberMag_r fiberMag_i-fiberMag_g fiberMag_i-fiberMag_u fiberMag_r-fiberMag_g fiberMag_r-fiberMag_u fiberMag_g-fiberMag_u psfMag_z-psfMag_i psfMag_z-psfMag_r psfMag_z-psfMag_g psfMag_z-psfMag_u psfMag_i-psfMag_r psfMag_i-psfMag_g psfMag_i-psfMag_u psfMag_r-psfMag_g psfMag_r-psfMag_u psfMag_g-psfMag_u
Profile	fracDeV_u fracDeV_g fracDeV_r fracDeV_i fracDeV_z
Ellipticity	expAB_u expAB_g expAB_r expAB_i expAB_z deVAB_u deVAB_g deVAB_r deVAB_i deVAB_z
Means Stokes	q_u u_u q_g u_g q_r u_r q_i u_i q_z u_z

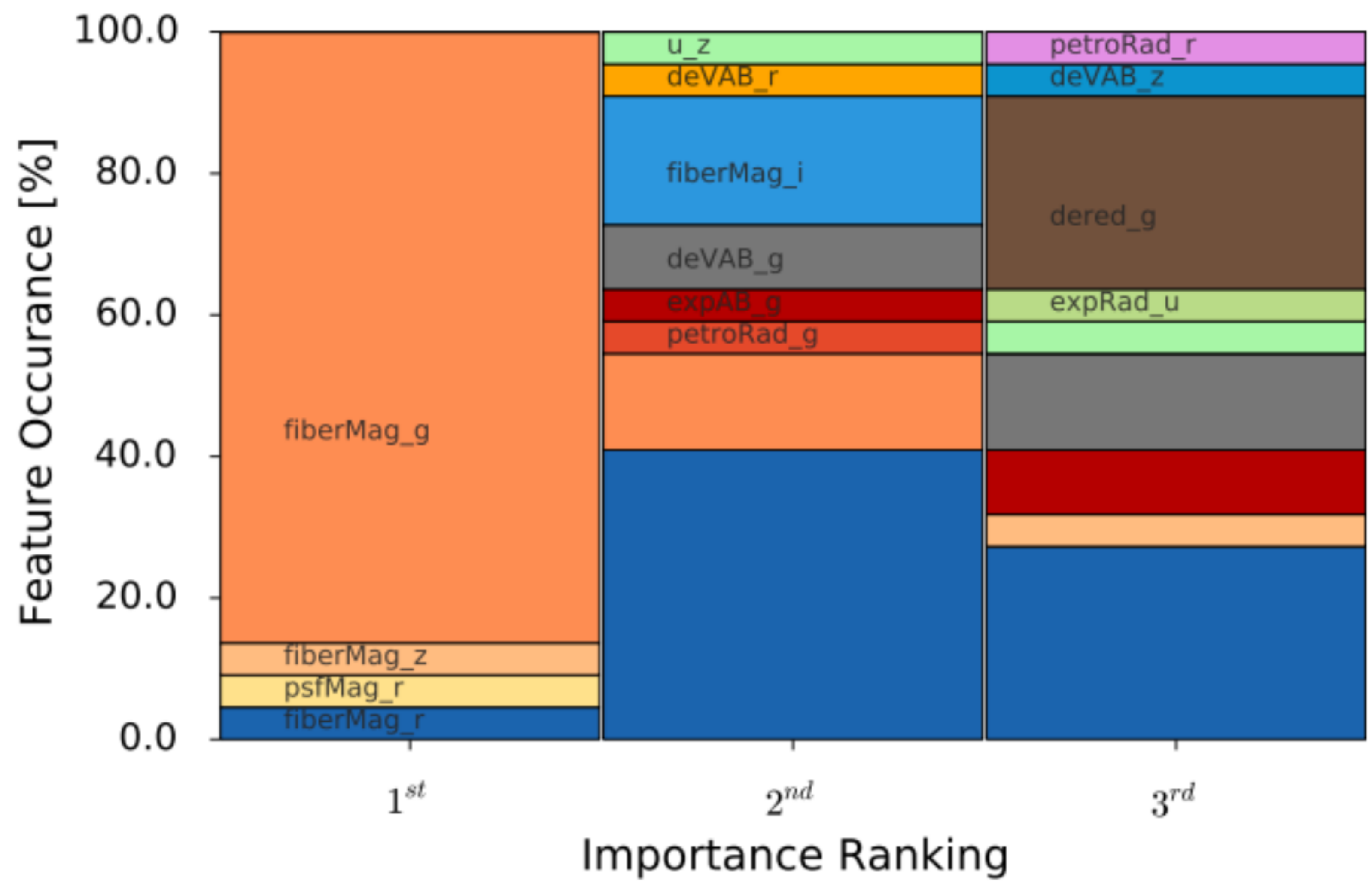
arXiv:1410.4696 [pdf, other]

Feature importance for machine learning redshifts applied to SDSS galaxies

Ben Hoyle, Markus Michael Rau, Roman Zitlau, Stella Seitz, Jochen Weller

Comments: 10 pages, 4 figures, updated to match version accepted in MNRAS

Subjects: **Instrumentation and Methods for Astrophysics (astro-ph.IM)**; Cosmology and Nongalactic Astrophysics (astro-ph.CO)



Feature importance applied to photo-z

Description	Feature name
Magnitudes	dered_u dered_g dered_r dered_i dered_z psfMag_u psfMag_g psfMag_r psfMag_i psfMag_z fiberMag_u fiberMag_g fiberMag_r fiberMag_i fiberMag_z
Radii	petroRad_u petroRad_g petroRad_r petroRad_i petroRad_z expRad_u expRad_g expRad_r expRad_i expRad_z deVRad_u deVRad_g deVRad_r deVRad_i deVRad_z
Colors	dered_z-dered_i dered_z-dered_r dered_z-dered_g dered_z-dered_u dered_i-dered_r dered_i-dered_g dered_i-dered_u dered_r-dered_g dered_r-dered_u dered_g-dered_u fiberMag_z-fiberMag_i fiberMag_z-fiberMag_r fiberMag_z-fiberMag_g fiberMag_z-fiberMag_u fiberMag_i-fiberMag_r fiberMag_i-fiberMag_g fiberMag_i-fiberMag_u fiberMag_r-fiberMag_g fiberMag_r-fiberMag_u fiberMag_g-fiberMag_u psfMag_z-psfMag_i psfMag_z-psfMag_r psfMag_z-psfMag_g psfMag_z-psfMag_u psfMag_i-psfMag_r psfMag_i-psfMag_g psfMag_i-psfMag_u psfMag_r-psfMag_g psfMag_r-psfMag_u psfMag_g-psfMag_u
Profile	fracDeV_u fracDeV_g fracDeV_r fracDeV_i fracDeV_z
Ellipticity	expAB_u expAB_g expAB_r expAB_i expAB_z deVAB_u deVAB_g deVAB_r deVAB_i deVAB_z
Means Stokes	q_u u_u q_g u_g q_r u_r q_i u_i q_z u_z

arXiv:1410.4696 [pdf, other]

Feature importance for machine learning redshifts applied to SDSS galaxies

Ben Hoyle, Markus Michael Rau, Roman Zitlau, Stella Seitz, Jochen Weller

Comments: 10 pages, 4 figures, updated to match version accepted in MNRAS

Subjects: **Instrumentation and Methods for Astrophysics (astro-ph.IM)**; Cosmology and Nongalactic Astrophysics (astro-ph.CO)

A catalogue of photometric redshifts for the SDSS-DR9 galaxies *(Research Note)*

M. Brescia[1], S. Cavuoti[1], G. Longo[2], and V. De Stefano[2]

a two hidden layer network, using a combination of the 4 SDSS colors (obtained from the SDSS *psfMag*) plus the pivot magnitude *psfMag* in the *r* band. This gi uncertainty of σ = 0.023 with a ver $\sim 3\times10^{-5}$, a low *NMAD*, and to a low f

PhotoRaptor

Feature importance applied to photo-z

Description	Feature name
Magnitudes	dered_u dered_g dered_r dered_i dered_z psfMag_u psfMag_g psfMag_r psfMag_i psfMag_z fiberMag_u fiberMag_g fiberMag_r fiberMag_i fiberMag_z
Radii	petroRad_u petroRad_g petroRad_r petroRad_i petroRad_z expRad_u expRad_g expRad_r expRad_i expRad_z deVRad_u deVRad_g deVRad_r deVRad_i deVRad_z
Colors	dered_z-dered_i dered_z-dered_r dered_z-dered_g dered_z-dered_u dered_i-dered_r dered_i-dered_g dered_i-dered_u dered_r-dered_g dered_r-dered_u dered_g-dered_u fiberMag_z-fiberMag_i fiberMag_z-fiberMag_r fiberMag_z-fiberMag_g fiberMag_z-fiberMag_u fiberMag_i-fiberMag_r fiberMag_i-fiberMag_g fiberMag_i-fiberMag_u fiberMag_r-fiberMag_g fiberMag_r-fiberMag_u fiberMag_g-fiberMag_u psfMag_z-psfMag_i psfMag_z-psfMag_r psfMag_z-psfMag_g psfMag_z-psfMag_u psfMag_i-psfMag_r psfMag_i-psfMag_g psfMag_i-psfMag_u psfMag_r-psfMag_g psfMag_r-psfMag_u psfMag_g-psfMag_u
Profile	fracDeV_u fracDeV_g fracDeV_r fracDeV_i fracDeV_z
Ellipticity	expAB_u expAB_g expAB_r expAB_i expAB_z deVAB_u deVAB_g deVAB_r deVAB_i deVAB_z
Means Stokes	q_u u_u q_g u_g q_r u_r q_i u_i q_z u_z

$$\Delta = z_{spec} - z_{predict}$$

arXiv:1410.4696 [pdf, other]

Feature importance for machine learning redshifts applied to SDSS galaxies

Ben Hoyle, Markus Michael Rau, Roman Zitlau, Stella Seitz, Jochen Weller

Comments: 10 pages, 4 figures, updated to match version accepted in MNRAS

Subjects: **Instrumentation and Methods for Astrophysics (astro-ph.IM)**; Cosmology and Nongalactic Astrophysics (astro-ph.CO)

A catalogue of photometric redshifts for the SDSS-DR9 galaxies *(Research Note)*

M. Brescia[1], S. Cavuoti[1], G. Longo[2], and V. De Stefano[2]

a two hidden layer network, using a combination of the 4 SDSS colors (obtained from the SDSS *psfMag*) plus the pivot magnitude *psfMag* in the *r* band. This gi[...] uncertainty of $\sigma = 0.023$ with a ver[...] $\sim 3\times10^{-5}$, a low *NMAD*, and to a low f[...]

PhotoRaptor

Input features	$\mu_{\Delta_z} \pm \sigma_{\Delta_z}$
PSF mag/cols	-0.0001±0.075
Top1&2	
Standard&top1&2	
Standard&top1&2&3	

Feature importance applied to photo-z

Description	Feature name
Magnitudes	dered_u dered_g dered_r dered_i dered_z psfMag_u psfMag_g psfMag_r psfMag_i psfMag_z fiberMag_u fiberMag_g fiberMag_r fiberMag_i fiberMag_z
Radii	petroRad_u petroRad_g petroRad_r petroRad_i petroRad_z expRad_u expRad_g expRad_r expRad_i expRad_z deVRad_u deVRad_g deVRad_r deVRad_i deVRad_z
Colors	dered_z-dered_i dered_z-dered_r dered_z-dered_g dered_z-dered_u dered_i-dered_r dered_i-dered_g dered_i-dered_u dered_r-dered_g dered_r-dered_u dered_g-dered_u fiberMag_z-fiberMag_i fiberMag_z-fiberMag_r fiberMag_z-fiberMag_g fiberMag_z-fiberMag_u fiberMag_i-fiberMag_r fiberMag_i-fiberMag_g fiberMag_i-fiberMag_u fiberMag_r-fiberMag_g fiberMag_r-fiberMag_u fiberMag_g-fiberMag_u psfMag_z-psfMag_i psfMag_z-psfMag_r psfMag_z-psfMag_g psfMag_z-psfMag_u psfMag_i-psfMag_r psfMag_i-psfMag_g psfMag_i-psfMag_u psfMag_r-psfMag_g psfMag_r-psfMag_u psfMag_g-psfMag_u
Profile	fracDeV_u fracDeV_g fracDeV_r fracDeV_i fracDeV_z
Ellipticity	expAB_u expAB_g expAB_r expAB_i expAB_z deVAB_u deVAB_g deVAB_r deVAB_i deVAB_z
Means Stokes	q_u u_u q_g u_g q_r u_r q_i u_i q_z u_z

$$\Delta = z_{spec} - z_{predict}$$

arXiv:1410.4696 [pdf, other]

Feature importance for machine learning redshifts applied to SDSS galaxies

Ben Hoyle, Markus Michael Rau, Roman Zitlau, Stella Seitz, Jochen Weller

Comments: 10 pages, 4 figures, updated to match version accepted in MNRAS

Subjects: **Instrumentation and Methods for Astrophysics (astro-ph.IM)**; Cosmology and Nongalactic Astrophysics (astro-ph.CO)

A catalogue of photometric redshifts for the SDSS-DR9 galaxies *(Research Note)*

M. Brescia[1], S. Cavuoti[1], G. Longo[2], and V. De Stefano[2]

a two hidden layer network, using a combination of the 4 SDSS colors (obtained from the SDSS *psfMag*) plus the pivot magnitude *psfMag* in the *r* band. This gi
uncertainty of $\sigma = 0.023$ with a ver
$\sim 3\times10^{-5}$, a low *NMAD*, and to a low f

PhotoRaptor

Input features	$\mu_{\Delta_z} \pm \sigma_{\Delta_z}$
PSF mag/cols	-0.0001±0.075
Top1&2	0.0001±0.068
Standard&top1&2	-0.0001±0.066
Standard&top1&2&3	0.0001±0.065

My Supervised Machine learning workflow

Examine the training / test / science sample data.
Is the test data representative of the science sample data?

Feature generation.
What has been used before, can we include it?

Feature pre-selection / feature importance
Random Forests / M.I.N.T. (see He et al 2013)

Training
Use heaps of algorithms & randomly explore hyper-parameter space.
Don’t have a favourite algorithm (mine is AdaBoost!). Use as benchmark.

Calibrate predictions
Do ML “pdfs” have the statistical properties of pdfs.

Application:
Apply calibrated predictions to “test” sample to estimate predictive power.
Apply calibrated predictions to “science sample” use this for science analysis.

Calibrating the machine's predictions

ML codes produce (P)DFs "conditional" on the data and algorithm. We can normalise them. We often call them PDFs. I always call them "PDFs".

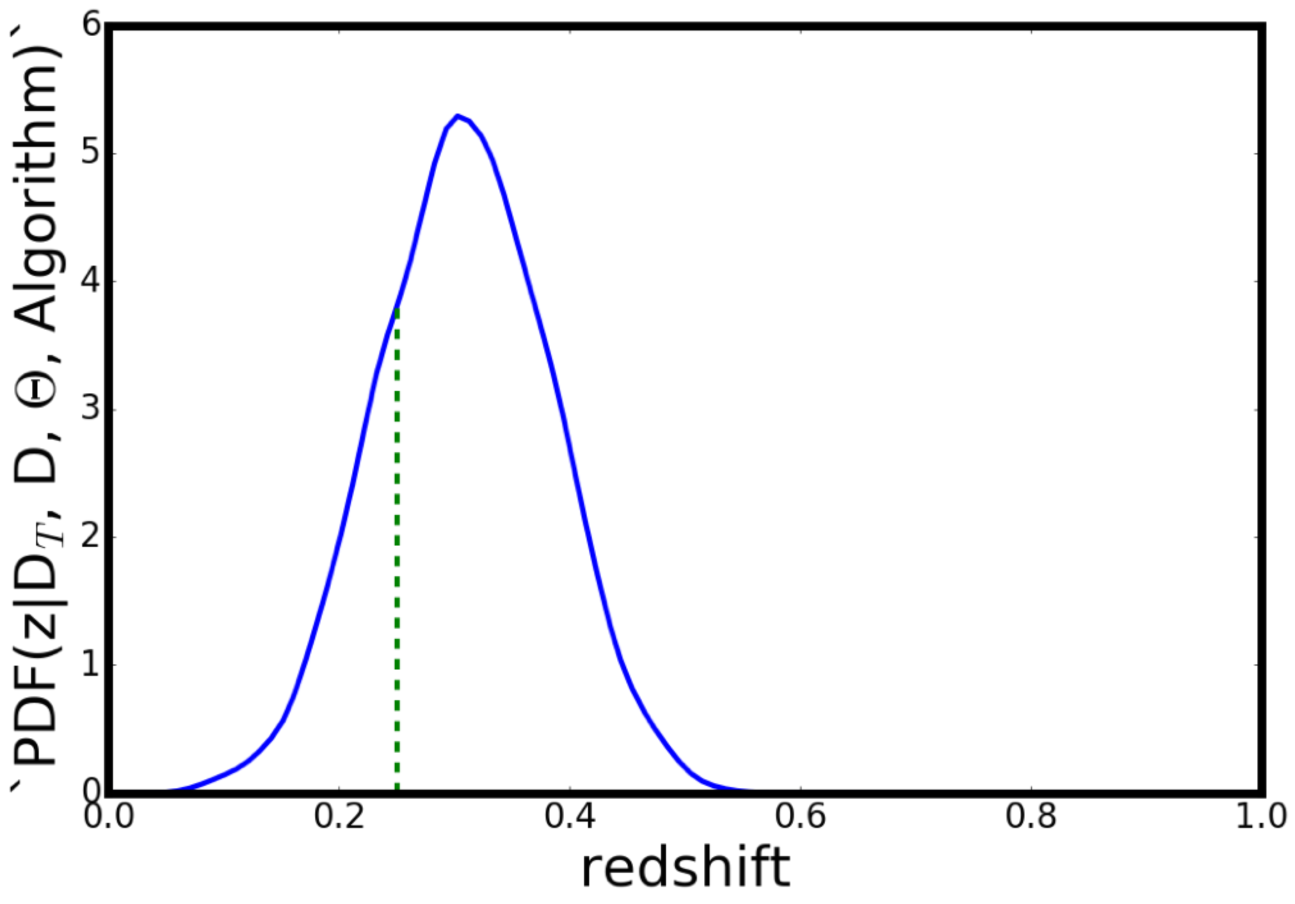
There are statistical methods to test if a PDF behaves like a PDF. [Brier 1950, Dawid 1984, Bordolio et al 2009, Polsterer 2016, BH et al 2017 —photo-z setting].

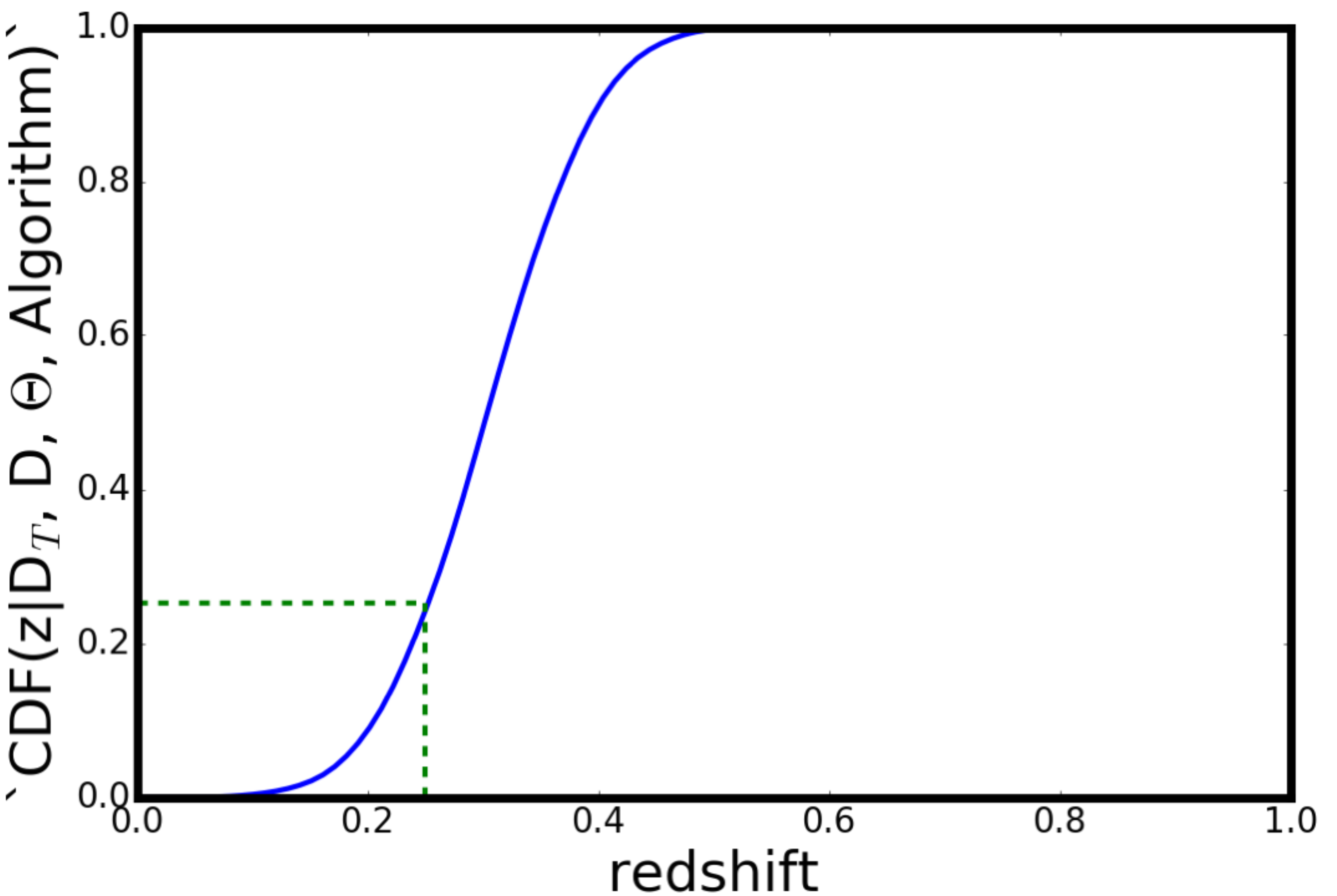
Calibrating the machine's predictions

ML codes produce (P)DFs "conditional" on the data and algorithm. We can normalise them. We often call them PDFs. I always call them "PDFs".

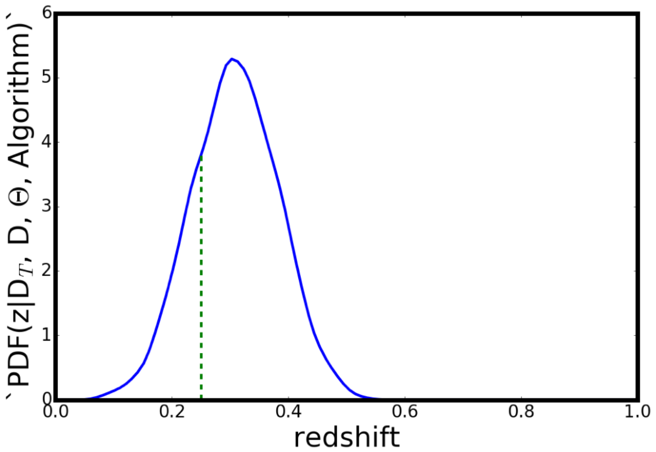
There are statistical methods to test if a PDF behaves like a PDF. [Brier 1950, Dawid 1984, Bordolio et al 2009, Polsterer 2016, BH et al 2017 —photo-z setting].

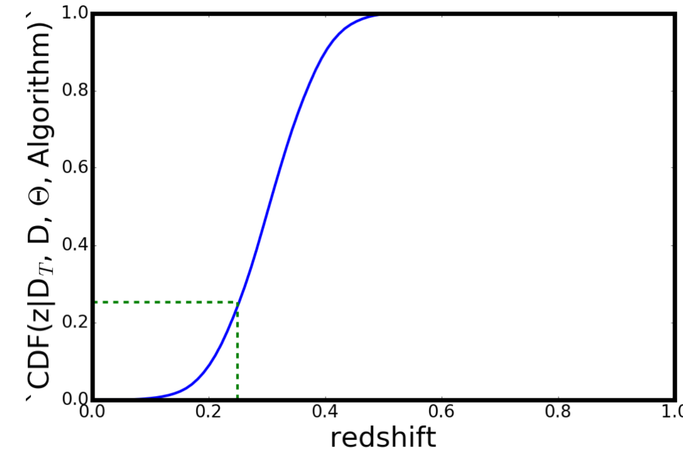
Reframe the question: If the pdf is correct, the truth value for one object should be consistent with being a random draw from the pdf. For an ensemble of objects we can measure the properties of the truth values w.r.t their pdfs. Probability Integral Transform (PIT).





Calibrating the machine's predictions





For an ensemble of objects, does the distributions of CDF values evaluated at the truth values have the expected shape.

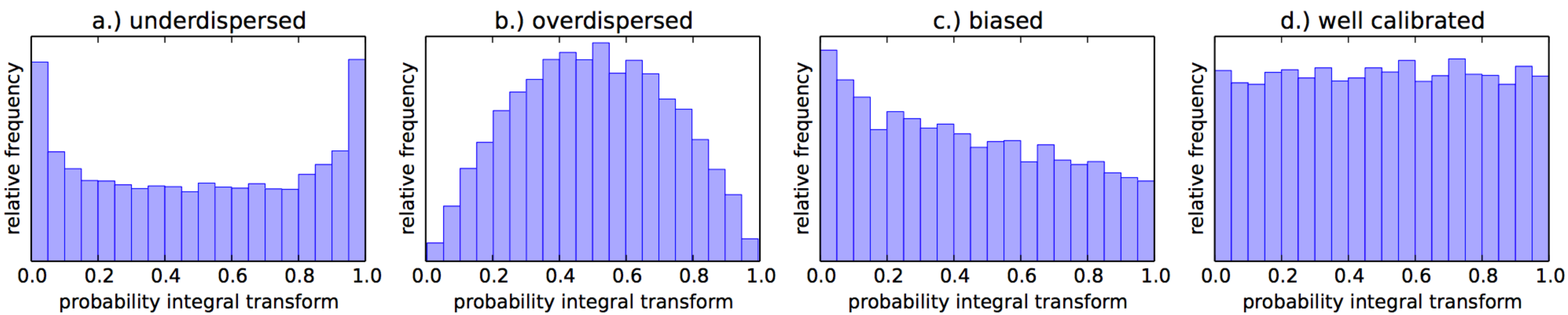
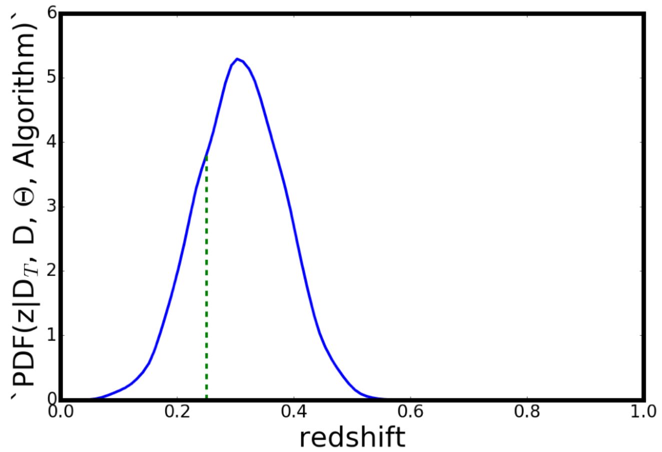
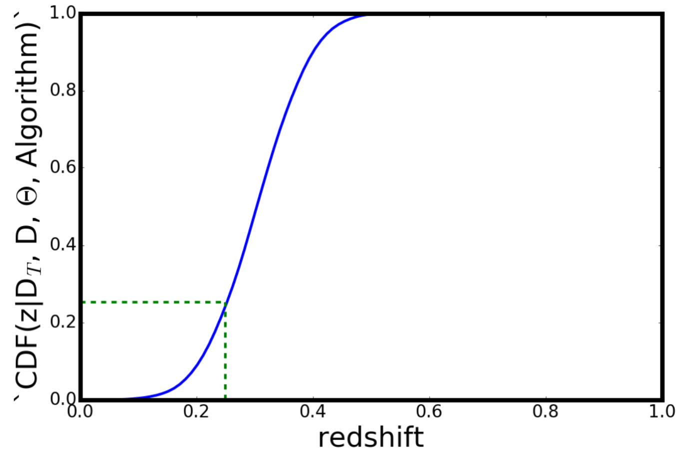


Figure 3. Four different probability integral transforms (PITs). In the case of underdispersed PDFs an u-shaped, concave distribution is observed (a). Overdispersed PDFs result in a peaked, convex distribution (b). When a slope in the PIT is observed, the analysed PDFs are biased (c). Only when the PIT exhibits a flat distribution, the PDFs are well calibrated (d).

Polsterer 1608.08016

Calibrating the machine's predictions





For an ensemble of objects, does the distributions of CDF values evaluated at the truth values have the expected shape.

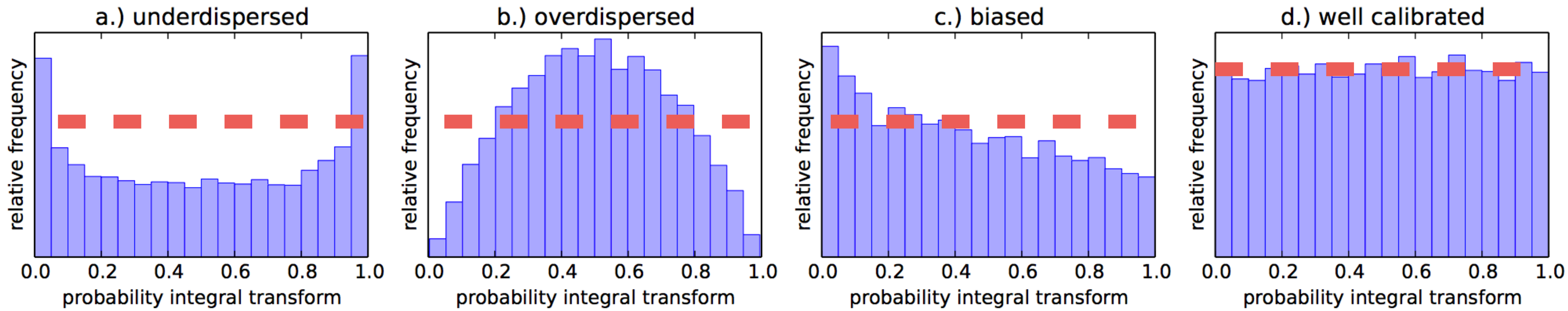
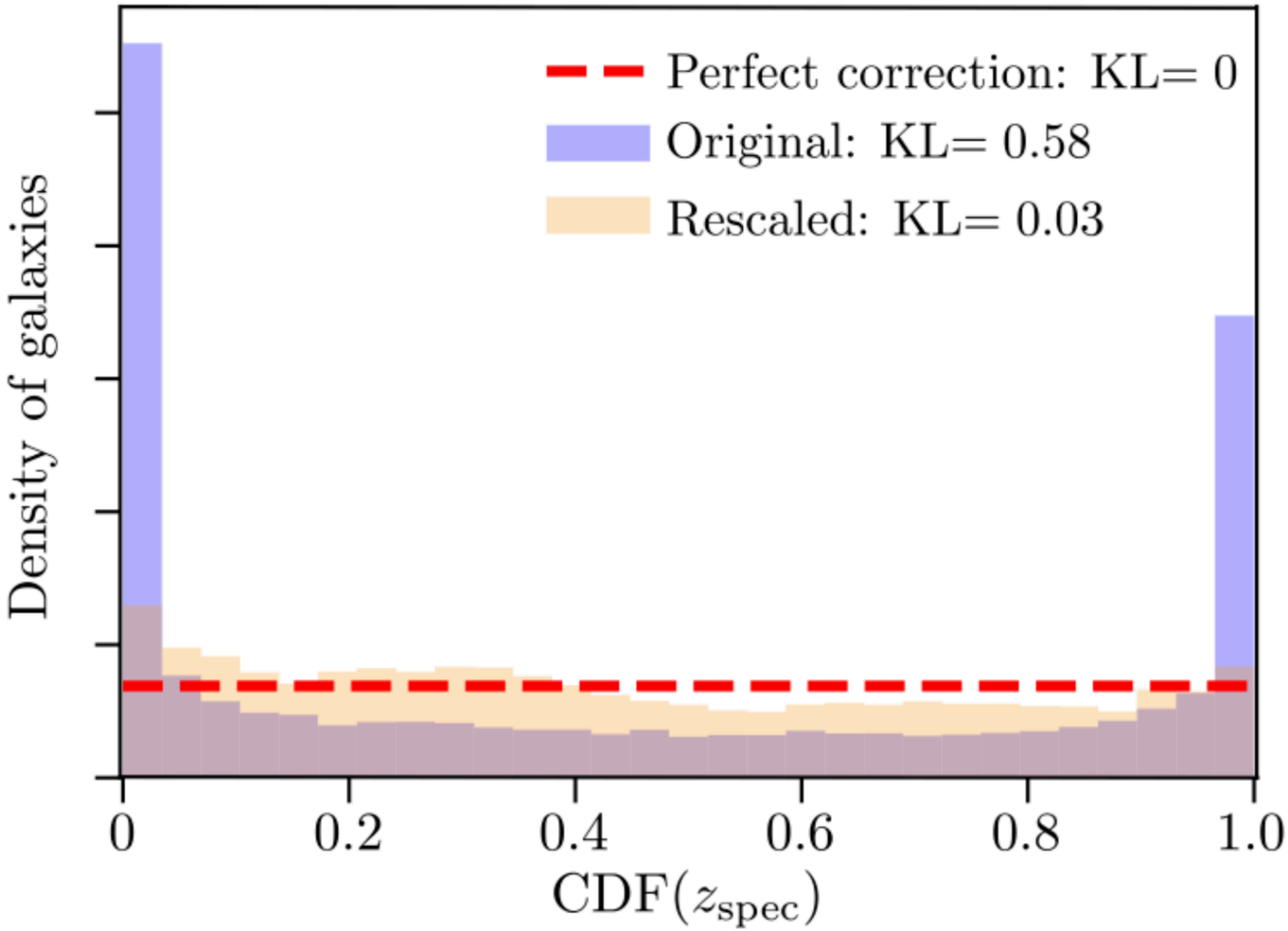


Figure 3. Four different probability integral transforms (PITs). In the case of underdispersed PDFs an u-shaped, concave distribution is observed (a). Overdispersed PDFs result in a peaked, convex distribution (b). When a slope in the PIT is observed, the analysed PDFs are biased (c). Only when the PIT exhibits a flat distribution, the PDFs are well calibrated (d).

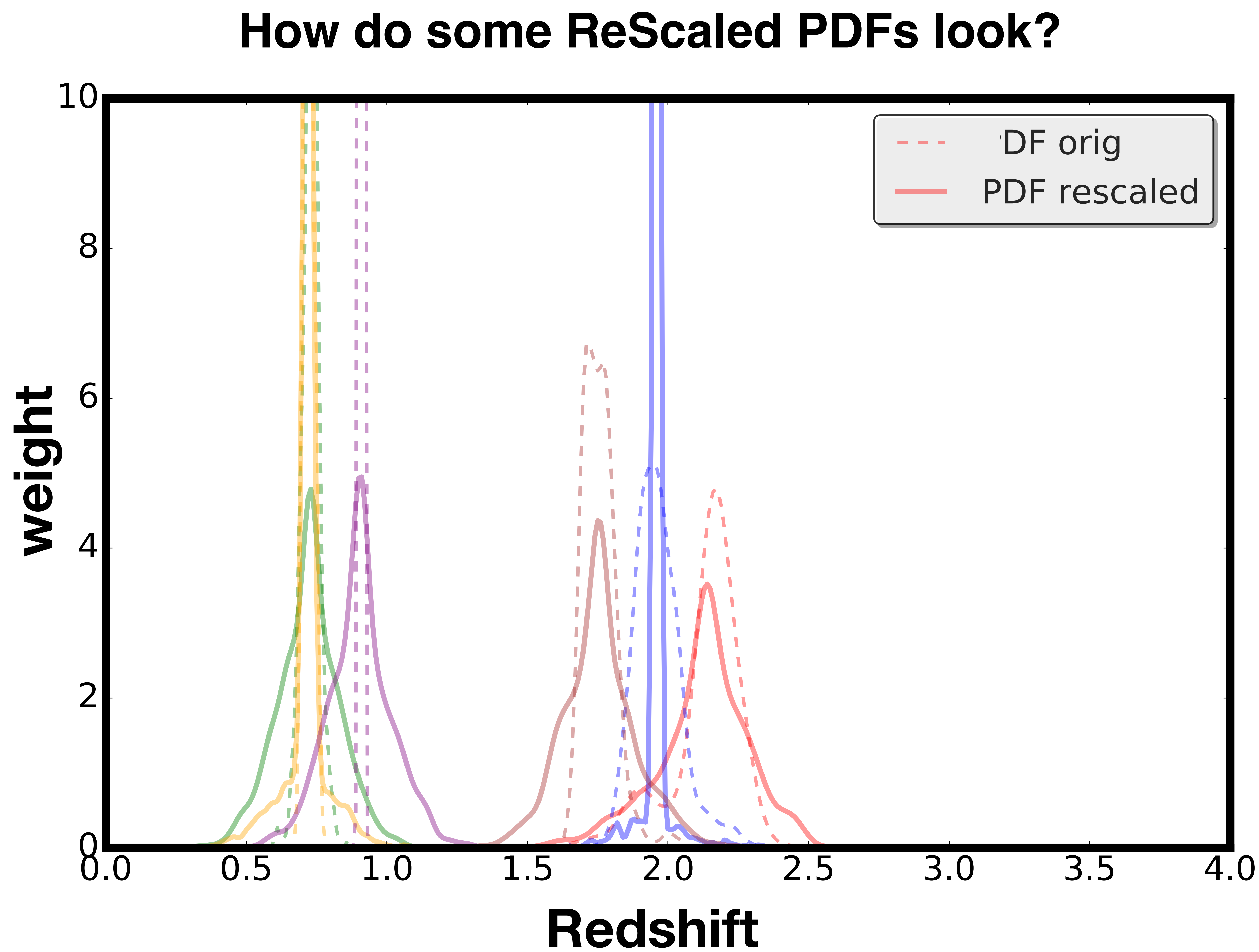
Polsterer 1608.08016

Photometric redshift DFs -> PDFs



PIT [CDF value @ truth value]

BH et al: 1708.01532



How do some ReScaled PDFs look?
PDF orig
PDF rescaled
weight
Redshift
0
2
4
6
8
10
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0

Overview

The supervised ML framework

An introduction to photometric redshifts

My typical ML workflow

A common ML application:

Photometric redshifts

The biggest problem for ML in cosmology:

Unrepresentative labelled data

Dealing with unrepresentative labelled data

Other common applications of ML

Recent, novel applications of ML

Conclusions

Incomplete spectroscopic samples

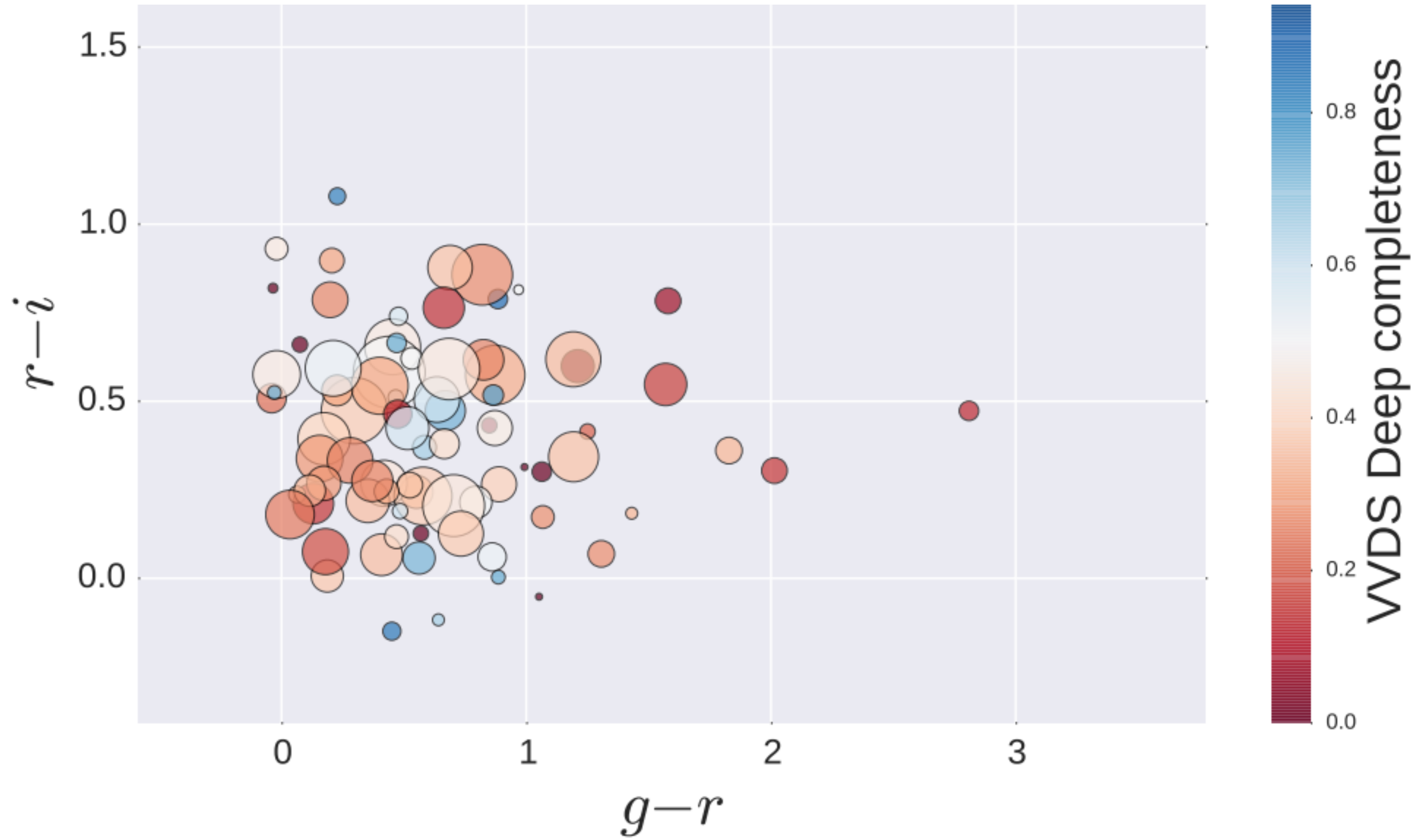
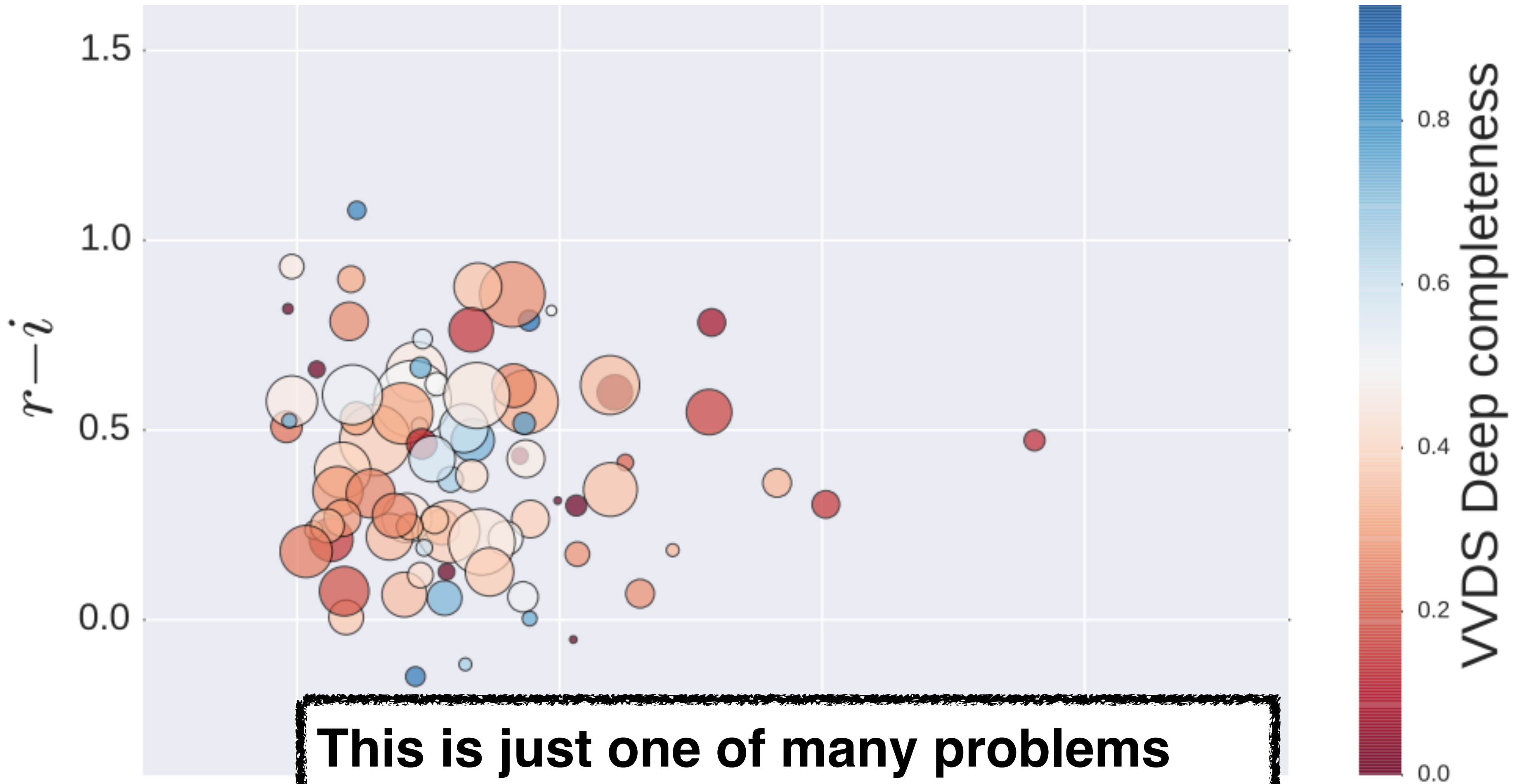


Figure 7. Spectroscopic completeness of the VVDS Deep sample in $g-r$ vs $r-i$ colour space. Each point represents the centre of a 4-D colour-magnitude k-means cell containing a similar number of galaxies from the DES SV NGMIX catalogue. The size of the point represents the number of targeted objects, while the colour indicates the fraction that returned a reliable redshift. The three magnitude ranges (as labelled) cover the i-band magnitude range that contains the majority of galaxies in the weak lensing sample – see Fig. 1 for the distribution in the catalogues.

Bonnet et al 1507.05909

Incomplete spectroscopic samples



This is just one of many problems that make the labelled data unrepresentative of the science sample data.

Figure 7. Spectroscopic completeness of the VVDS Deep sample in $g-r$ vs $r-i$ colour space. Each point represents the centre of a 4-D colour-magnitude k-means cell containing a similar number of galaxies from the DES SV NGMIX catalogue. The size of the point represents the number of targeted objects, while the colour indicates the fraction that returned a reliable redshift. The three magnitude ranges (as labelled) cover the i-band magnitude range that contains the majority of galaxies in the weak lensing sample – see Fig. 1 for the distribution in the catalogues.

Bonnet et al 1507.05909

Overview

The supervised ML framework

An introduction to photometric redshifts

My typical ML workflow

A common ML application:

    Photometric redshifts

The biggest problem for ML in cosmology:

    Unrepresentative labelled data

Dealing with unrepresentative labelled data

Other common applications of ML

Recent, novel applications of ML

Conclusions

Validating photo-z distribution in Y1 Dark Energy Survey

Value	Bin 1	Bin 2	Bin 3	Bin 4
z^{PZ} range	0.20–0.43	0.43–0.63	0.63–0.90	0.90-1.30
COSMOS final Δz^i, tomographic uncertainty	$+0.001 \pm 0.020$	-0.014 ± 0.021	$+0.008 \pm 0.018$	-0.057 ± 0.022
WZ final Δz^i	$+0.008 \pm 0.026$	-0.031 ± 0.017	-0.010 ± 0.014	—
Combined final Δz^i	$+0.004 \pm 0.0$ [illegible]	[illegible]	[illegible]	[illegible] 022

Δ_z and it's uncertainty

Δ_z = <z_true> - <z-photz>

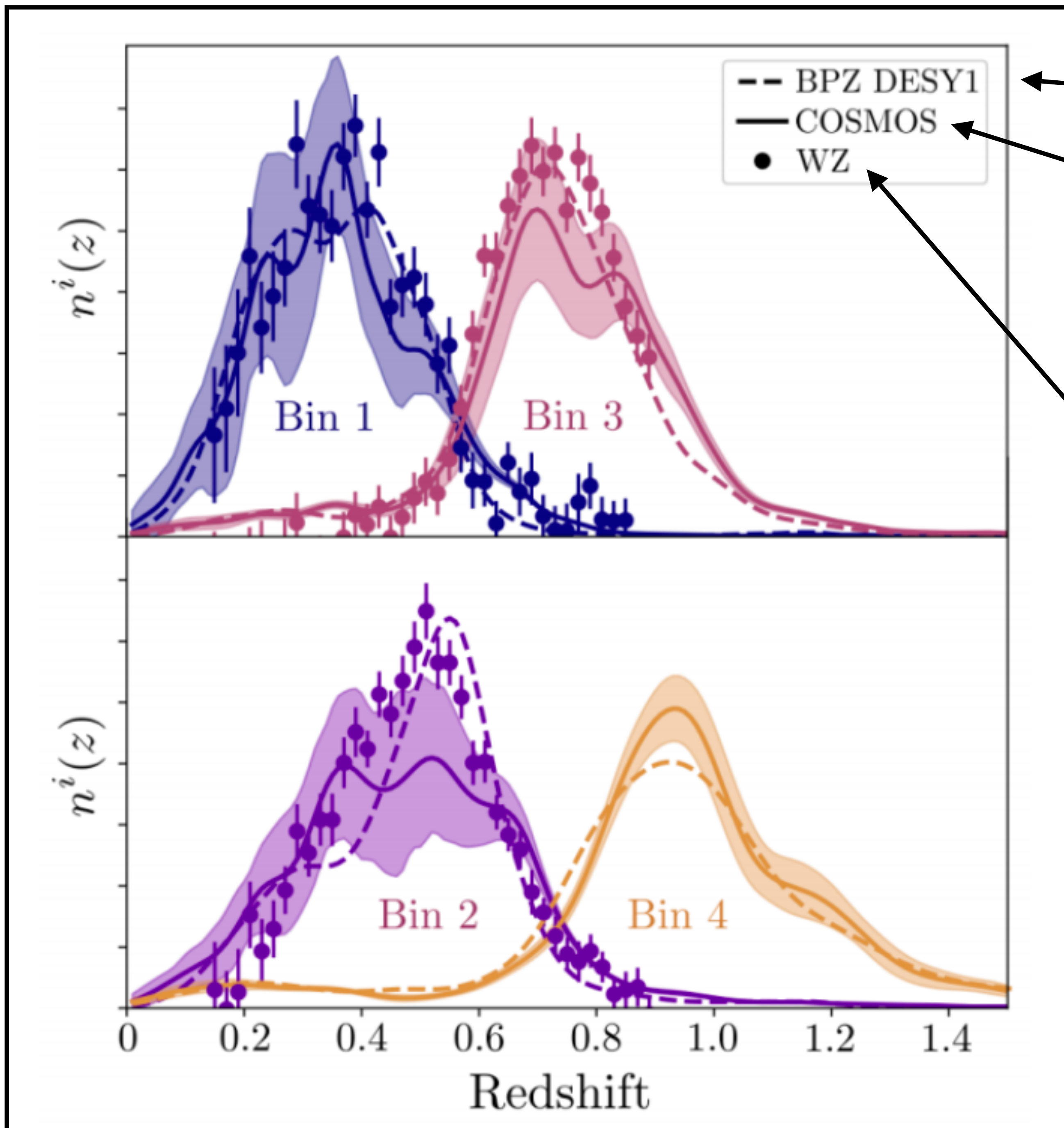


Photo-z predictions

Method 1:
Color-redshift mapping using
30 band photo-z [cosmic variance]

Method 2:
Estimation of dndz of a sample using the clustering technique (i.e, cross correlate with a sample of objects with known redshifts)

Hoyle, Grün & DES et al 2017

Overview

The supervised ML framework

An introduction to photometric redshifts

My typical ML workflow

A common ML application:

Photometric redshifts

The biggest problem for ML in cosmology:

Unrepresentative labelled data

Dealing with unrepresentative labelled data

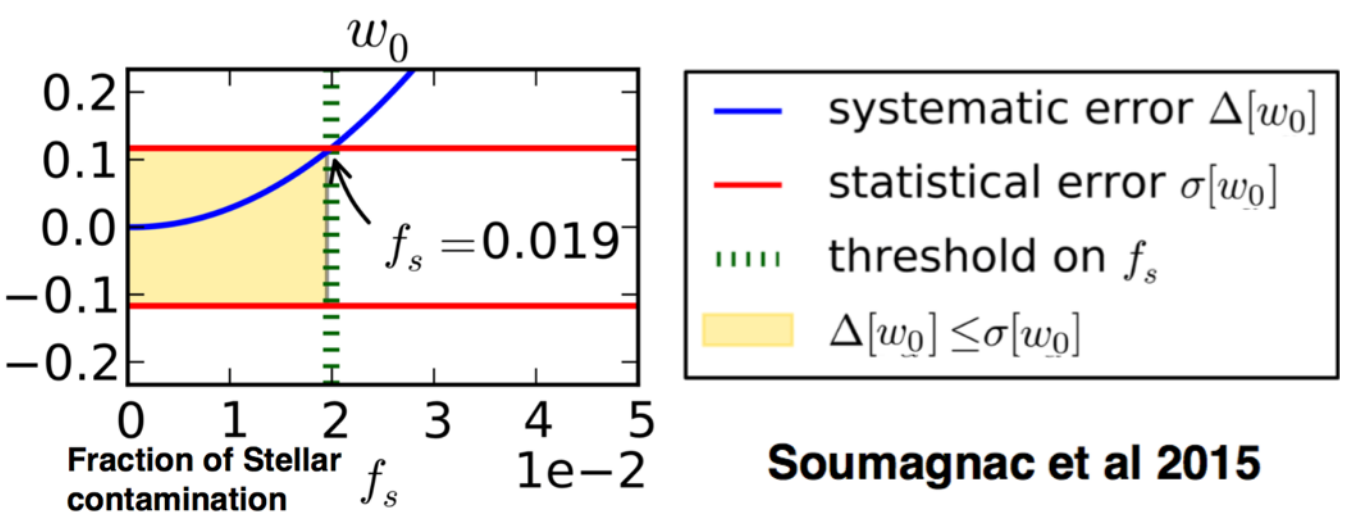
Other common applications of ML

Recent, novel applications of ML

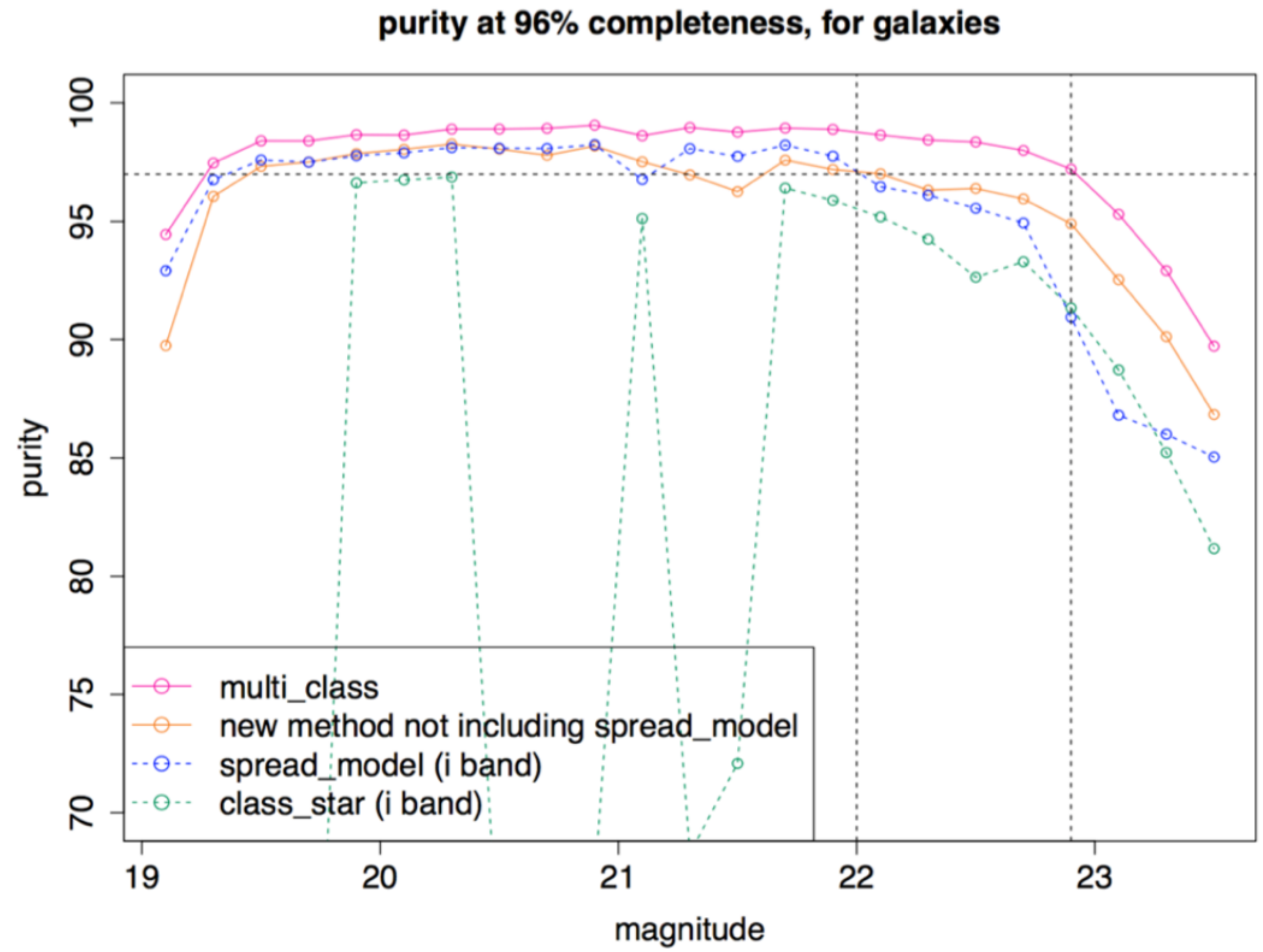
Conclusions

Star Galaxy separation

Given an image of the night sky, is an object a star in our galaxy, or a far away galaxy? Improvement in star-galaxy classification leads to reduced errors in cosmological analysis e.g. DES SV analysis:



Soumagnac et al 2015

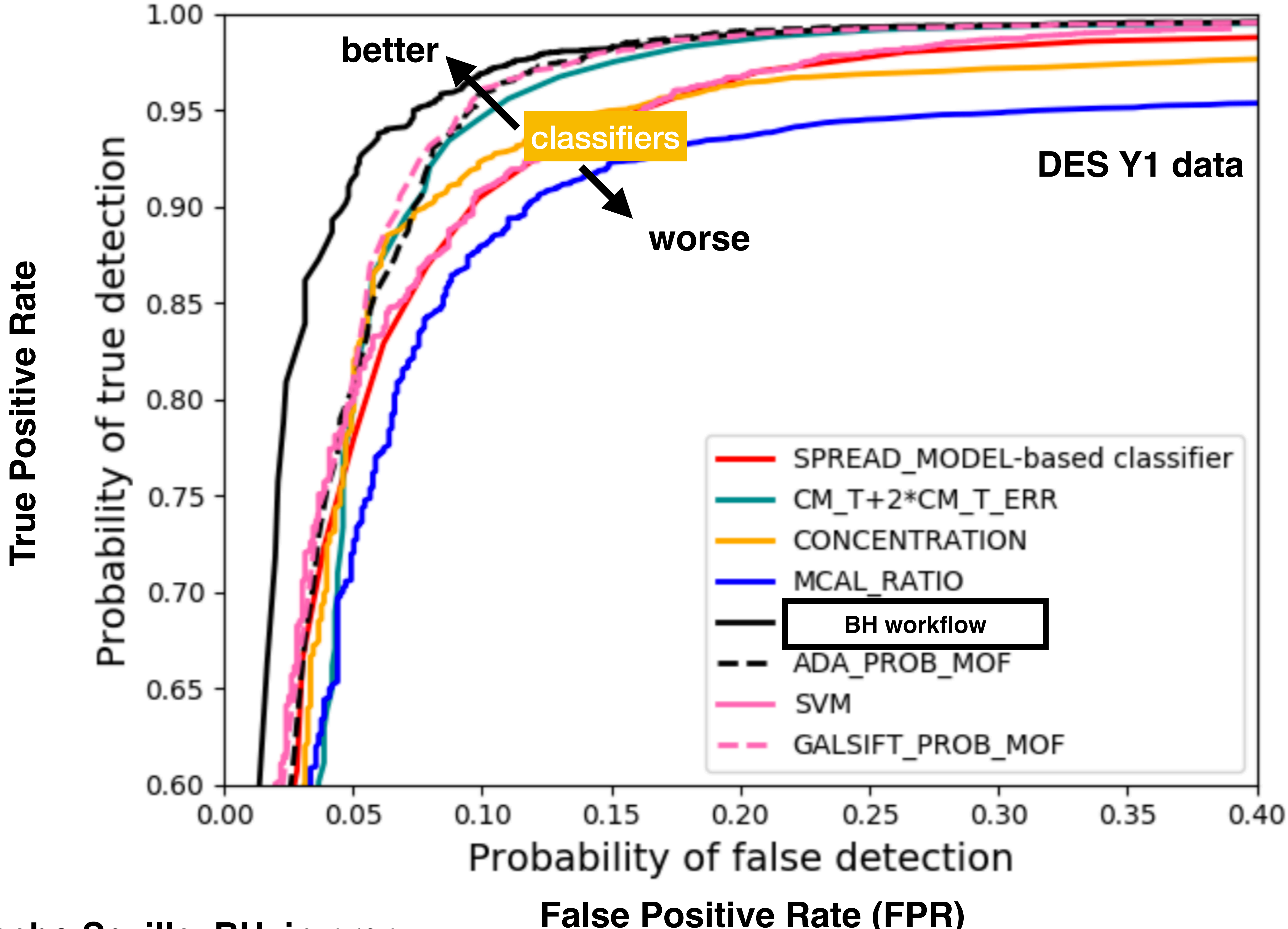


Using a PCA method to select features

In Y1 we face a similar problem as before labelled data is biased!

Moving towards higher order measurements of the predicted signal. e.g. does the number density of stars increase as one approaches the LMC / our Galaxy disk (Nacho Sevilla, BH, DES et al in prep)

Feature Importance Applied to Star-Gal Sep.



Nacho Sevilla, BH, in prep.

Overview

The supervised ML framework

An introduction to photometric redshifts

My typical ML workflow

A common ML application:

Photometric redshifts

The biggest problem for ML in cosmology:

Unrepresentative labelled data

Dealing with unrepresentative labelled data

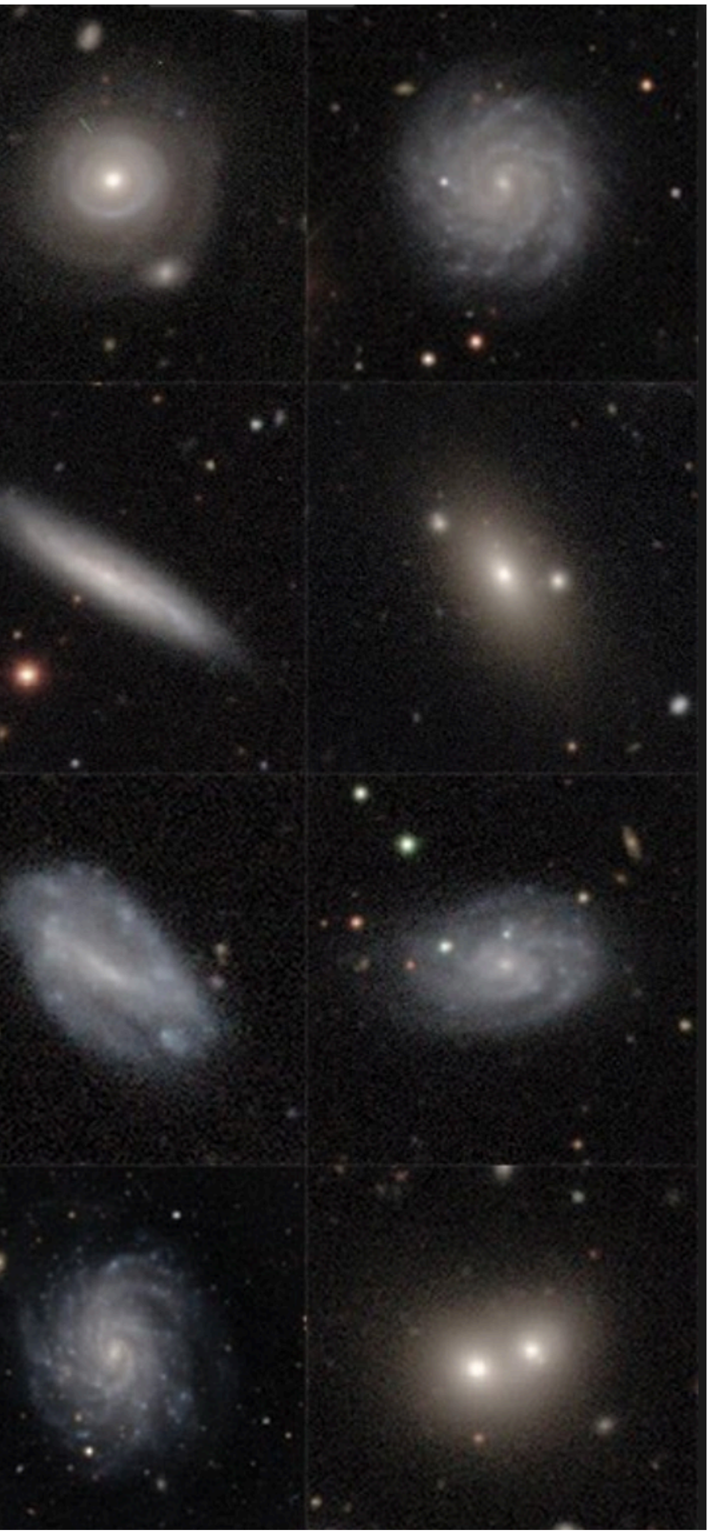
Other common applications of ML

Recent, novel applications of ML I/II

Conclusions

Convolutional Neural Networks

Galaxy Zoo: A massive program to train members of the public to visually inspect 1 Million galaxies more than 50 times each



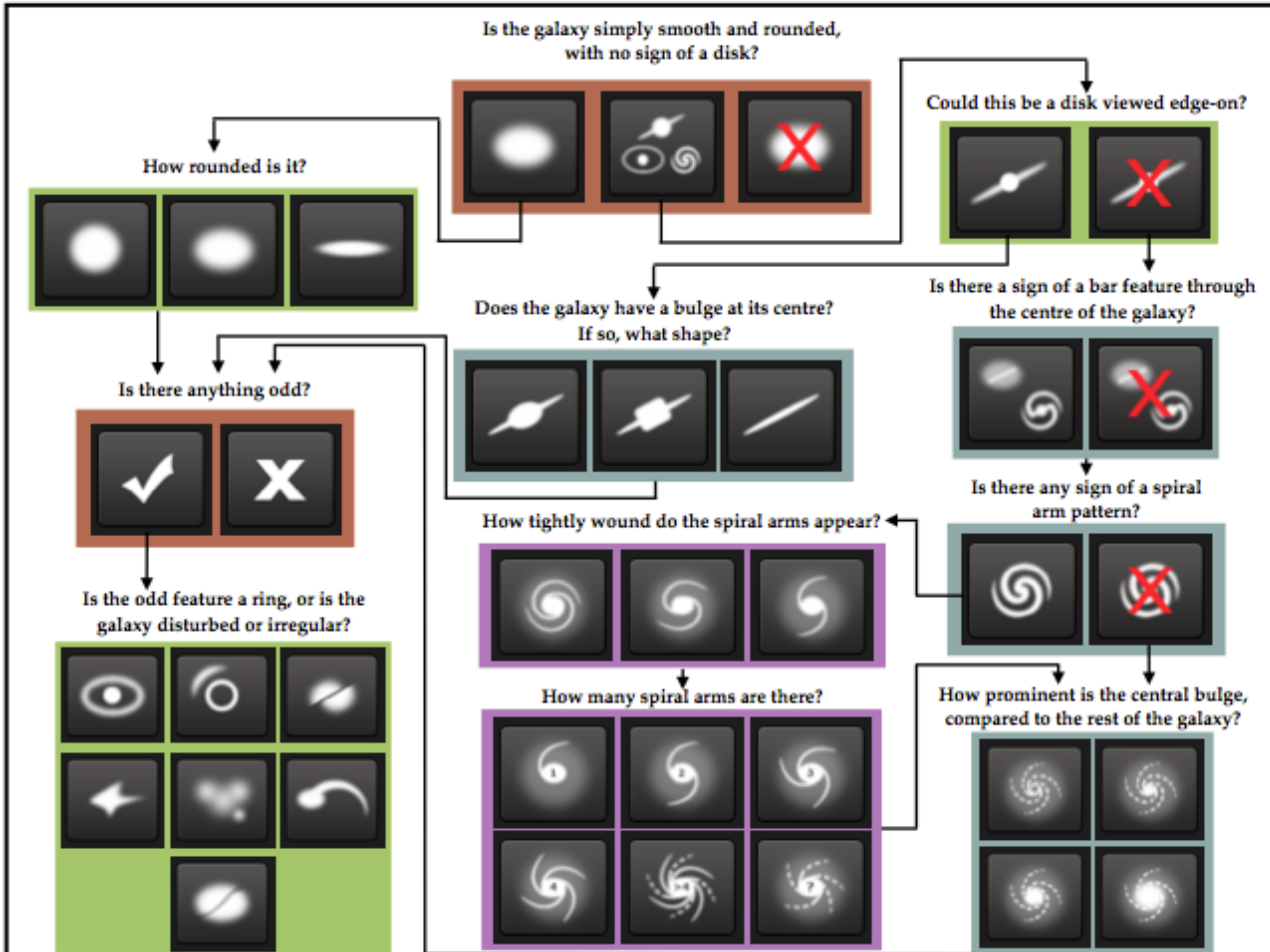


Figure 1. Flowchart of the classification tasks for GZ2, beginning at the top centre. Tasks are colour-coded by their relative depths in the decision tree. Tasks outlined in brown are asked of every galaxy. Tasks outlined in green, blue, and purple are (respectively) one, two or three steps below branching points in the decision tree. Table 2 describes the responses that correspond to the icons in this diagram.

Willet et al 2013

Convolutional Neural Networks

Galaxy Zoo: A massive program to train members of the public to visually inspect 1 Million galaxies more than 50 times each

Kaggle-contest:
use ML to reproduce the classifications of humans.

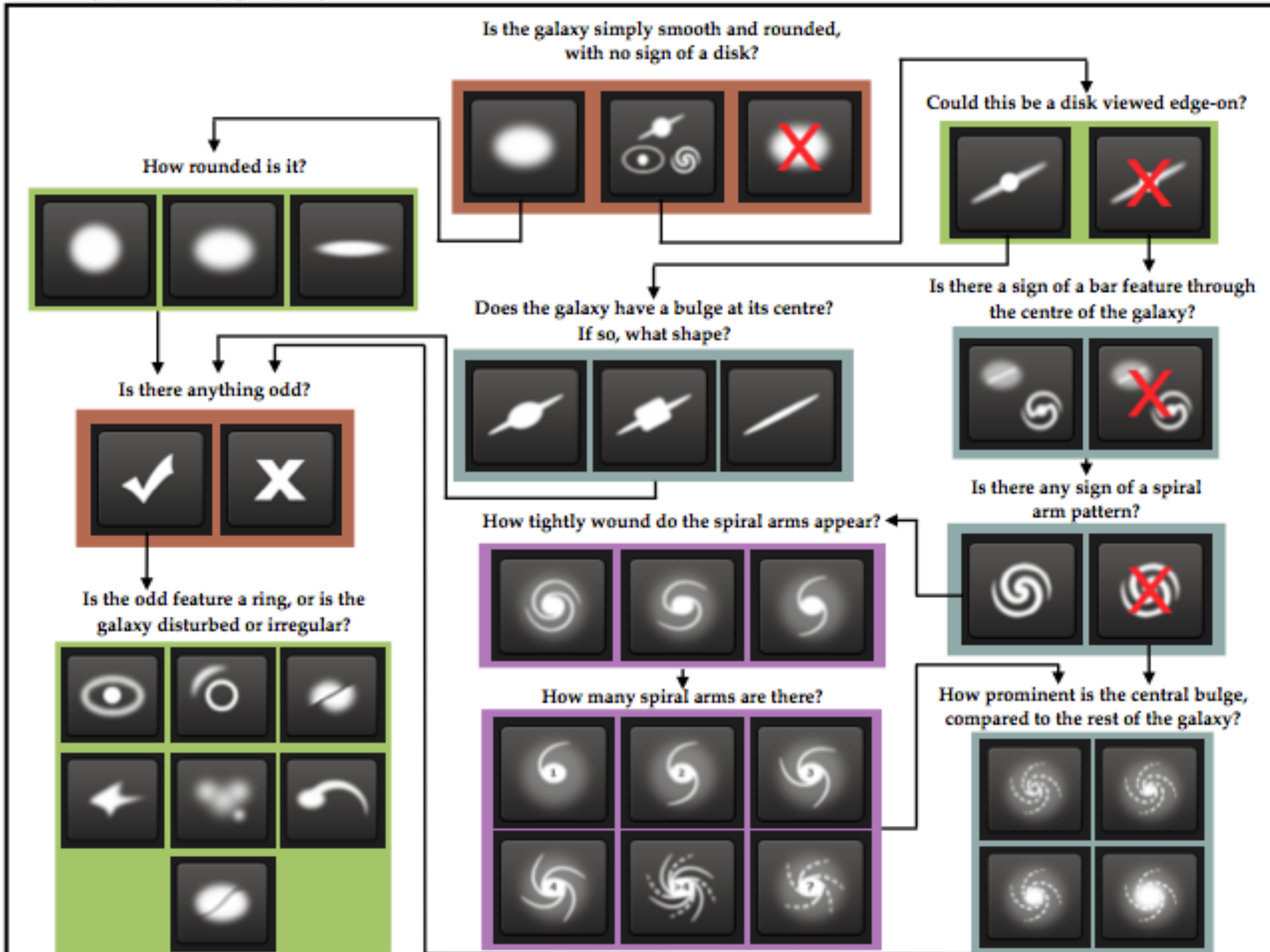


Figure 1. Flowchart of the classification tasks for GZ2, beginning at the top centre. Tasks are colour-coded by their relative depths in the decision tree. Tasks outlined in brown are asked of every galaxy. Tasks outlined in green, blue, and purple are (respectively) one, two or three steps below branching points in the decision tree. Table 2 describes the responses that correspond to the icons in this diagram.

https://www.kaggle.com/c/galaxy-zoo-the-galaxy-challenge **Willet et al 2013**

Convolutional Neural Networks

Galaxy Zoo: A massive program to train members of the public to visually inspect 1 Million galaxies more than 50 times each

**Kaggle-contest:
use ML to reproduce the classifications of humans.**

Could apply results to the 100's million of galaxies and repeat for new surveys

First application of Deep ML with 2d-CovNets in Astrophysics (Dieleman et al 2015)

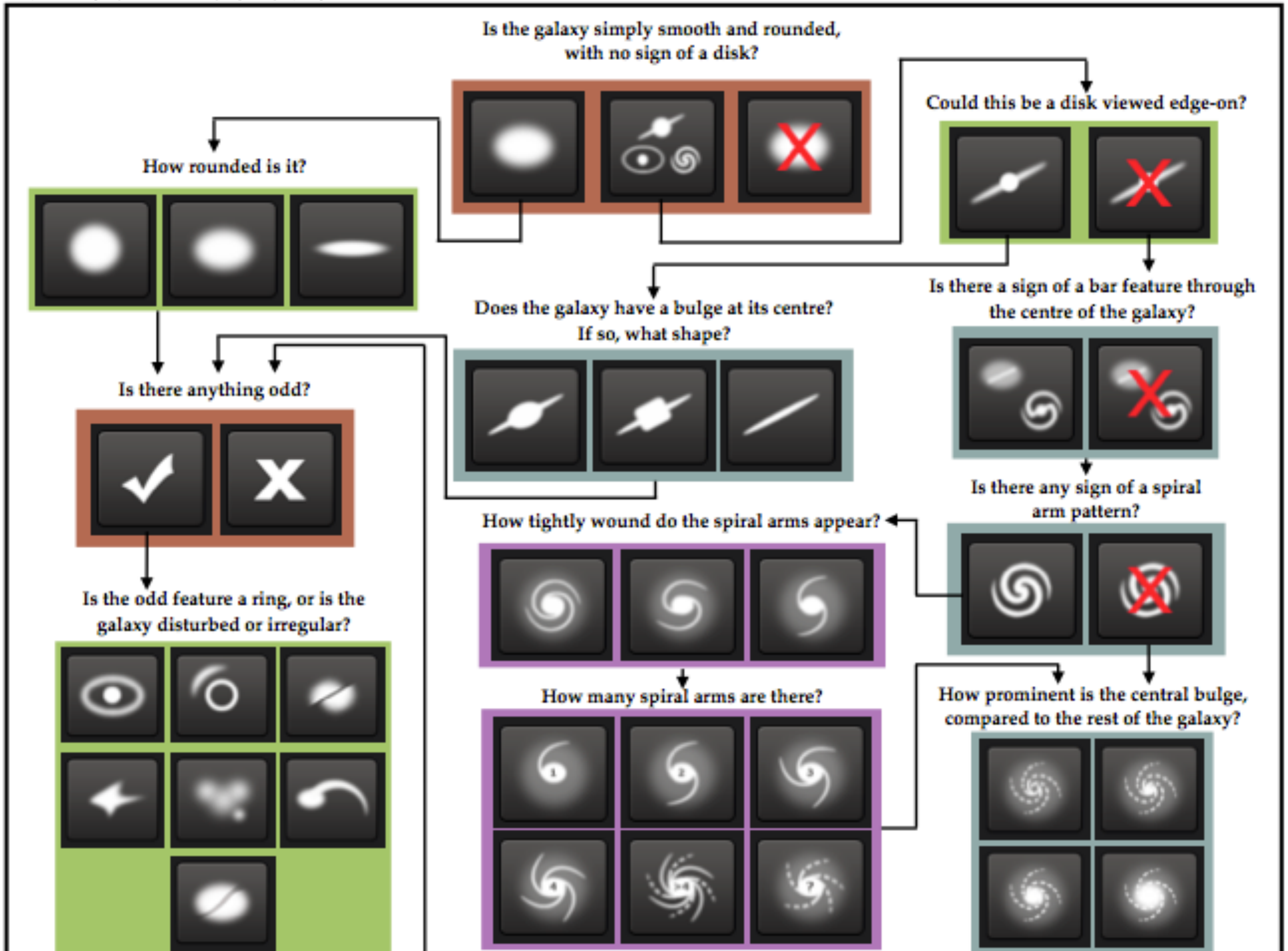


Figure 1. Flowchart of the classification tasks for GZ2, beginning at the top centre. Tasks are colour-coded by their relative depths in the decision tree. Tasks outlined in brown are asked of every galaxy. Tasks outlined in green, blue, and purple are (respectively) one, two or three steps below branching points in the decision tree. Table 2 describes the responses that correspond to the icons in this diagram.

https://www.kaggle.com/c/galaxy-zoo-the-galaxy-challenge

Willet et al 2013

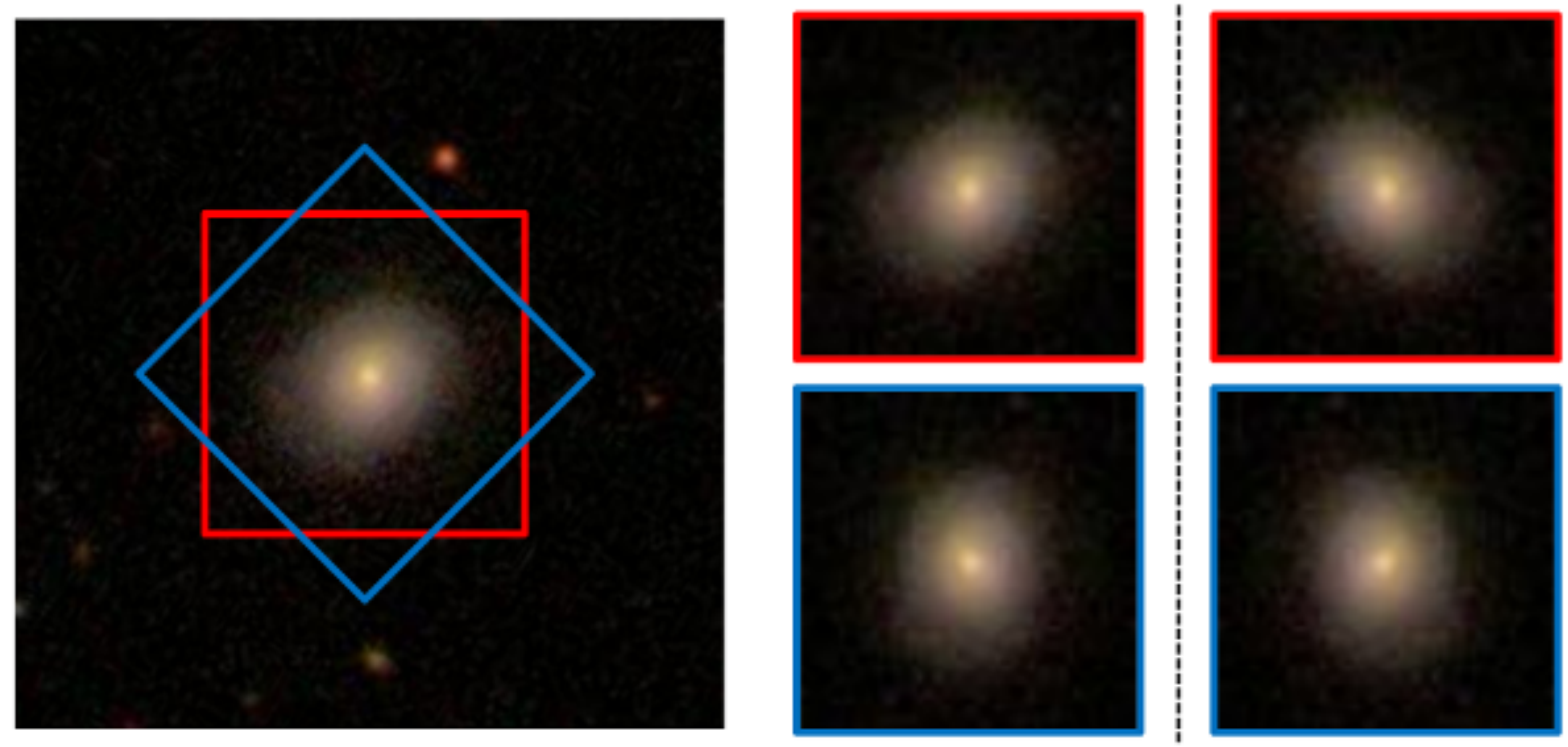
CNNs for Galaxy Zoo

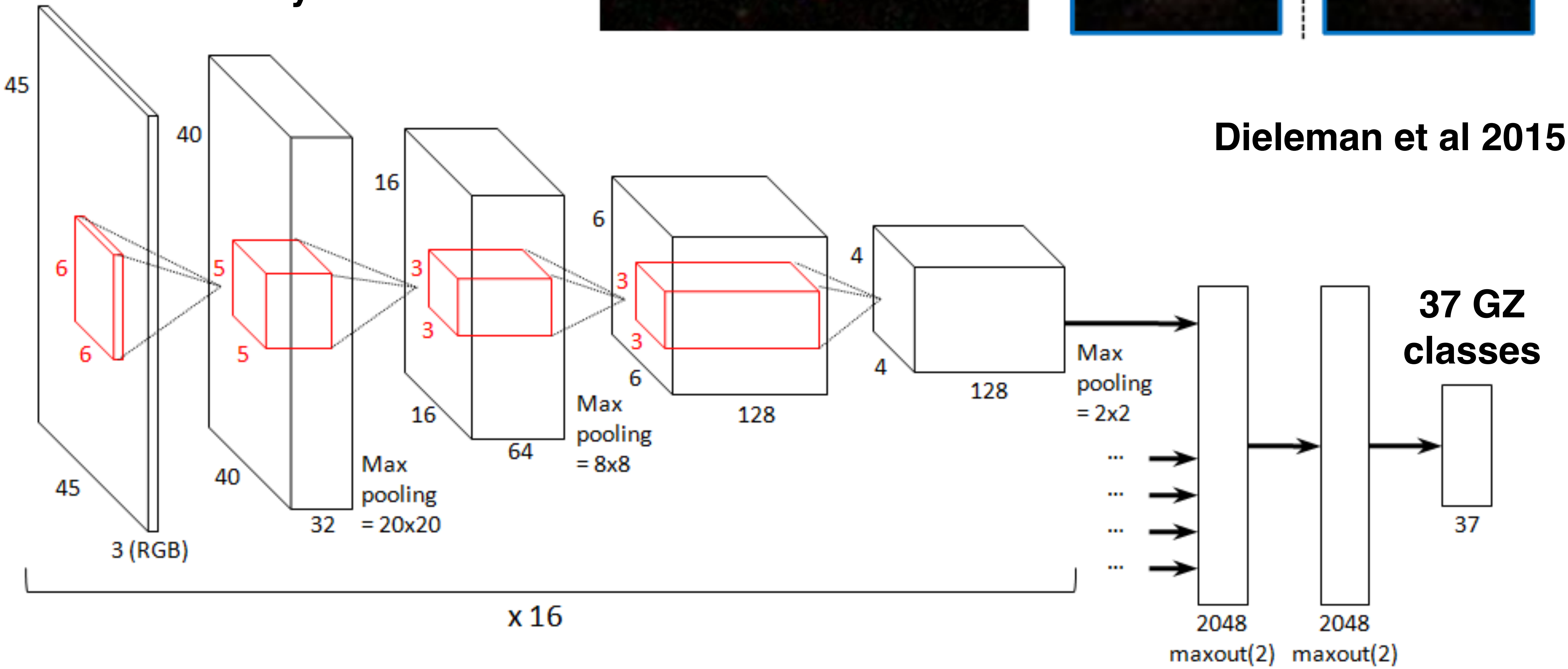
Extract centre of image
=> the galaxy,
rescaled to 45x45 pixels

Data augmentation

Dropout/Max pooling

Combined many networks





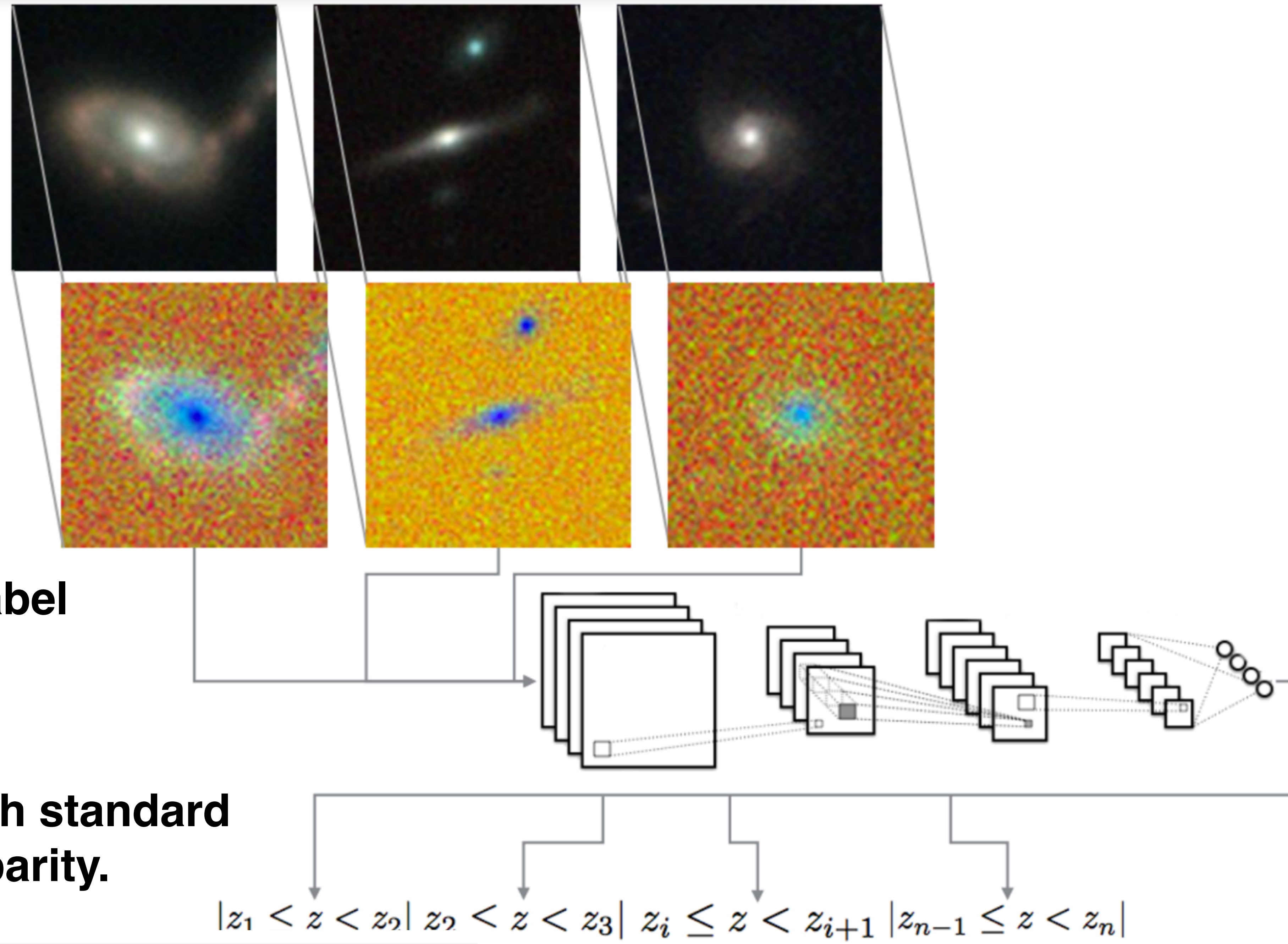
Dieleman et al 2015

http://benanne.github.io/2014/04/05/galaxy-zoo.html

CNNs for redshift estimates

. arXiv:1504.07255 [pdf, other]
Measuring photometric redshifts using galaxy images and Deep Neural Networks
Ben Hoyle

Inputs: galaxy image
->
ImageNet architecture
->
Targets: spec-z



everything about biased label data is still a problem

Compared performance with standard ML algorithms, and found parity.

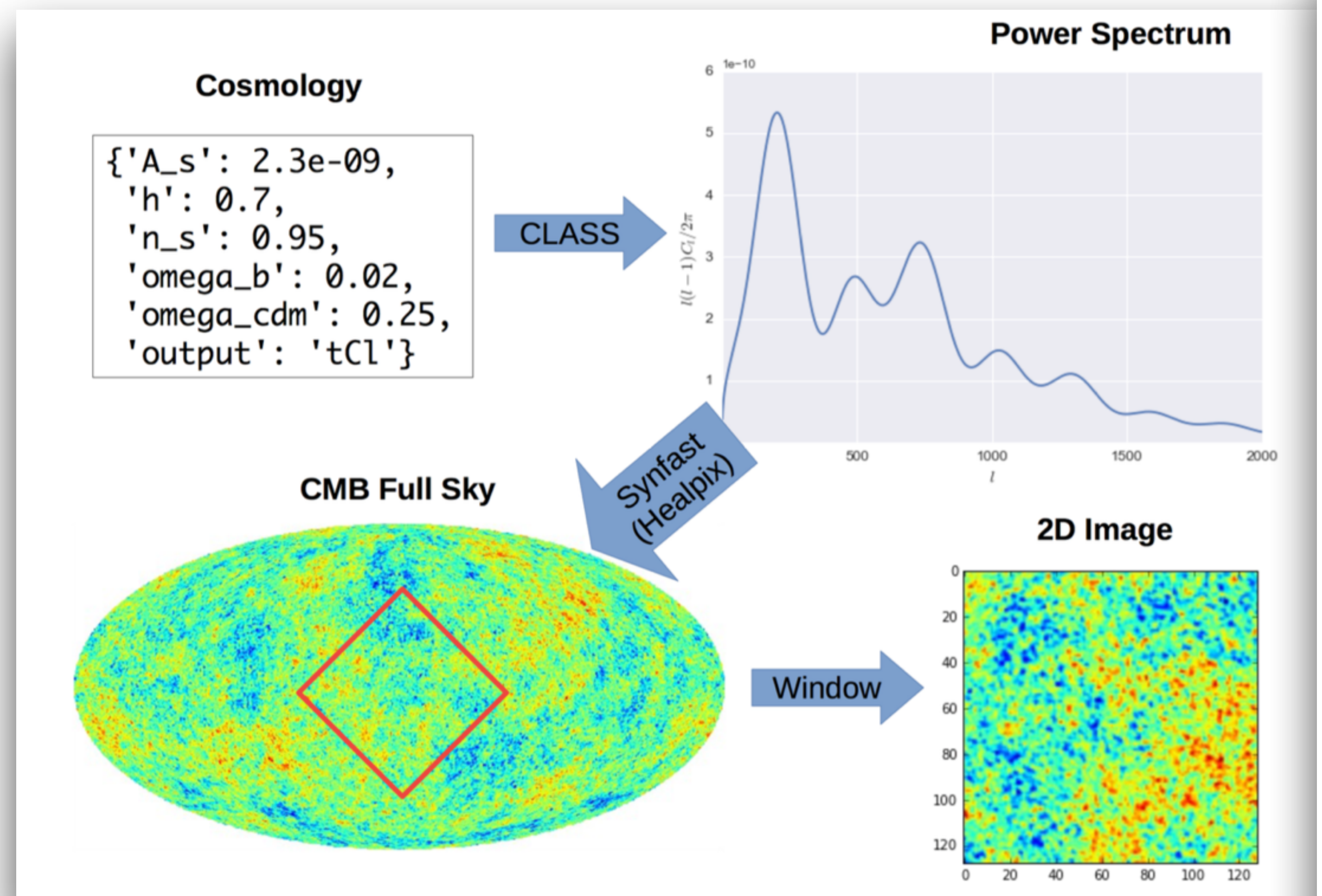
MLA	μ	σ_{68}	σ_{95}	$\lvert\Delta\ /(1+z_{spec})\rvert > 0.15$
DNNs	0.00	0.030	0.10	1.71%
AdaBoost	−0.001	0.030	0.10	1.56%

$$\Delta = z_{spec} - z_{predict}$$

CNNs for Cosmic Microwave Background radiation

Measuring Cosmological Parameters from Simulated CMB Images with Convolutional Neural Networks

Is there information in the CMB that is not contained in Cls? E.g. Higher order moments, such as non-Gaussianities.



2D CNN Configuration
input (128×128)
Conv2D (3×3) - 16 Conv2D (3×3) - 16
maxpool (2×2)
Conv2D (3×3) - 32 Conv2D (3×3) - 32
maxpool (2×2)
Conv2D (3×3) - 64 Conv2D (3×3) - 64
maxpool (2×2)
Conv2D (3×3) - 128 Conv2D (3×3) - 128
maxpool (2×2)
FC - 256
FC - 128
FC - 1 / FC - 2

1D CNN Configuration
input (16384)
Conv1D $(4, \mathit{Stride}\ 4)$ - 128 Conv1D $(4, \mathit{Stride}\ 4)$ - 128
maxpool (4)
Conv1D $(4, \mathit{Stride}\ 4)$ - 256 Conv1D $(4, \mathit{Stride}\ 4)$ - 256
maxpool (4)
FC - 256
FC - 128
FC - 1 / FC - 2

	ΔA_s	$\Delta\Omega_{CDM}$	$\Delta A_s^{(single)}$
PolSpice correlation function	$1.45 \cdot 10^{-10}$	0.025	$3.3 \cdot 10^{-11}$
2D CNN	$1.68 \cdot 10^{-10}$	0.0357	$7.19 \cdot 10^{-11}$
1D CNN	$1.91 \cdot 10^{-10}$	0.0437	-

Robert Lohmeyer Master thesis 2017

Supervisor BH

A random sample of CNN papers

Spectral classification using convolutional neural networks
https://arxiv.org › cs
by P Hála - 2014 - Cited by 2 - Related articles
Dec 29, 2014 - This thesis is about training a **convolutional neural network** (ConvNet) to ... neural networks and deep learning methods in **astrophysics**.

Fast Automated Analysis of Strong Gravitational Lenses with Convolutional Neural Networks

Yashar D. Hezaveh, Laurence Perreault Levasseur, Philip J. Marshall

arXiv:1704.02744 [pdf, other]

Finding strong lenses in CFHTLS using convolutional neural networks
Colin Jacobs, Karl Glazebrook, Thomas Collett, Anupreeta More, Christopher McCarthy
Comments: 16 pages, 8 figures. Accepted by MNRAS
Subjects: **Instrumentation and Methods for Astrophysics (astro-ph.IM)**; Astrophysics of Galaxies (astro-p

A Convolutional Neural Network For Cosmic String Detection in CMB Temperature Maps

Razvan Ciuca, Oscar F. Hernández, Michael Wolman

(Submitted on 29 Aug 2017)

Overview

The supervised ML framework

An introduction to photometric redshifts

My typical ML workflow

A common ML application:

Photometric redshifts

The biggest problem for ML in cosmology:

Unrepresentative labelled data

Dealing with unrepresentative labelled data

Other common applications of ML

Recent, novel applications of ML II/II

Conclusions

Generative Adversarial Networks (GANs)

Generative:
Deep ML NN1: Input (random noise) vector -> output something / image

Adversarial:
Deep ML NN2: distinguish examples of training data examples from non-training data, e.g. that obtained from NN1

Networks:
Deep ML Convolution Neural Networks.

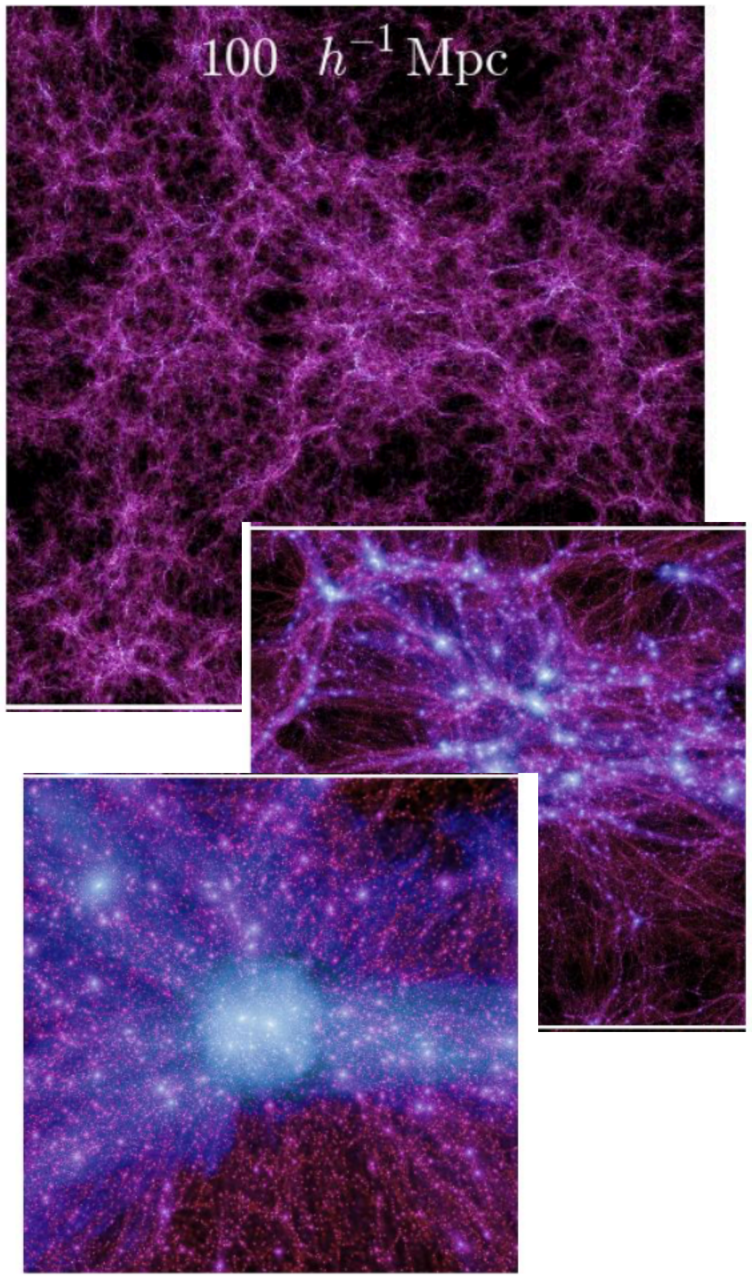
As training proceeds, NN1 generates more and more realistic “examples” from a random noise vector, and NN2 get better and better at distinguishing training data, from everything else, e.g that generated by NN1.

The problem with GANs:
Mode collapse. Difficult learning —> Wasserstein GAN.
https://arxiv.org/abs/1701.07875

https://github.com/bobchennan/Wasserstein-GAN-Keras/blob/master/mnist_wacgan.py
https://raw.githubusercontent.com/farizrahman4u/keras-contrib/master/examples/improved_wgan.py

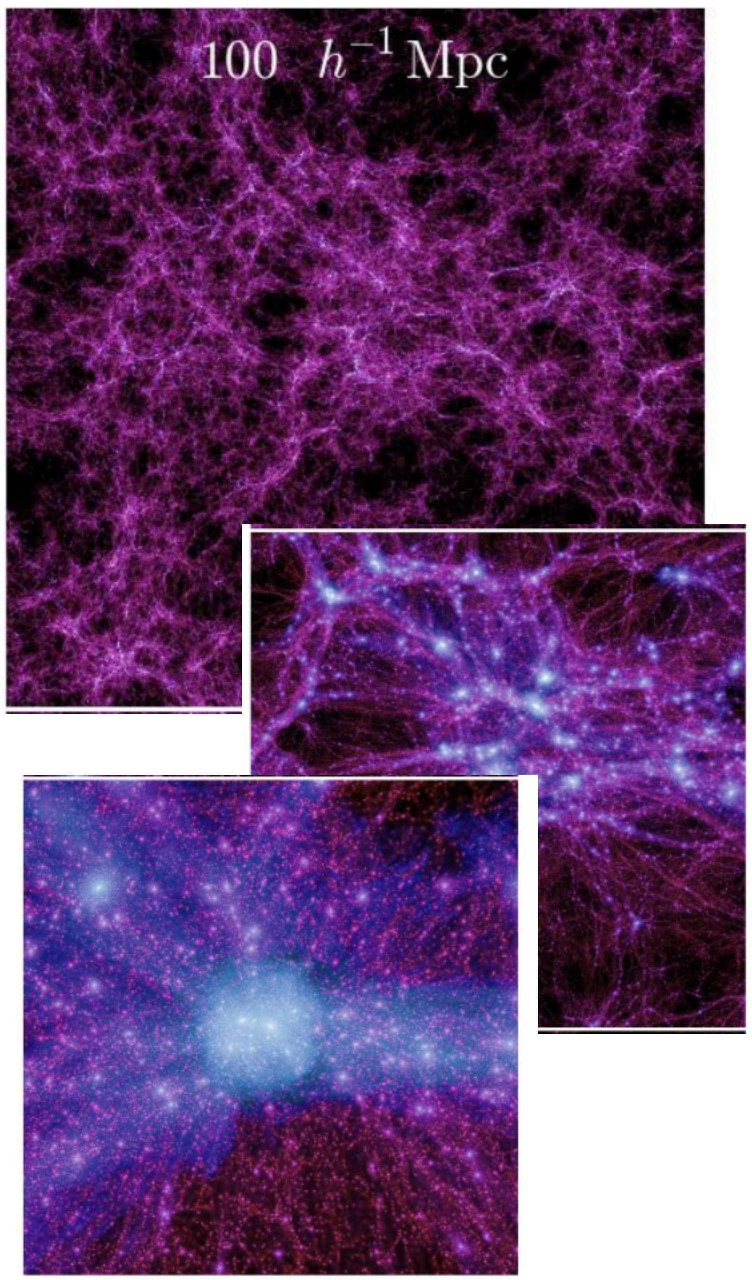
GANs generate realisations of a Dark-Matter N-body sim.

We want to estimate the covariance matrices for correlation functions analysis, e.g. for the Baryon Acoustic Oscillations. Currently we call a very expensive cosmological N-body code called Gadget many 100’s - 1000s of times.



GANs generate realisations of a Dark-Matter N-body sim.

We want to estimate the covariance matrices for correlation functions analysis, e.g. for the Baryon Acoustic Oscillations. Currently we call a very expensive cosmological N-body code called Gadget many 100’s - 1000s of times.



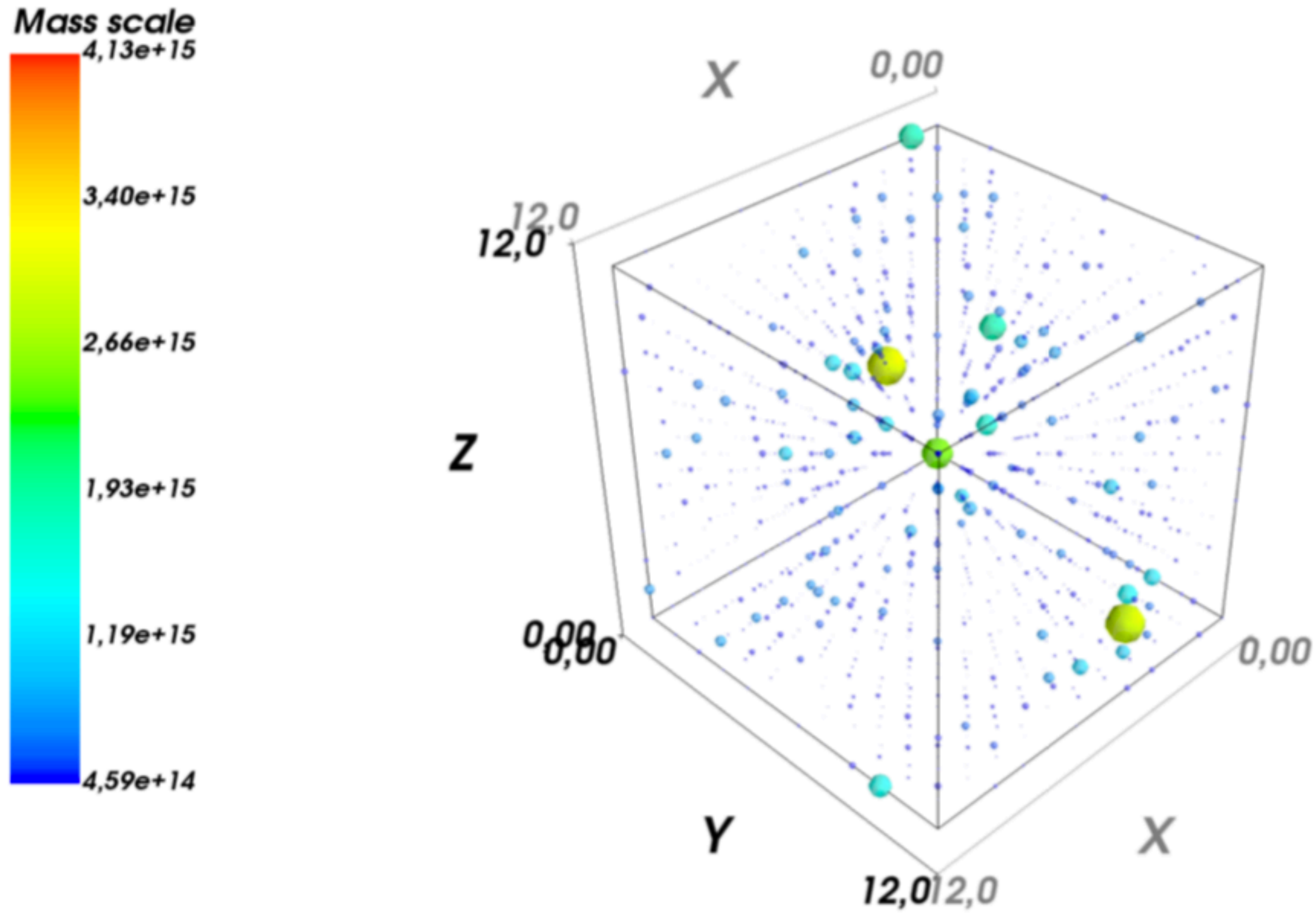


FIGURE 4.4: 3D representation of a randomly selected training example. The color gradient corresponds to the mass scale represented by the color bar in units of $[h^{-1}Mpc]$.

In essence we try to replace the Nbody simulation code with a Deep 3-d CovNet trained using a GAN.
Master thesis Julien Wolf 2018 (Supervisor BH)

GANs generate realisations of a Dark-Matter N-body sim.

Example of the type of data we want to generate (training data)

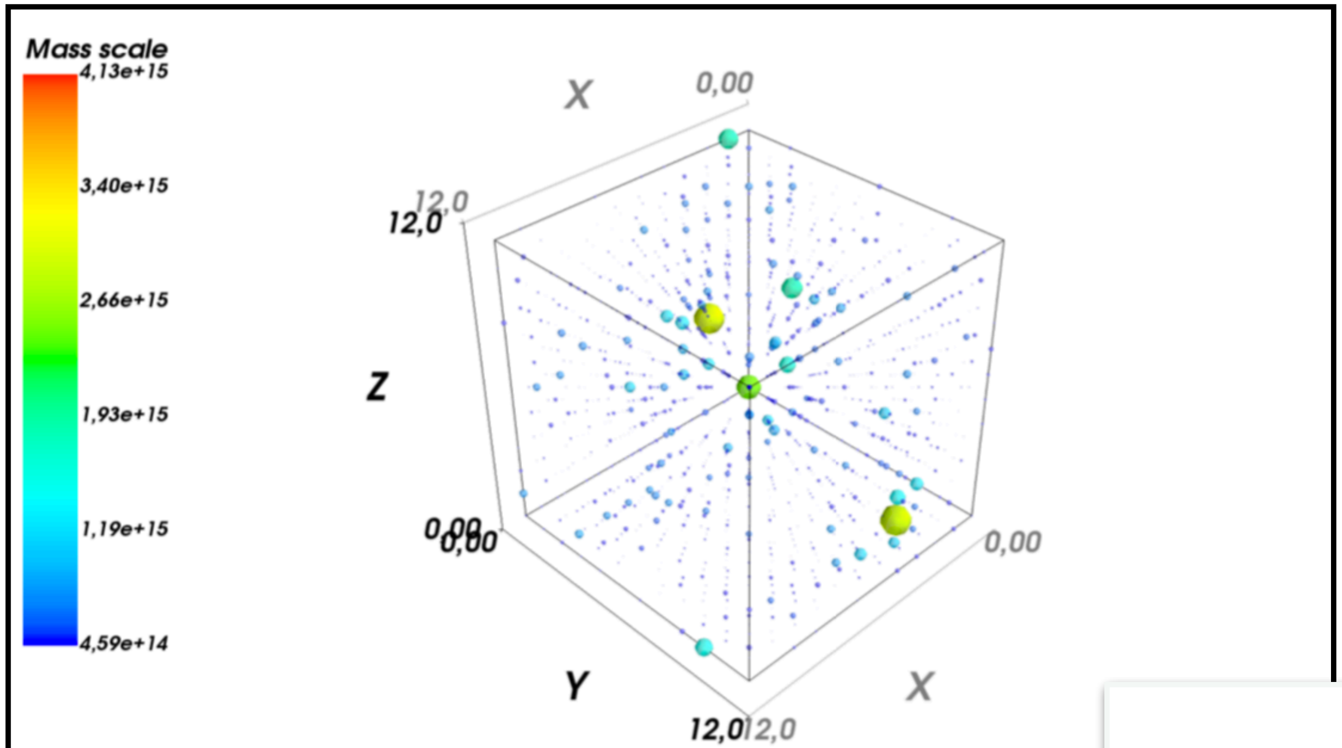
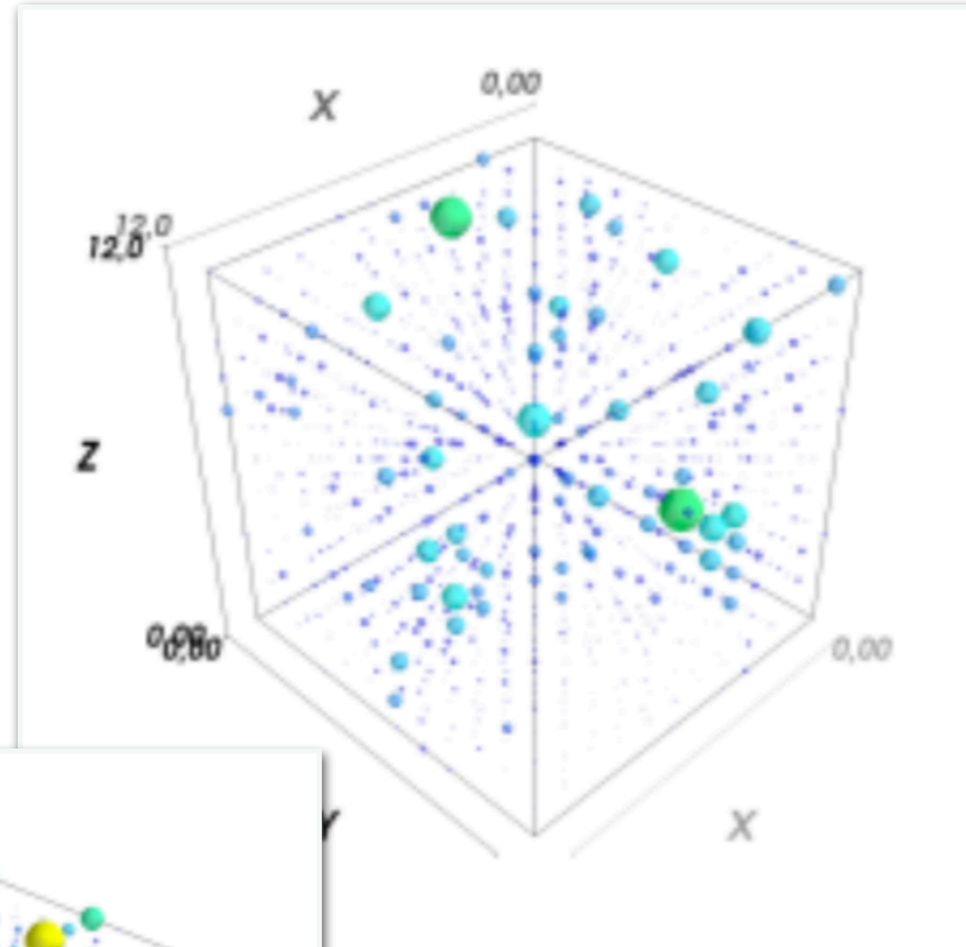
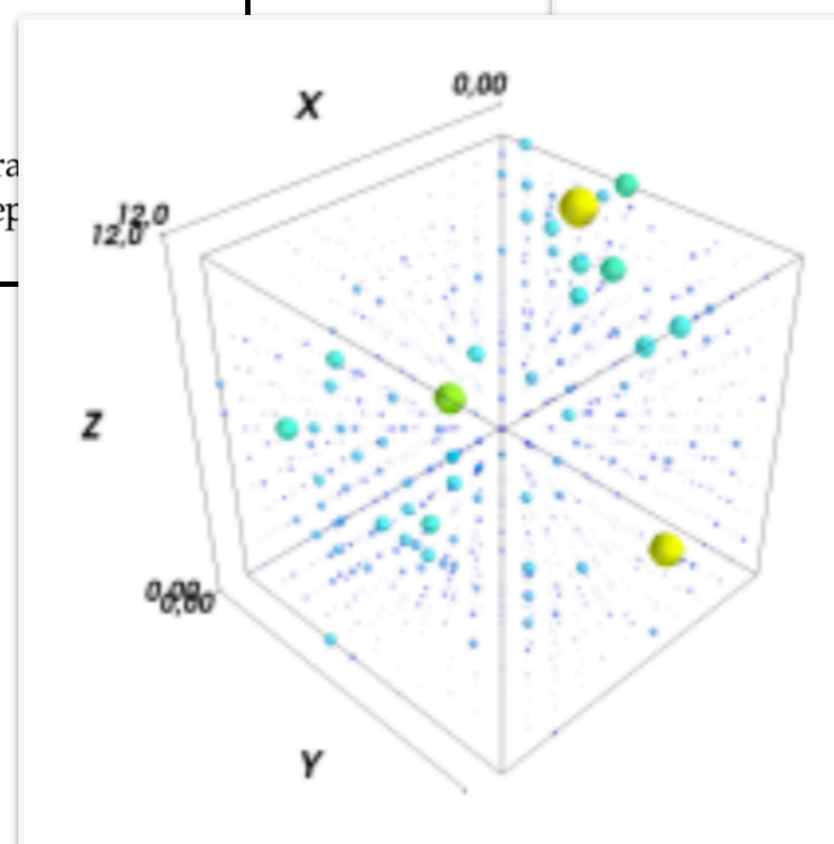


FIGURE 4.4: 3D representation of a randomly selected tra
ple. The color gradient corresponds to the mass scale rep
the color bar in units of $[h^{-1}Mpc]$.

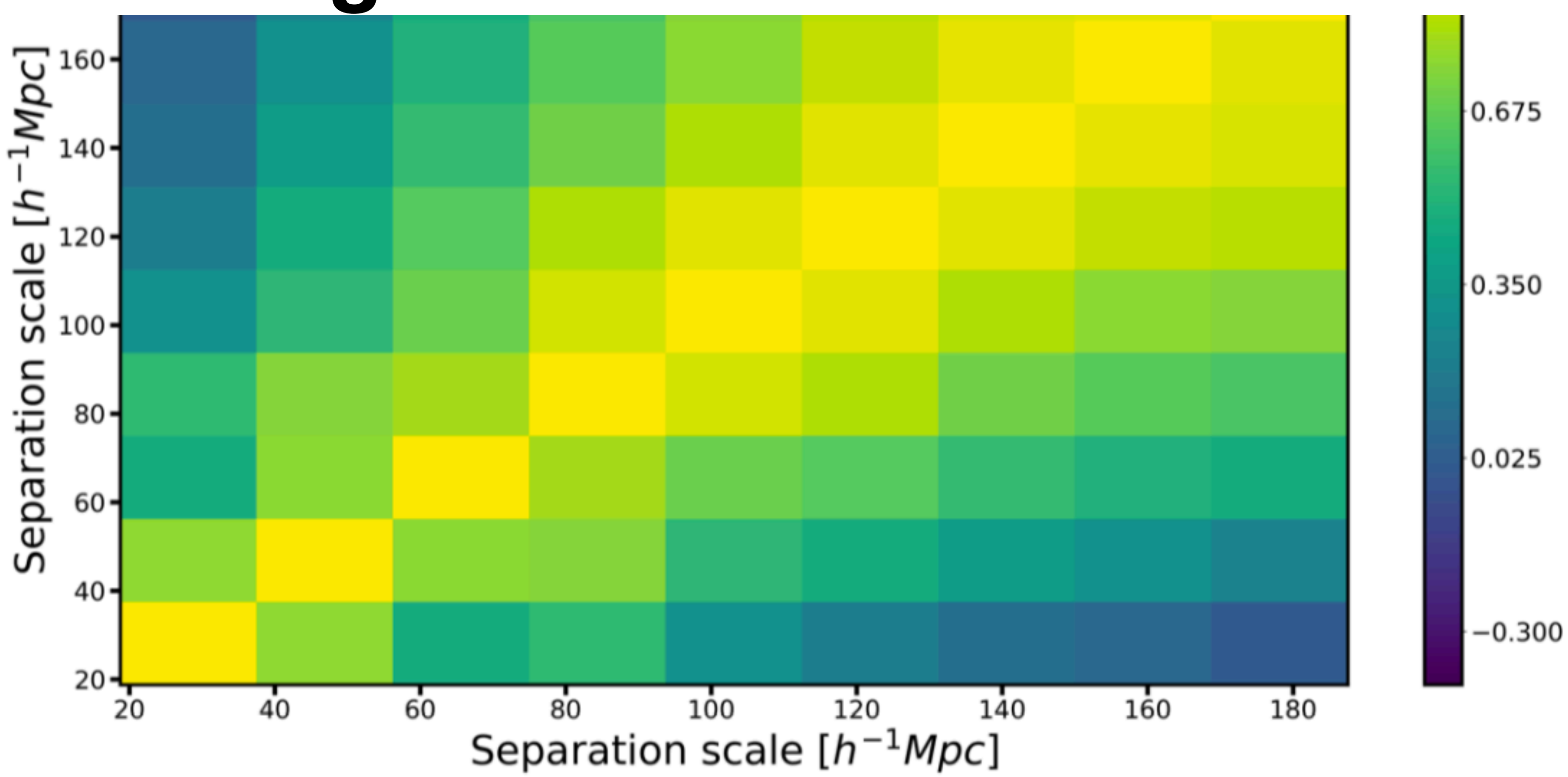




Examples of the GAN generated data

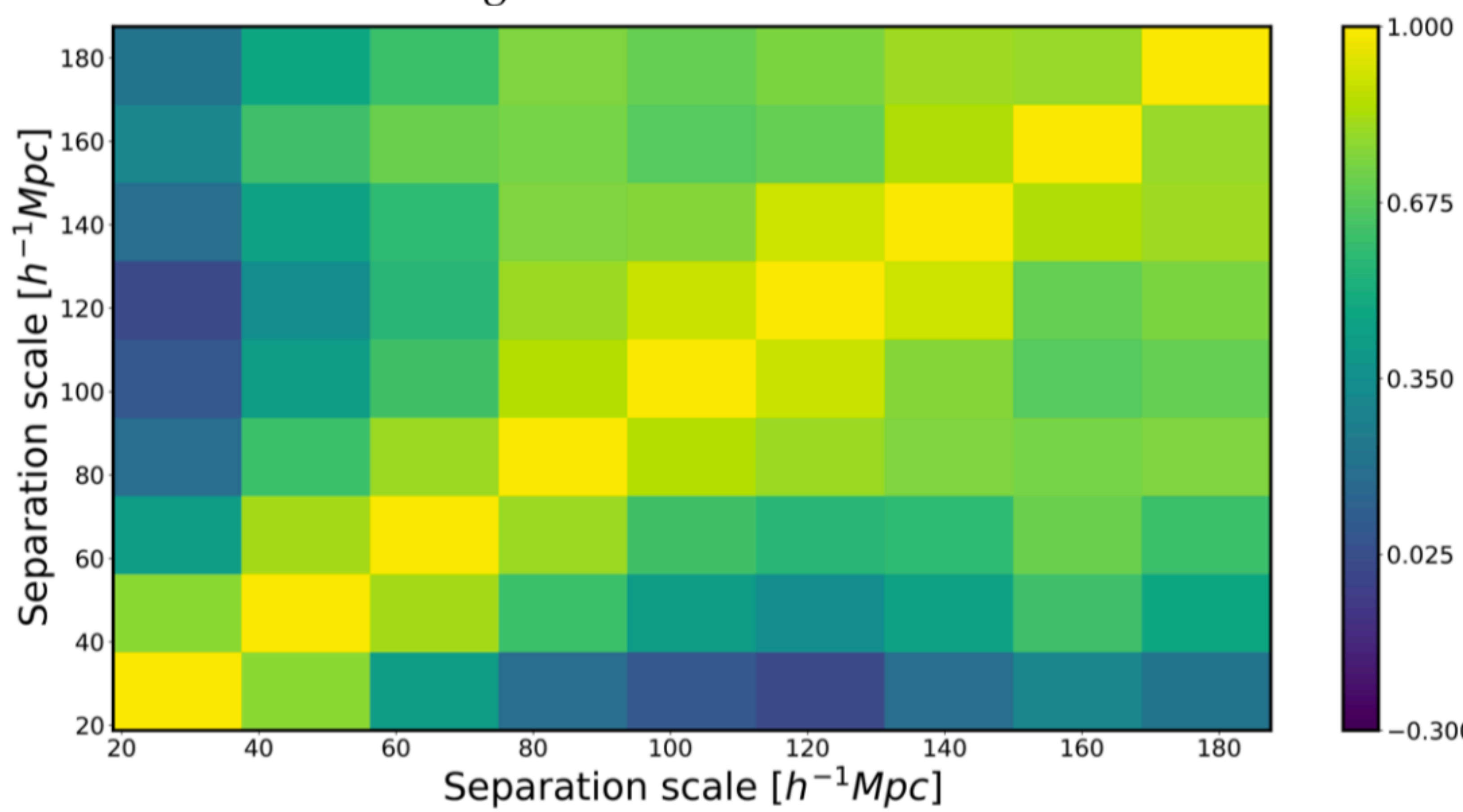
Master thesis Julien Wolf 2018

GANs generate realisations of a Dark-Matter N-body sim.



(a) Correlation matrix for the marked correlation in the training set

The covariance matrix of the correlation functions look reasonable



(b) Correlation matrix for the marked correlation in the GAN output

Master thesis Julien Wolf 2018

Overview

The supervised ML framework

When to use the Machine Learning (ML) hammer

An introduction to photometric redshifts

My typical ML workflow

A common ML application:

Photometric redshifts

The biggest problem for ML in cosmology:

Unrepresentative labelled data

Dealing with unrepresentative labelled data

Other common applications of ML

Recent, novel applications of ML

Conclusions

Conclusions / Summary

Cosmology is in the realm of “big data”; 100’s millions/ billions of galaxies. A subset of objects have target values. Many possibilities of applying machine learning in new and interesting ways.

My personal ML workflow.

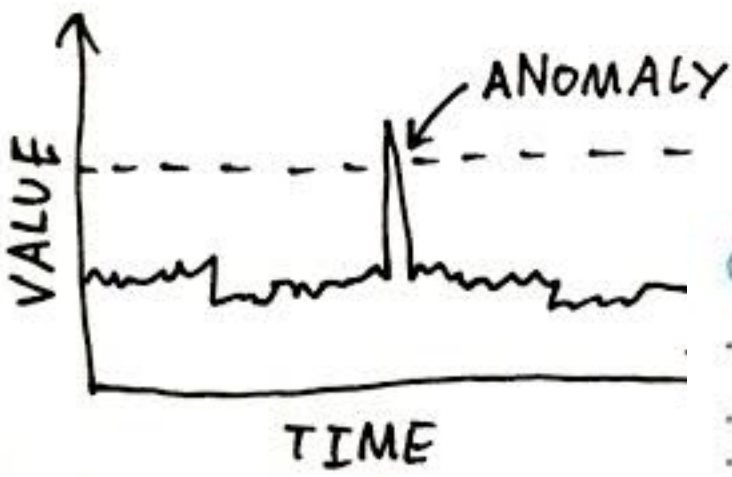
Some cosmological analysis is in a state of crisis:

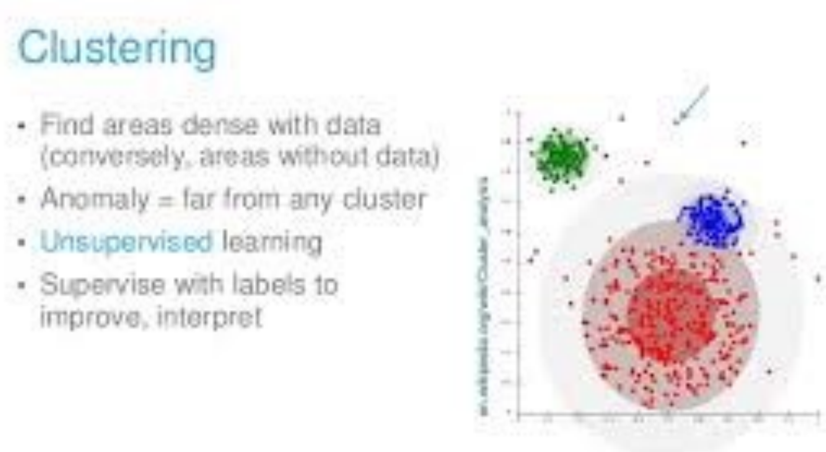
- **Unrepresentative labelled data means we need new ideas, and potentially new algorithms.**
- **Higher order measurements of predictions is one way to proceed.**

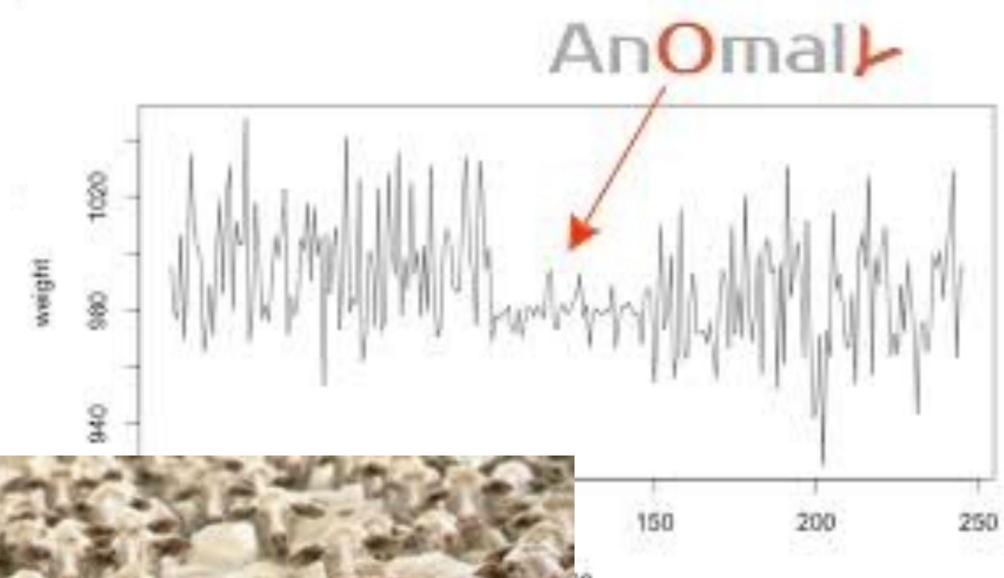
Cutting edge algorithms being implemented in astrophysics/cosmology Deep ML: CNNs / GANs.

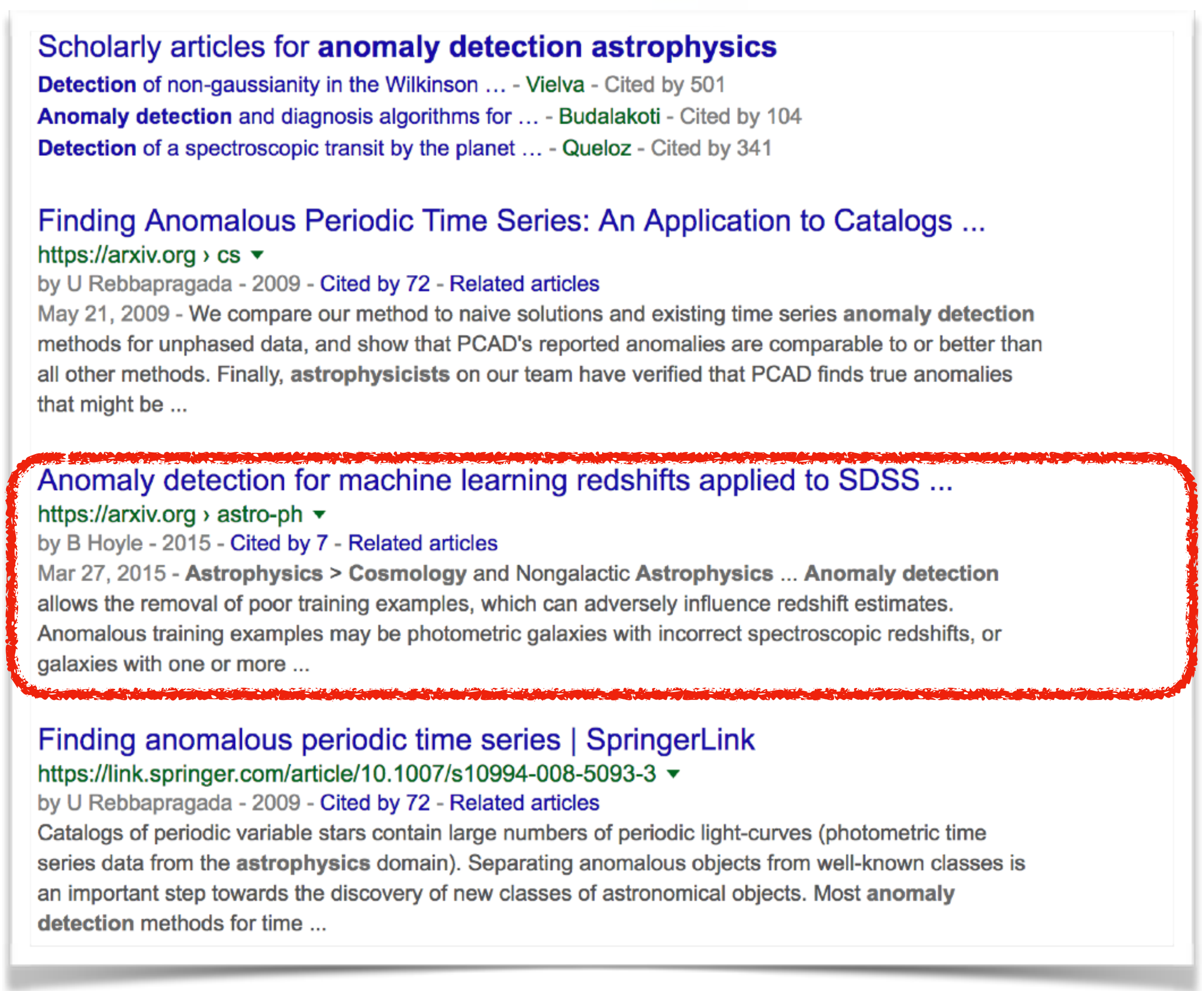
Shameless self-plug: benhoyle1212@gmail.com
I always have ML projects for dedicated students
I co-supervise PhD and Master students

What is Anomaly Detection?









Scholarly articles for **anomaly detection astrophysics**

Detection of non-gaussianity in the Wilkinson … - Vielva - Cited by 501

Anomaly detection and diagnosis algorithms for … - Budalakoti - Cited by 104

Detection of a spectroscopic transit by the planet … - Queloz - Cited by 341

Finding Anomalous Periodic Time Series: An Application to Catalogs ...

https://arxiv.org › cs

by U Rebbapragada - 2009 - Cited by 72 - Related articles

May 21, 2009 - We compare our method to naive solutions and existing time series **anomaly detection** methods for unphased data, and show that PCAD's reported anomalies are comparable to or better than all other methods. Finally, **astrophysicists** on our team have verified that PCAD finds true anomalies that might be ...

Anomaly detection for machine learning redshifts applied to SDSS ...

https://arxiv.org › astro-ph

by B Hoyle - 2015 - Cited by 7 - Related articles

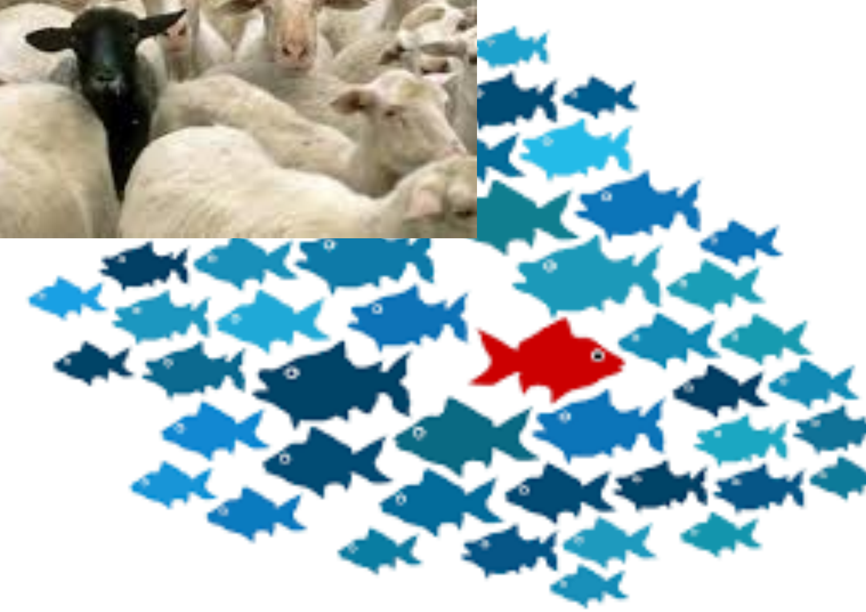
Mar 27, 2015 - **Astrophysics** > **Cosmology** and Nongalactic **Astrophysics** ... **Anomaly detection** allows the removal of poor training examples, which can adversely influence redshift estimates. Anomalous training examples may be photometric galaxies with incorrect spectroscopic redshifts, or galaxies with one or more ...

Finding anomalous periodic time series | SpringerLink

https://link.springer.com/article/10.1007/s10994-008-5093-3

by U Rebbapragada - 2009 - Cited by 72 - Related articles

Catalogs of periodic variable stars contain large numbers of periodic light-curves (photometric time series data from the **astrophysics** domain). Separating anomalous objects from well-known classes is an important step towards the discovery of new classes of astronomical objects. Most **anomaly detection** methods for time ...



ML dictionary

[Input] Features (X): — the input quantities (or independent variables), which are often easily measured for all data.

Targets (Y): The things we will want to predict (dependent variables), and have been measured for a subset of data.

Training: Fitting (or learning) a function to the training data, which maps features to targets Y = f(X)

Hyper-parameters: The tuning components of an algorithm, which modify its behaviour.

Training data: The data used to fit the algorithm

Validation data: The independent data used after training to check how the hyper-parameter choices have changed the predictive power.

Test data: A final independent data sample with target values, used to measure predictive power after all hyper-parameters have been fixed.

Science sample: the data set without target values, that we want to make predictions on.

Deep machine learning

keras.io http://caffe.berkeleyvision.org/ pylearn2 torch

Deep machine learning

keras.io http://caffe.berkeleyvision.org/ pylearn2 torch

Convolution neural networks (>2006):Images: Cat or Dog? Fbook face recognition.

Deep machine learning

keras.io http://caffe.berkeleyvision.org/ pylearn2 torch

Convolution neural networks (>2006):Images: Cat or Dog? Fbook face recognition.

Deep Recurrent neural networks: google translate, Shakespeare

http://karpathy.github.io/2015/05/21/rnn-effectiveness/

Deep machine learning

keras.io http://caffe.berkeleyvision.org/ pylearn2 torch

Convolution neural networks (>2006):Images: Cat or Dog? Fbook face recognition.

Reinforcement learning: AlphaGo, PacMan https://gym.openai.com/

http://karpathy.github.io/2016/05/31/rl/

Deep Recurrent neural networks: google translate, Shakespeare

http://karpathy.github.io/2015/05/21/rnn-effectiveness/

Deep machine learning

keras.io http://caffe.berkeleyvision.org/ pylearn2 torch

Convolution neural networks (>2006):Images: Cat or Dog? Fbook face recognition.

Reinforcement learning: AlphaGo, PacMan https://gym.openai.com/
http://karpathy.github.io/2016/05/31/rl/

Deep Recurrent neural networks: google translate, Shakespeare
http://karpathy.github.io/2015/05/21/rnn-effectiveness/

Standard machine learning

scikit-learn.org

Random forests, neural networks (ala 1980-2005's), support vector machines, K-means clustering, self organising maps

Deep machine learning

keras.io http://caffe.berkeleyvision.org/ pylearn2 torch

Convolution neural networks (>2006):Images: Cat or Dog? Fbook face recognition.

Reinforcement learning: AlphaGo, PacMan https://gym.openai.com/
http://karpathy.github.io/2016/05/31/rl/

Deep Recurrent neural networks: google translate, Shakespeare
http://karpathy.github.io/2015/05/21/rnn-effectiveness/

Standard machine learning

scikit-learn.org

Random forests, neural networks (ala 1980-2005's), support vector machines, K-means clustering, self organising maps

Supervised machine learning

You have some subset of data that you know the "truth" values for.

Regression / Classification

Regression: predict a floating point number. e.g. Galaxy redshifts.
Classification: prediction an integer. E.g. star/galaxy/quasar classfctn.

Deep machine learning

keras.io http://caffe.berkeleyvision.org/ pylearn2 torch

Convolution neural networks (>2006):Images: Cat or Dog? Fbook face recognition.

Reinforcement learning: AlphaGo, PacMan https://gym.openai.com/
http://karpathy.github.io/2016/05/31/rl/

Deep Recurrent neural networks: google translate, Shakespeare
http://karpathy.github.io/2015/05/21/rnn-effectiveness/

Standard machine learning

scikit-learn.org

Random forests, neural networks (ala 1980-2005's), support vector machines, K-means clustering, self organising maps

Supervised machine learning

You have some subset of data that you know the "truth" values for.

Regression / Classification

Regression: predict a floating point number. e.g. Galaxy redshifts.
Classification: prediction an integer. E.g. star/galaxy/quasar classfctn.

Unsupervised machine learning

You don't have a subset of data that you know the "truth" values for [Or you don't care].

Data Clustering algorithms: How can you cluster your data into "like" objects

Anomalous data identification: Does any new data look very different from the data you already have. E.g., fraud detection.

Overview

The supervised Machine Learning (ML) framework

When to use the ML hammer

An introduction to photometric redshifts

My typical ML workflow

A common ML application:

Photometric redshifts

The biggest problem for ML in cosmology:

Unrepresentative labelled data

Dealing with unrepresentative labelled data

Other common applications of ML

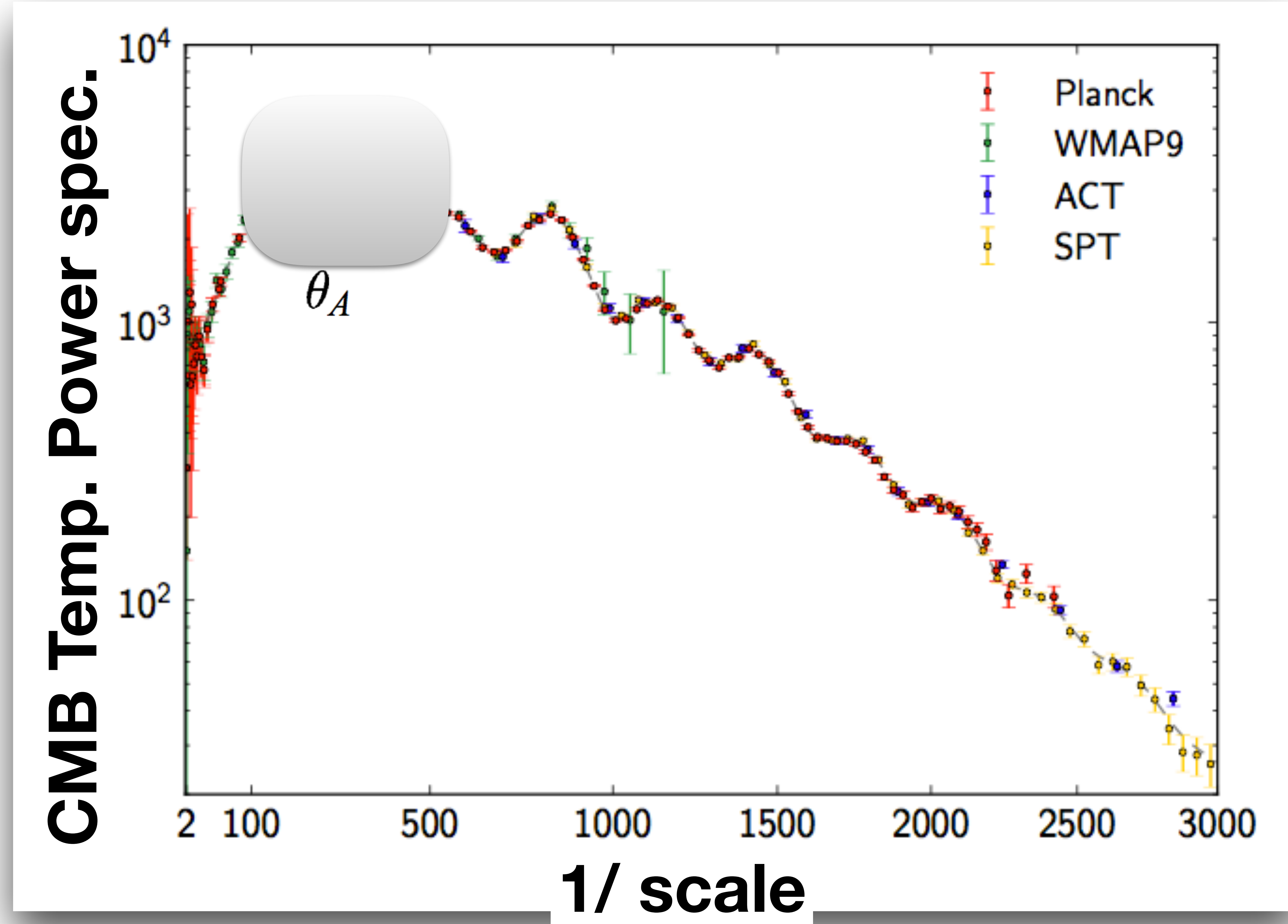
Recent, novel applications of ML

Conclusions

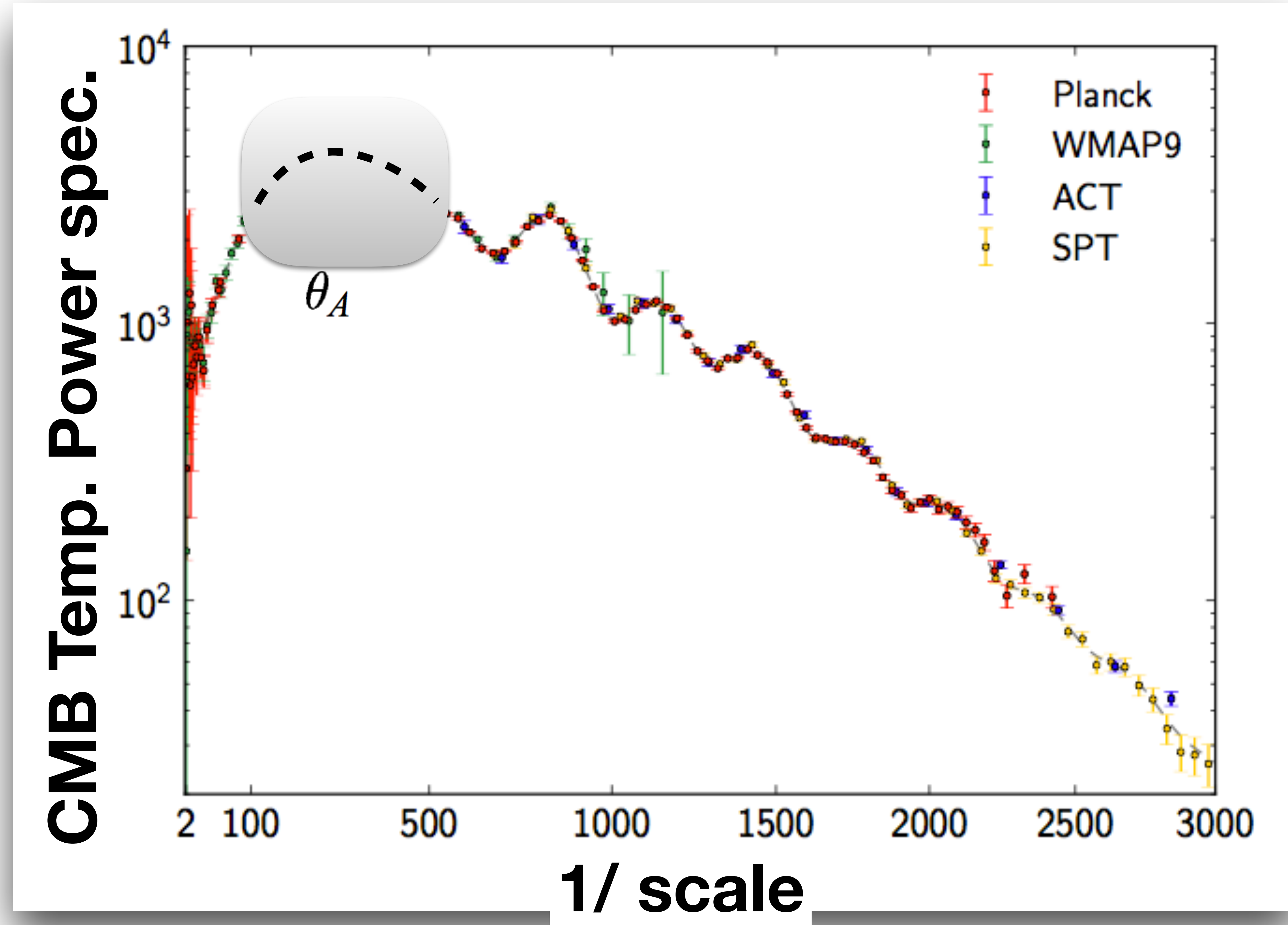
When/why is ML suited to astrophysics/ cosmology?

When we are in a "data poor" and "model rich" regime e.g. Correlation function analysis of Cosmic Microwave Background maps, we should not use ML, rather rely on the predictive model [s].

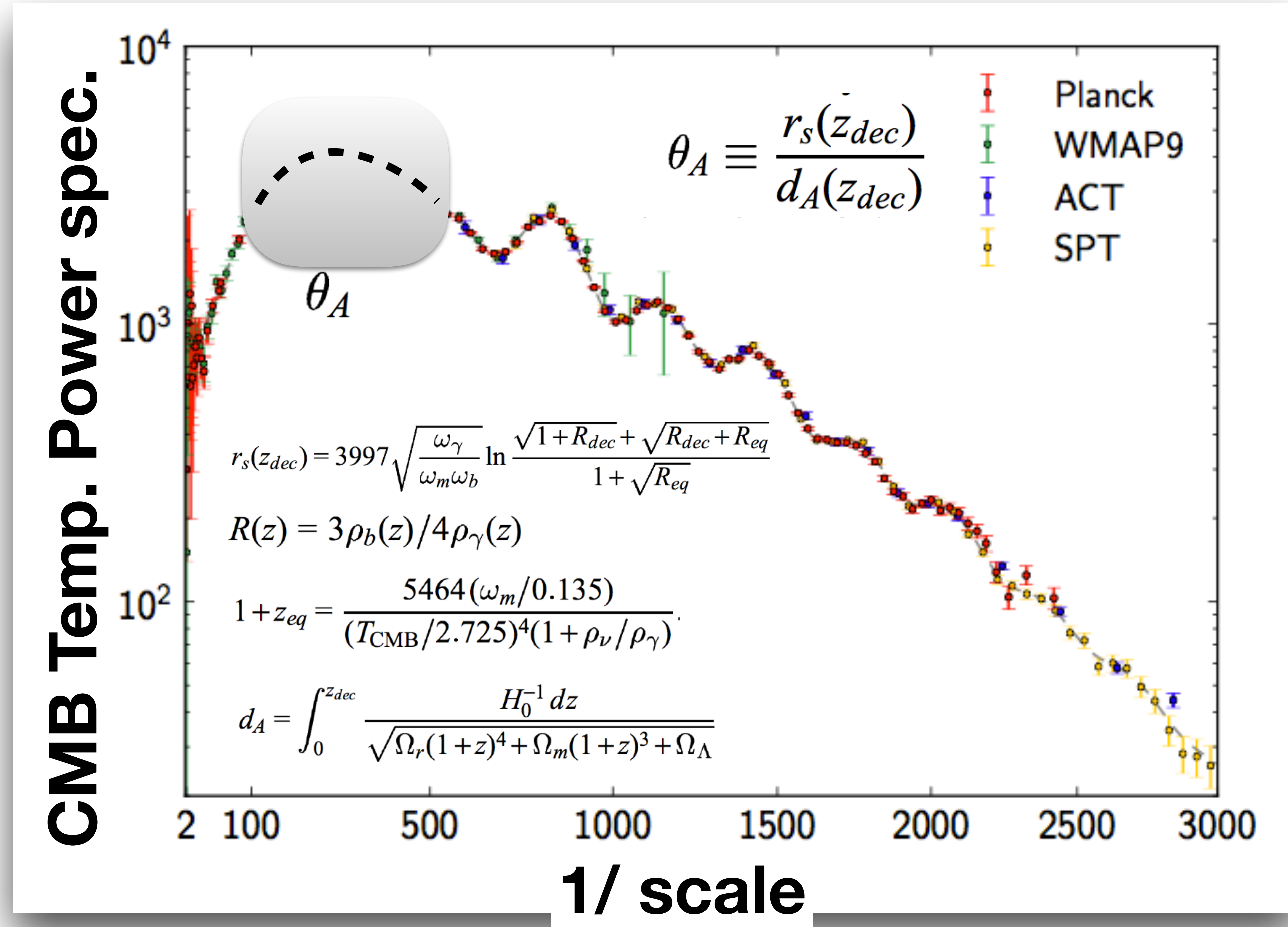
When *not* to use the ML hammer?



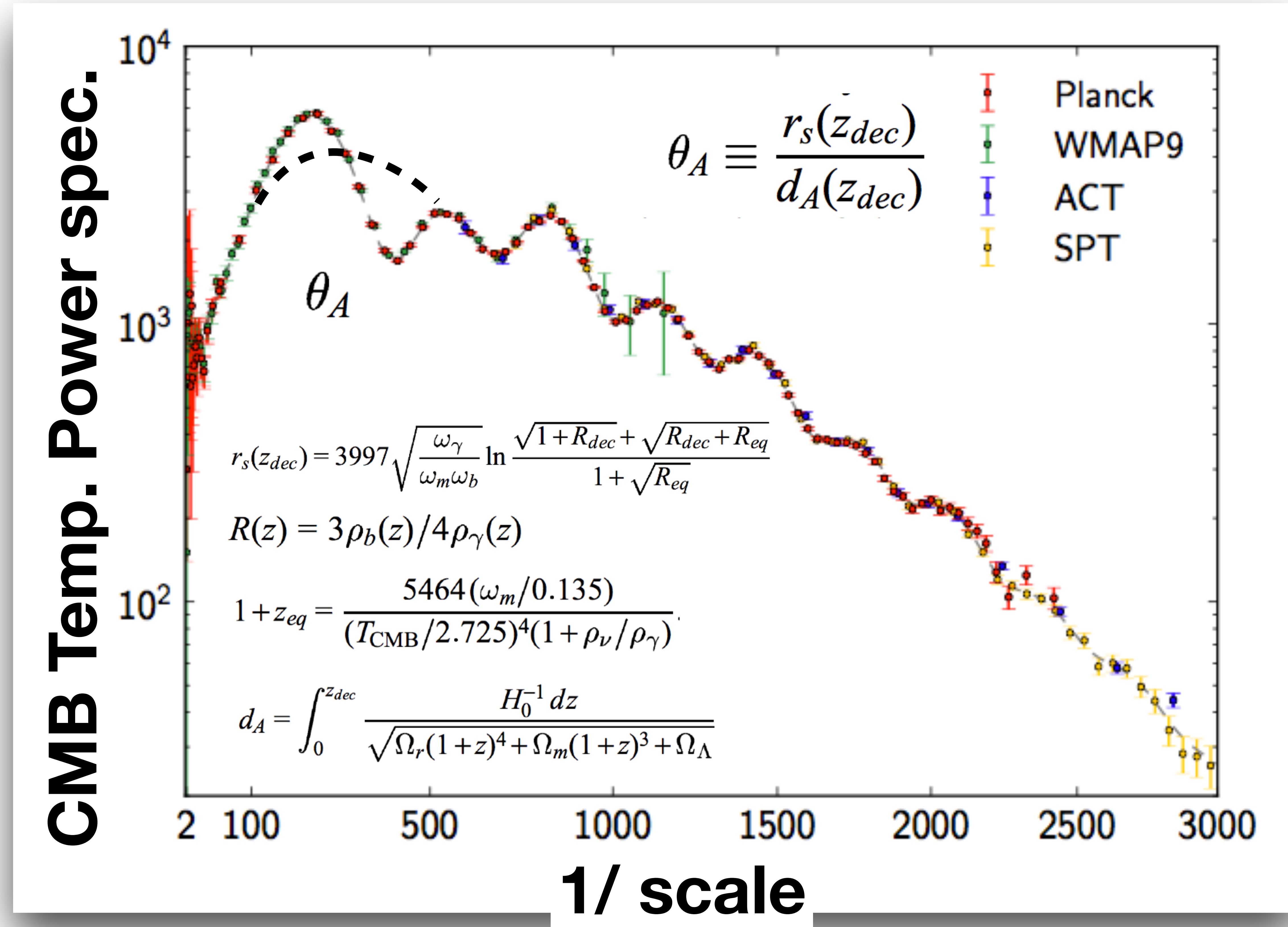
When *not* to use the ML hammer?



When *not* to use the ML hammer?



When *not* to use the ML hammer?



If you have an accurate predictive model, you should use it if possible

When/why is ML suited to astrophysics/ cosmology?

When we are in a "data poor" and "model rich" regime e.g. Correlation function analysis of Cosmic Microwave Background maps, we should not use ML, rather rely on the predictive model [s].

When we are in a "data rich" and "model poor" regime, and still want to approximate some unknown model y=f(x); we can use machine learning to learn (or fit) an arbitrarily complex model (e.g. non-functional curves) of the data.

When/why is ML suited to astrophysics/ cosmology?

When we are in a "data poor" and "model rich" regime e.g. Correlation function analysis of Cosmic Microwave Background maps, we should not use ML, rather rely on the predictive model [s].

When we are in a "data rich" and "model poor" regime, and still want to approximate some unknown model y=f(x); we can use machine learning to learn (or fit) an arbitrarily complex model (e.g. non-functional curves) of the data.

Cosmology is firmly in the data "rich" regime:

1) SDSS has 100 million photometrically identified objects (stars/galaxies) and 3 million spectroscopic "truth" values, for e.g. redshift, and galaxy/ stellar type

2) DES has 300 million objects with photometry, and ~400k objects with spectra

3) Gaia has >1.2 billion sources [stellar maps of the Milky Way]

4) Euclid with have 3 billion objects...

When/why is ML suited to astrophysics/ cosmology?

When we are in a "data poor" and "model rich" regime e.g. Correlation function analysis of Cosmic Microwave Background maps, we should not use ML, rather rely on the predictive model [s].

When we are in a "data rich" and "model poor" regime, and still want to approximate some unknown model y=f(x); we can use machine learning to learn (or fit) an arbitrarily complex model (e.g. non-functional curves) of the data.

Cosmology is firmly in the data "rich" regime, and often in the "model-poor" regime:

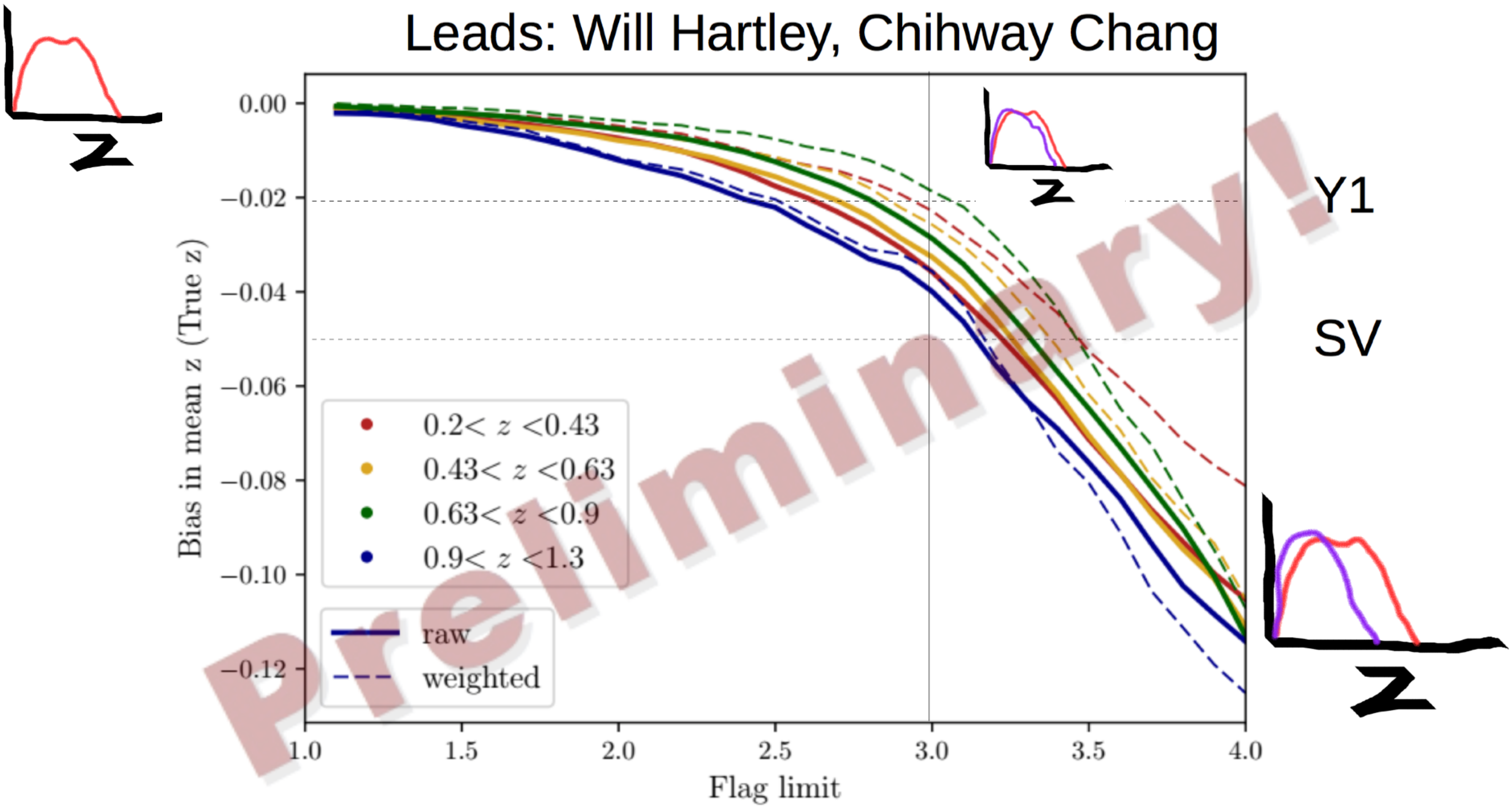
1) The exact mapping between galaxies observed in long exposure photographs and their true distance (redshift) depends on stellar population physics, initial stellar mass functions, feedback from exploding stars and black holes, the dust in our galaxy,...

2) Is an object found in photometric images a faint star in our galaxy, or a high redshift galaxy?

Redshifting is tough!

We extract 1-d spectra from simulations (known redshift), added realistic noise. Ask observers to redshift the spectra, using their common analysis tools.

Leads: Will Hartley, Chihway Chang



We cannot validate photo-z performance on data which is biased w.r.t the science sample

Feature pre-selection

We are now swimming in ‘M’ input features. Most algorithms don’t work well with many 10’s or 100’s of input features. Which ’S’ of those ’M’ features should we use?

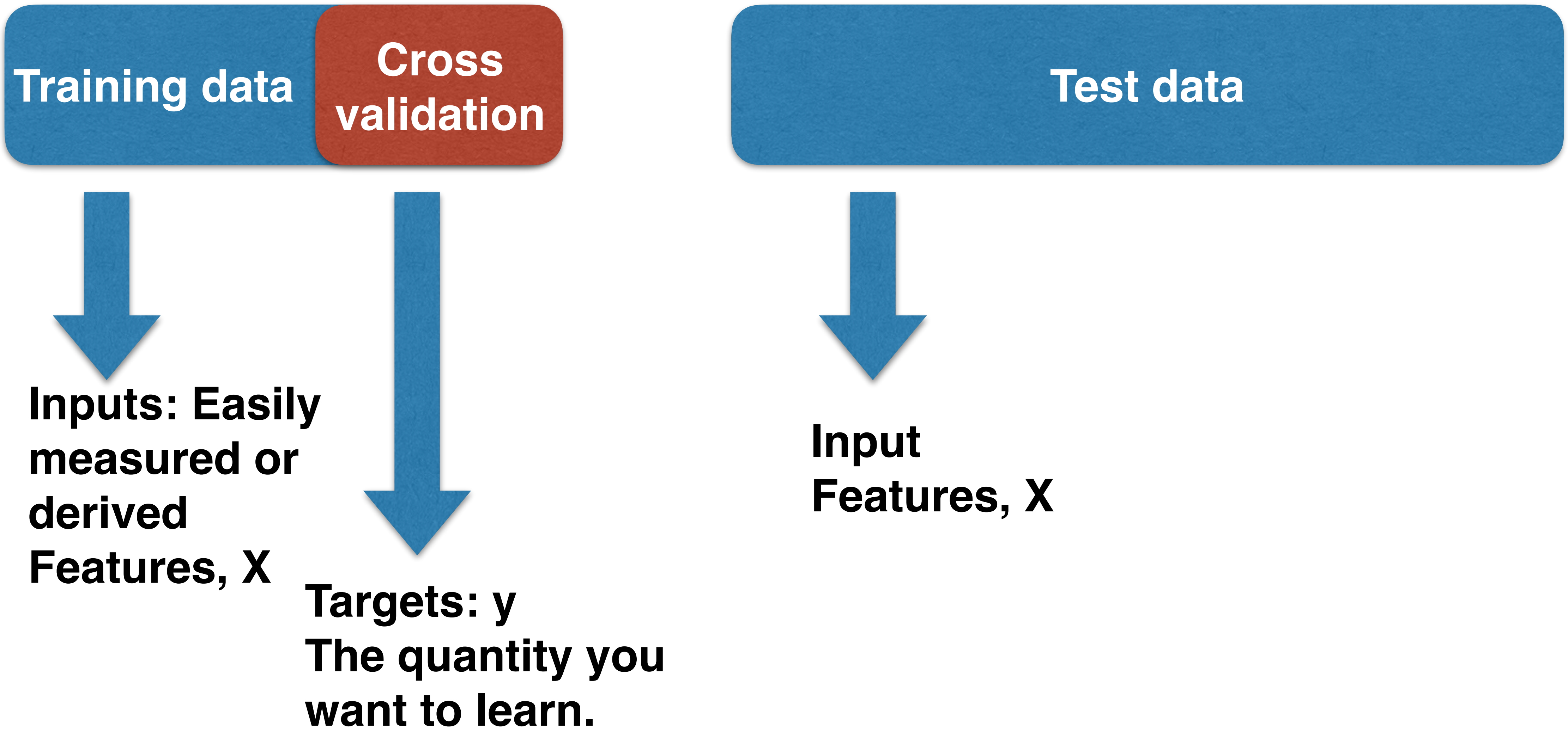
We could give everything to a Random Forest + feature importance.

- **Memory Limitations.**
- **Correlations between features ignored.**
- **Shape of the “test/science sample data” is ignored.**

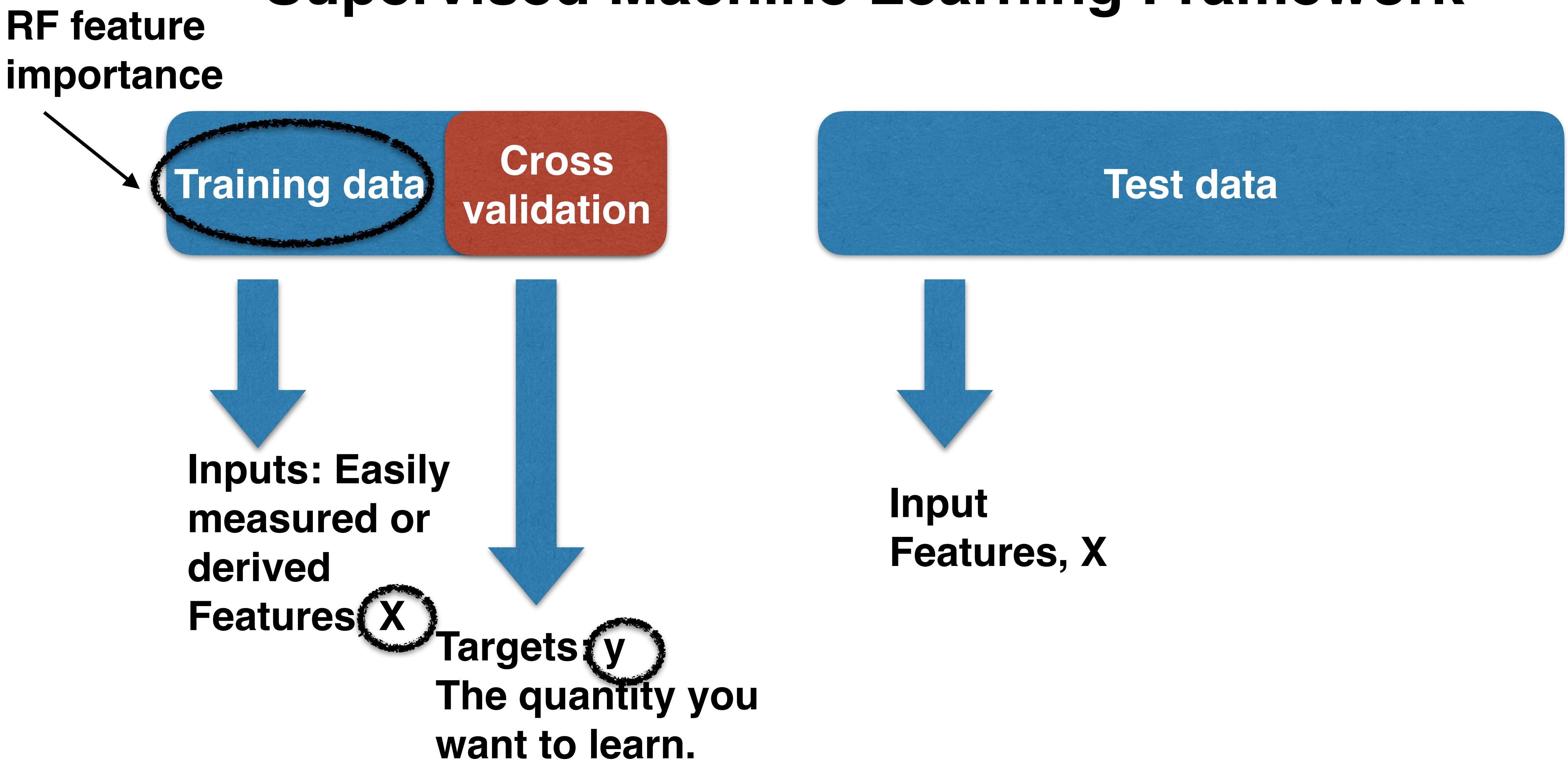
Feature pre-selection: MINT

Which features should I choose to feed into my algorithms?
https://arxiv.org/abs/1310.1659

Supervised Machine Learning Framework



Supervised Machine Learning Framework



Supervised Machine Learning Framework

RF feature importance

Training data

Cross validation

Test data

Inputs: Easily measured or derived Features, X

Targets: y
The quantity you want to learn.

Input Features, X

Relevant (RL) features

$$RL = \frac{1}{S}\sum_{i} MI(X_i;Y)$$

Supervised Machine Learning Framework

RF feature importance

Training data

Cross validation

Test data

Inputs: Easily measured or derived Features, X

Targets: y
The quantity you want to learn.

Input Features, X

$$MI(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right)$$

Relevant (RL) features

$$RL = \frac{1}{S} \sum_i MI(X_i;Y)$$

Supervised Machine Learning Framework

RF feature importance

Training data

Cross validation

Test data

Inputs: Easily measured or derived Features, X

Targets: y
The quantity you want to learn.

Input Features, X

$$MI(X;Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right)$$

Relevant (RL) features

$$RL = \frac{1}{S} \sum_{i} MI(X_i;Y)$$

Redundant (RD) features

$$RD = \frac{1}{S^2} \sum_{i,j} MI(X_i;X_j)$$

Supervised Machine Learning Framework

RF feature importance

Training data

Cross validation

Test data

Inputs: Easily measured or derived Features, X

Targets: y
The quantity you want to learn.

Input Features, X

$$MI(X;Y) = \sum_{x\in X}\sum_{y\in Y} p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right)$$

Relevant (RL) features

$$RL = \frac{1}{S}\sum_{i} MI(X_i;Y)$$

Redundant (RD) features

$$RD = \frac{1}{S^2}\sum_{i,j} MI(X_i;X_j)$$

$$\Phi_S(RL,RD) = RL - RD$$

(also Peng et al. 2005)

Choose S features, so that we maximise Φ_S

Explore all M! / (M-S)! combinations of features using a greedy search algorithm.

Feature pre-selection

We are now swimming in 'M' input features. Most algorithms don't work well with many 10's or 100's of input features. Which of those 'S' generated features should we use?

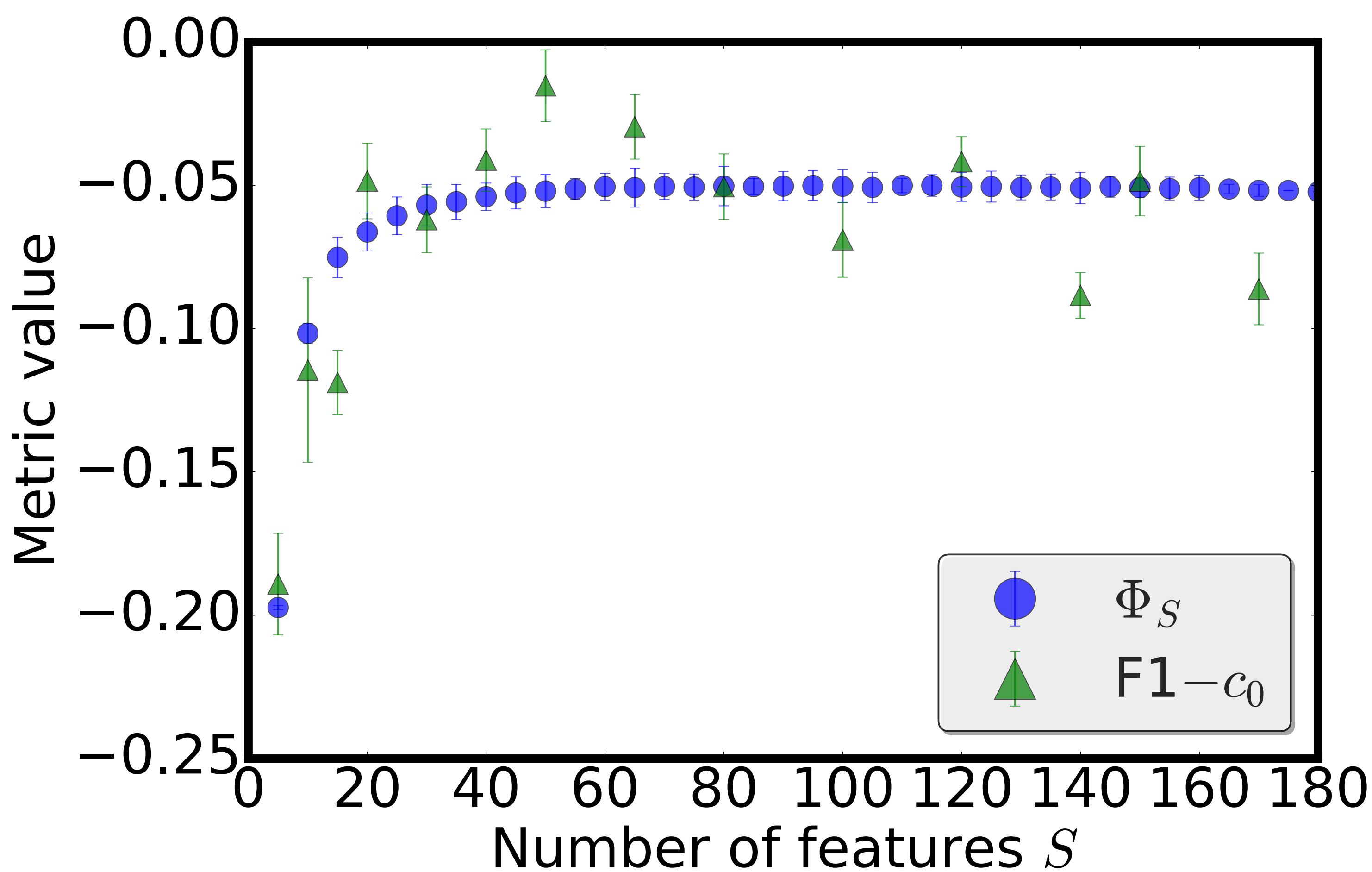
We could give everything to a Random Forest

— Correlations between features ignored.

— Shape of the "test input data" is ignored.

Mutual INformation based Transductive feature selection MINT (He et al 2013)

— Identify the set of 'S' (out of 'M') features which have the largest combined correlation with the target (measured in the training data), and the smallest correlation with each other (as measured in the test data).



Recent GAN applications

GANs to peer within a galaxy image: sub PSF properties of galaxies. Schawinski et al 2017

GANs produce one realisation of what the input galaxy could look like.
http://space.ml/supp/GalaxyGAN.html

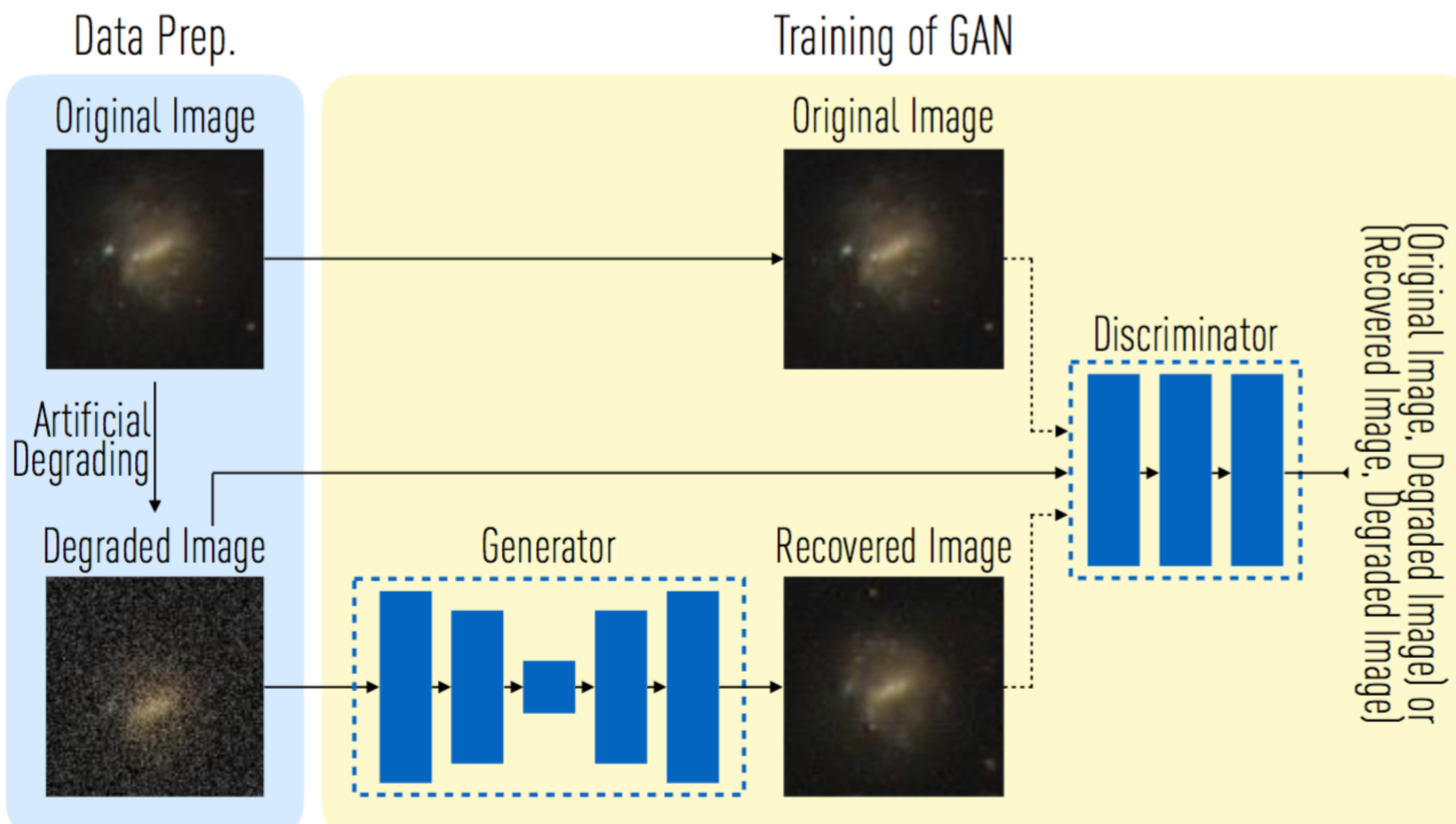
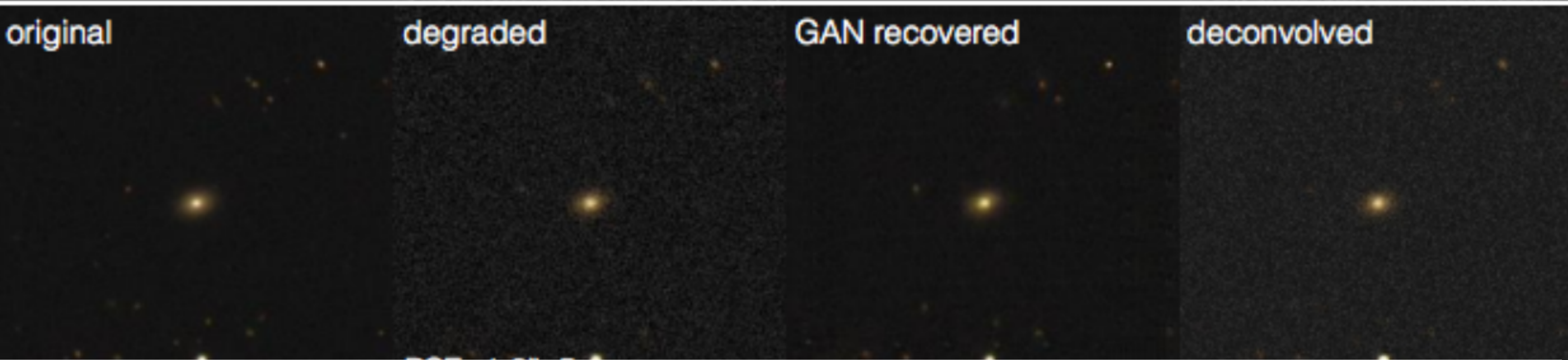


Figure 1. Schematic illustration of the training process of our method. The input is a set of original images. From these we automatically generate degraded images, and train a Generative Adversarial Network. In the testing phase, only the generator will be used to recover images.



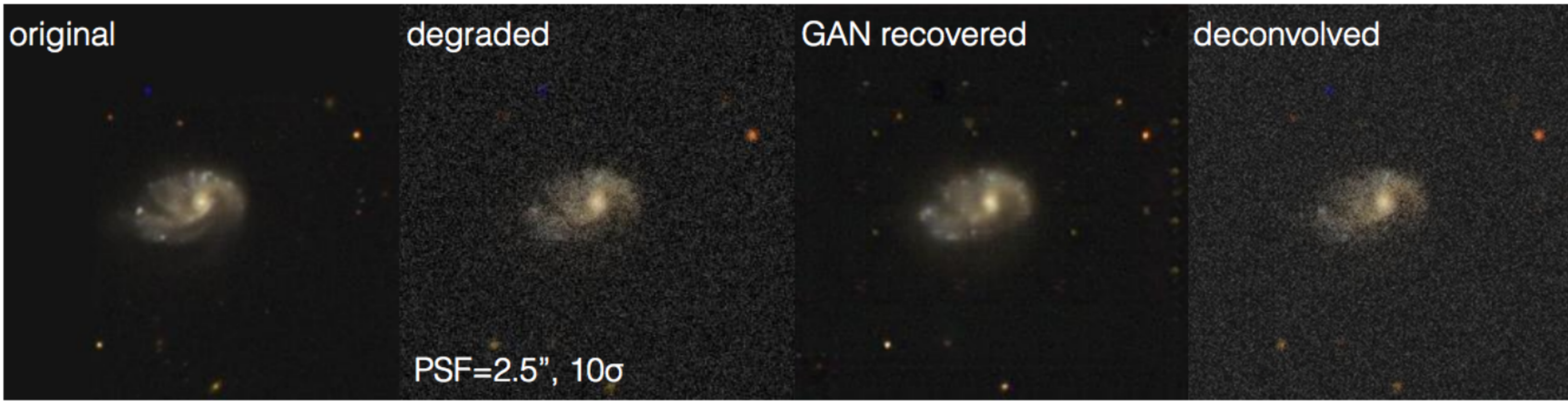
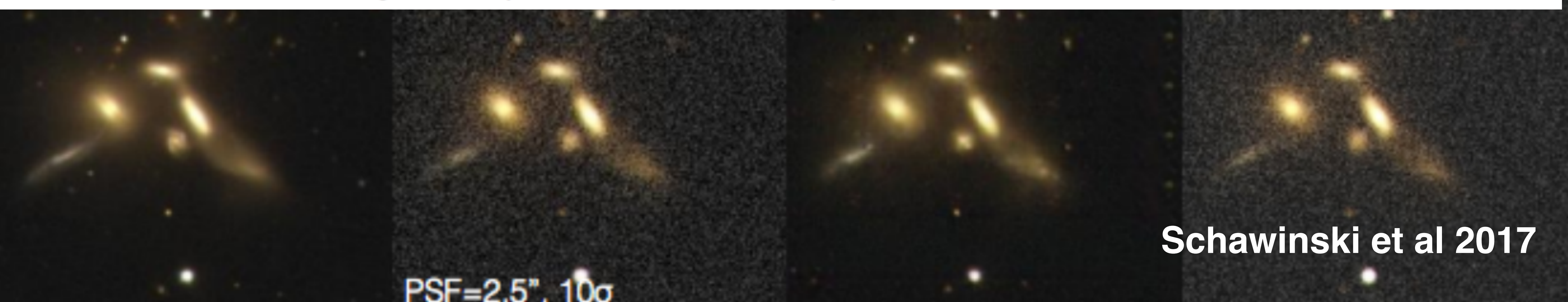


Figure 2. We show the results obtained for one example galaxy. From left to right: the original SDSS image, the degraded image with a worse PSF and higher noise level (indicating the PSF and noise level used), the image as recovered by the GAN, and for comparison, the result of a deconvolution. This figure visually illustrates the GAN's ability to recover features which conventional deconvolutions cannot.



Recent GAN applications

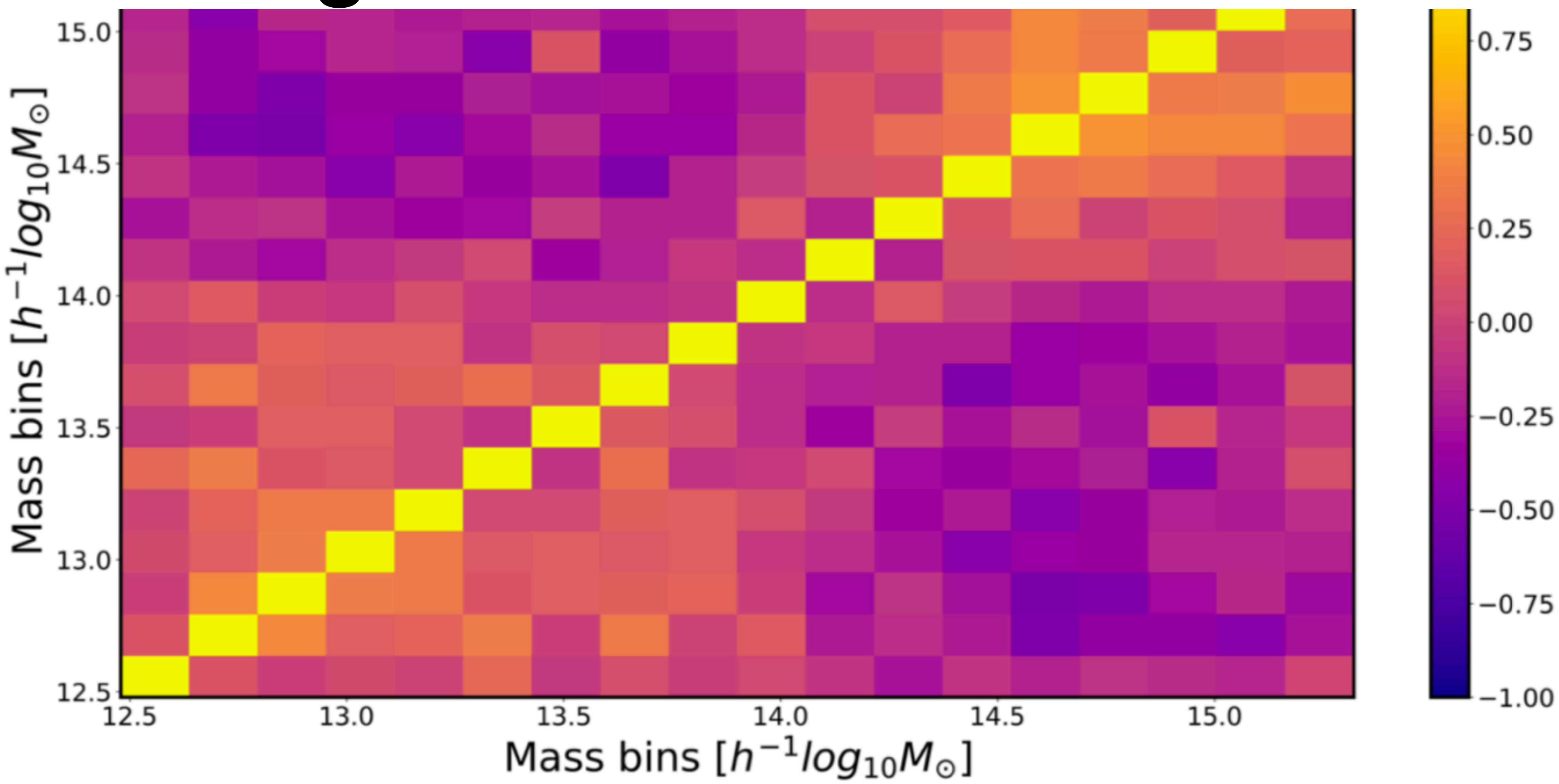
GANs to peer within a galaxy image: sub PSF properties of galaxies. Schawinski et al 2017

GANs produce one realisation of what the input galaxy could look like.
http://space.ml/supp/GalaxyGAN.html

Getting "labels" for the science sample data one cares about, is very challenging.

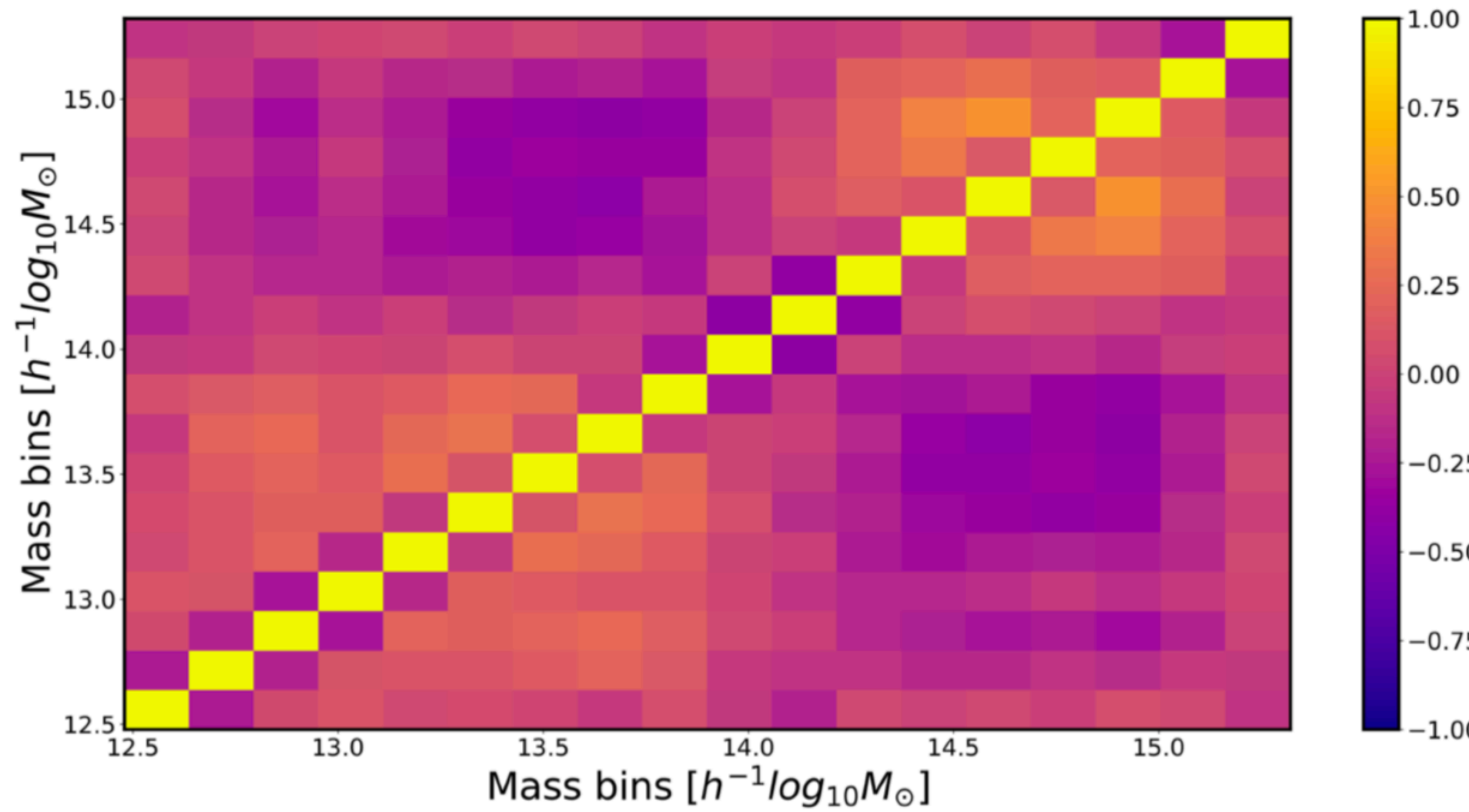
Again, move towards higher order measurements of the predicted signal:
E.g. does gas predicted to exist in some part of the galaxy/disk give off radiation which can be observed in other bands?

GANs generate realisations of a Dark-Matter N-body sim.



(a) Correlation matrix for the voxel mass count from the training set

The covariance matrix of the abundances of masses look reasonable



(b) Correlation matrix for the voxel mass count from the GAN output

Master thesis Julien Wolf 2018