Mitigation of Photometric Systematics in Galaxy Clustering with Deep Learning

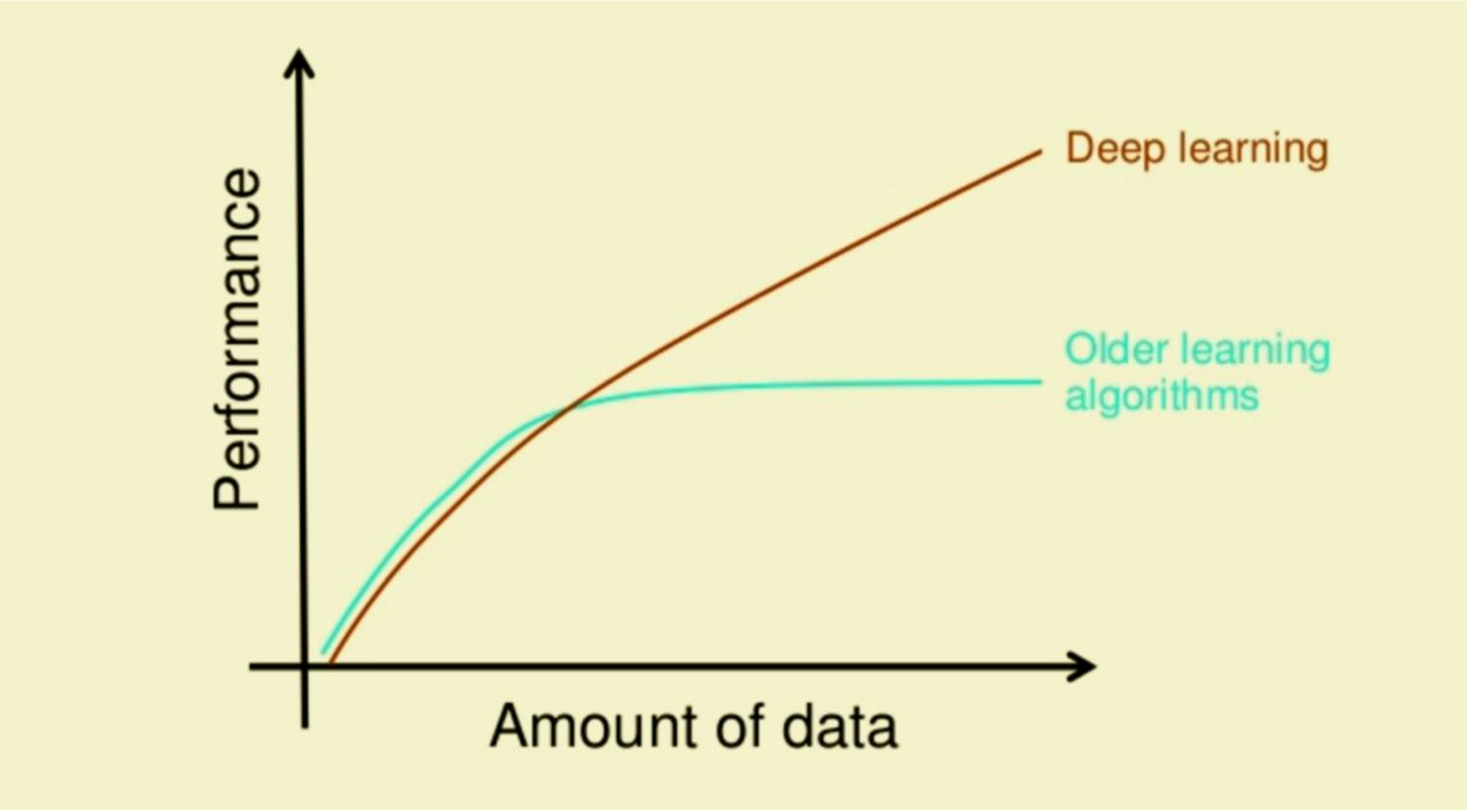
Mehdi Rezaie
Physics and Astronomy, Ohio University

Supervisor:
Hee-Jong Seo
In collaboration with:
Ashley J. Ross & Razvan Bunescu

DECam Community Science Workshop, Tucson, May 2018


THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS

Why Deep Learning?



Angular correlation function with systematics

Luminous Galaxies 0.4<z<0.7

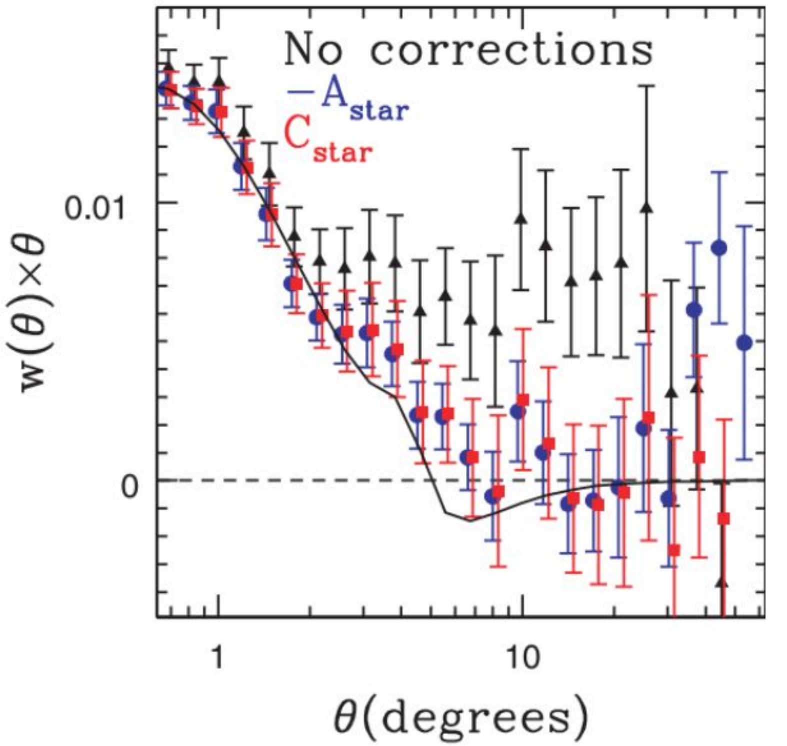
Example that stellar density affects the clustering signal at larger separation angles

© *A.J.Ross et. al, 2011*

Solid line ΛCDM with $\Omega_m = 0.27$

Black: no corrections
Blue: correcting for the area loss due to stars (A_{star})
Red: correcting for stellar density (n_{star})

Bottom line: We won't be able to get a robust clustering measurement unless we correct for systematic effects



Treatment of Systematics

5. NOVEL TREATMENT OF SYSTEMATICS

Assuming that residual systematics will exist best catalog we can construct without a serious loss of sky, how should we handle the remaining systematics? Without any evidence of possible non-linear effects of systematics on the observed density fields, we adopt the simplest approach: linear relationship between the systematics and the observed galaxy density fields.

We start from the following: Transforming fields from real space into spherical harmonic space (or l space in particular), so that $<\delta_g \delta_g>= C_l$:

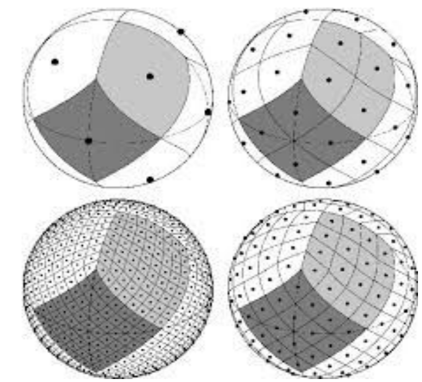
$$\delta_g^o = \delta_g^t + \sum_{i=1}^{N_{sys}} \epsilon_i \delta_i \quad (17)$$

Ho et. al. 2012

eBOSS with DR5

Table 1. eBOSS ELG Target Selection from Raichoor et al. (2017)

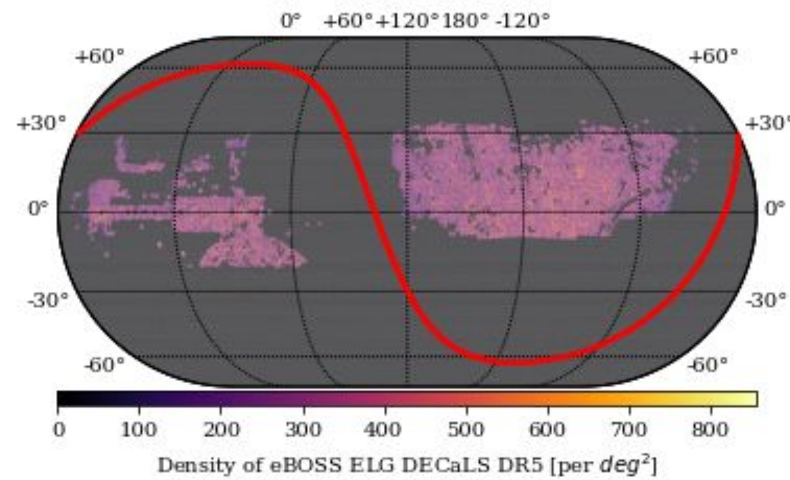
Criterion	eBOSS ELG
[OII] emitters	$21.825 < g < 22.825$
Redshift range	$-0.068(r\text{-}z) + 0.457 < g\text{-}r < 0.112\,(r\text{-}z) + 0.773$
	$0.218(g\text{-}r) + 0.571 < r\text{-}z < 0.555\,(g\text{-}r) + 1.901$



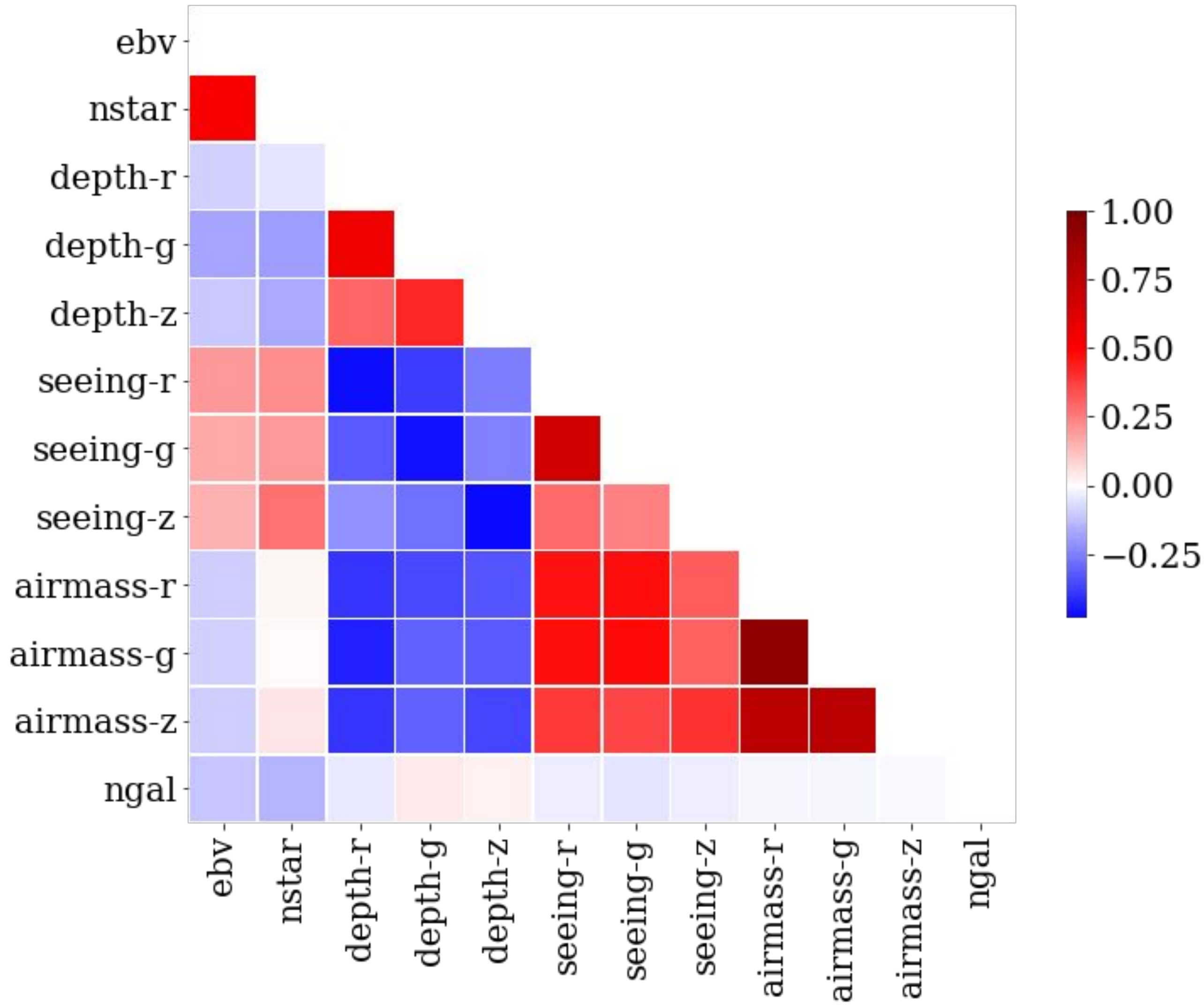
Our problem (*Regression*): model how the ELG density depends on imaging systematics

Use **HEALPIX** to pixelize sky
With **Nside = 256**, we have around **120k** healpix pixels

For each pixel, we have imaging systematics as "features"
And number of Emission Line Galaxies ELGs as "label"



Correlation Matrix of *systematics* and *ngal*



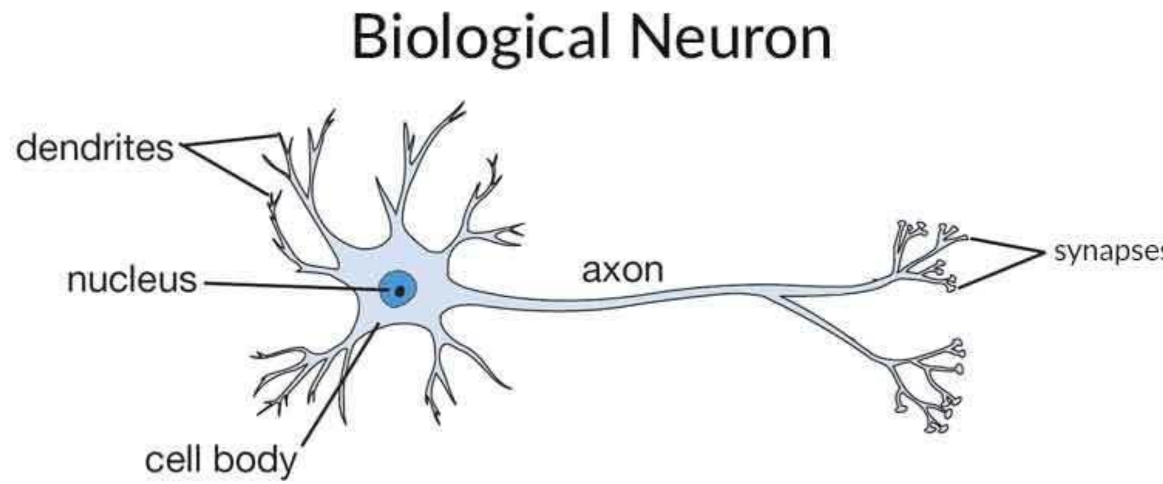
Artificial Neuron

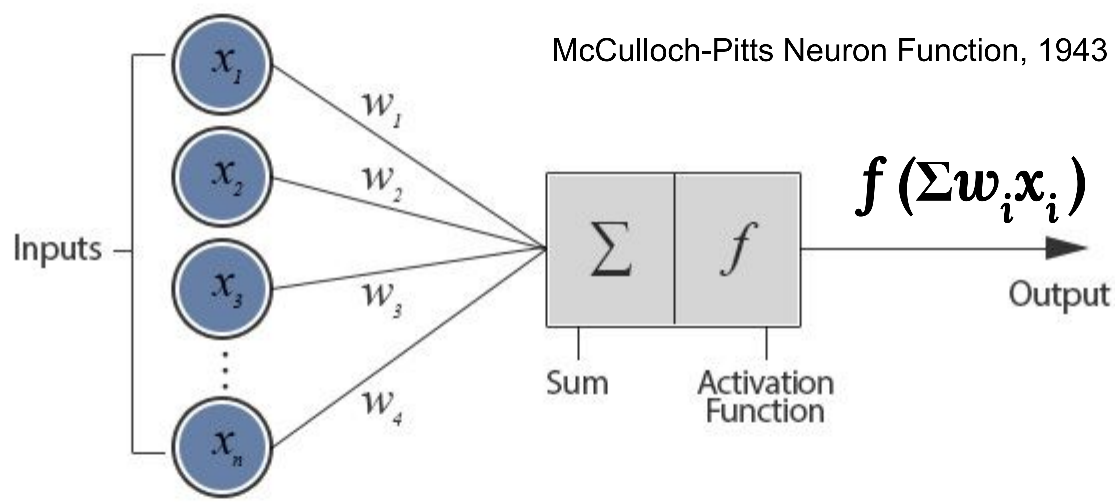
Inspired by Biological Neuron

The output is a non-linear function of the sum of the inputs

Can be used for regression; Estimating the relationships between variables

Universal Approximators
see *Hornik (1991), Cybenko (1989)*

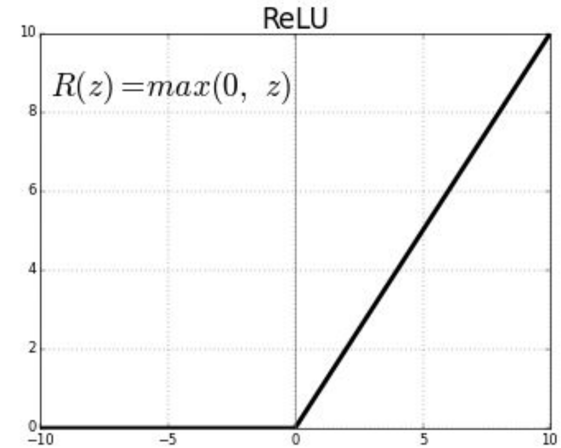




McCulloch-Pitts Neuron Function, 1943

Feed Forward Neural Net

11 features, 2 h. layers, 10 neurons on each layer



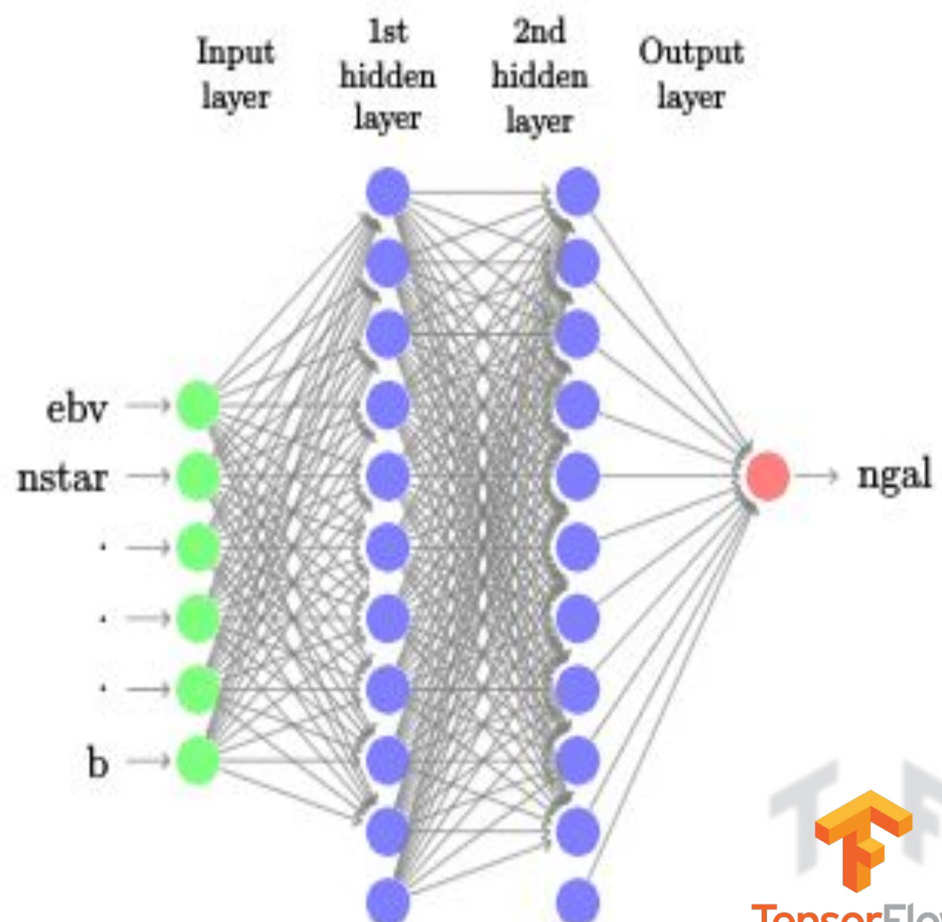
Inputs:
EBV, seeing/depth/airmass in rgz,
Stellar density 17<r<20

Output:
Number of galaxies per pixel

ReLU activation function on hidden layer neurons

Weights optimized with *ADAM* algorithm by minimizing the mean squared error

Num of parameters
(11+1)x10 + (10+1)x10 + 10+1

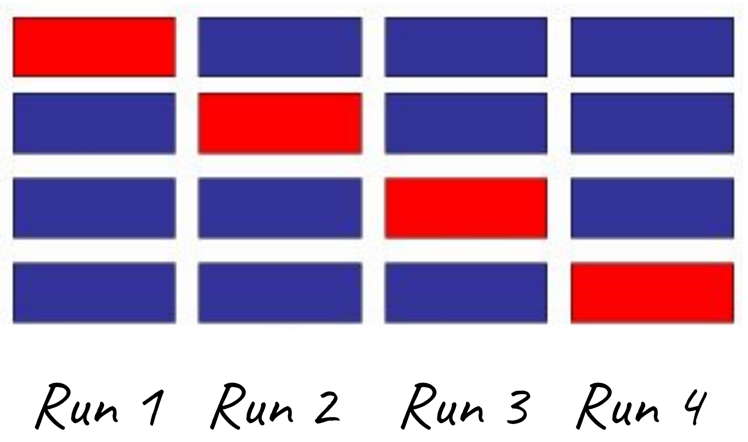


TensorFlow

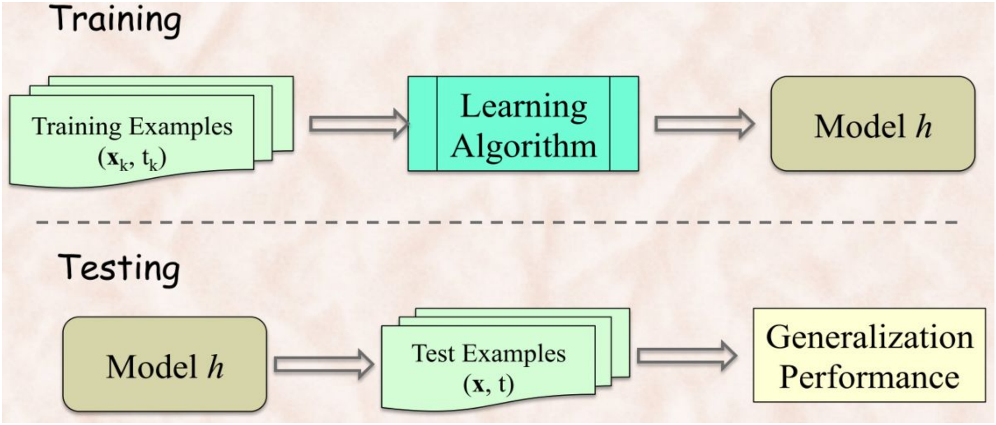
4-Fold Validation

Overfitting?

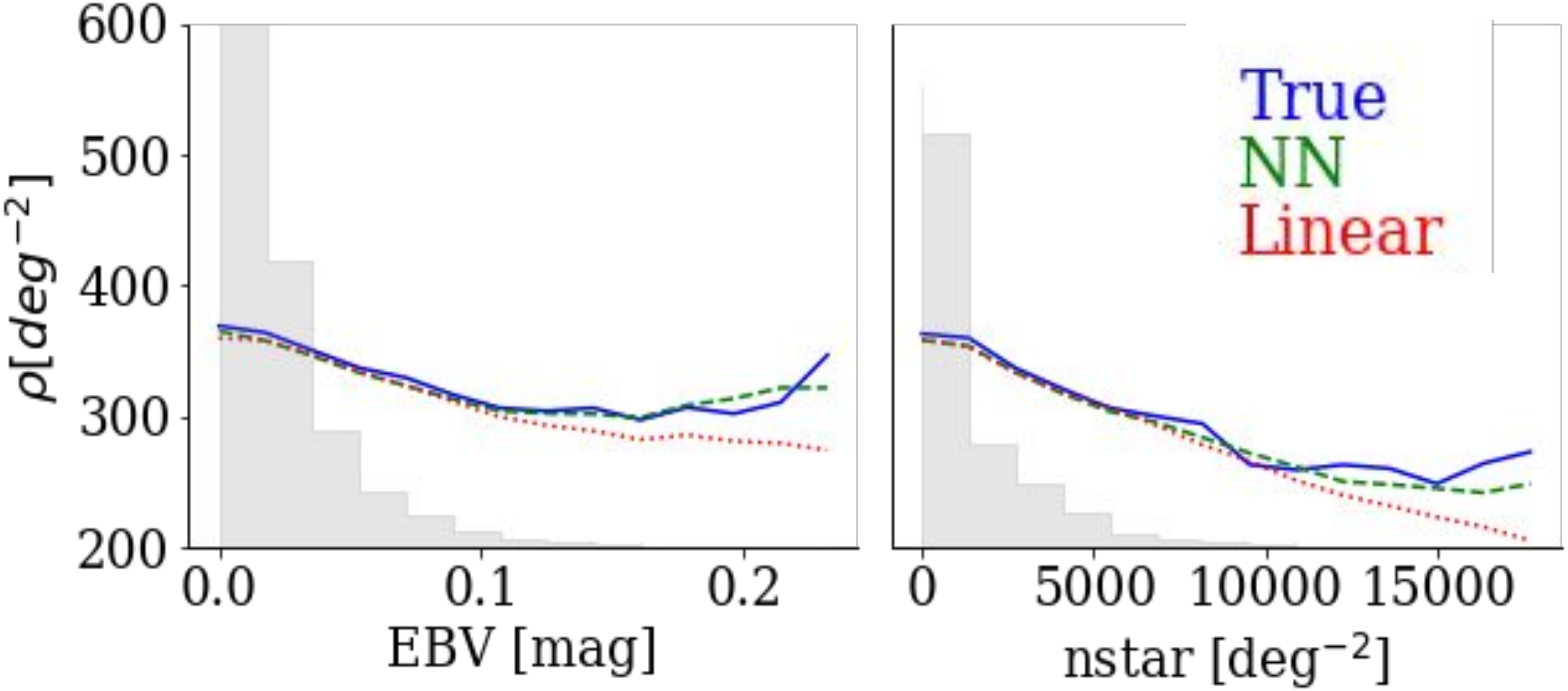
Split data into k chunks
3 chunks for training and 1 chunk for testing
Shuffle the chunks
Repeat 4 times



A technique to have prediction
for the entire footprint

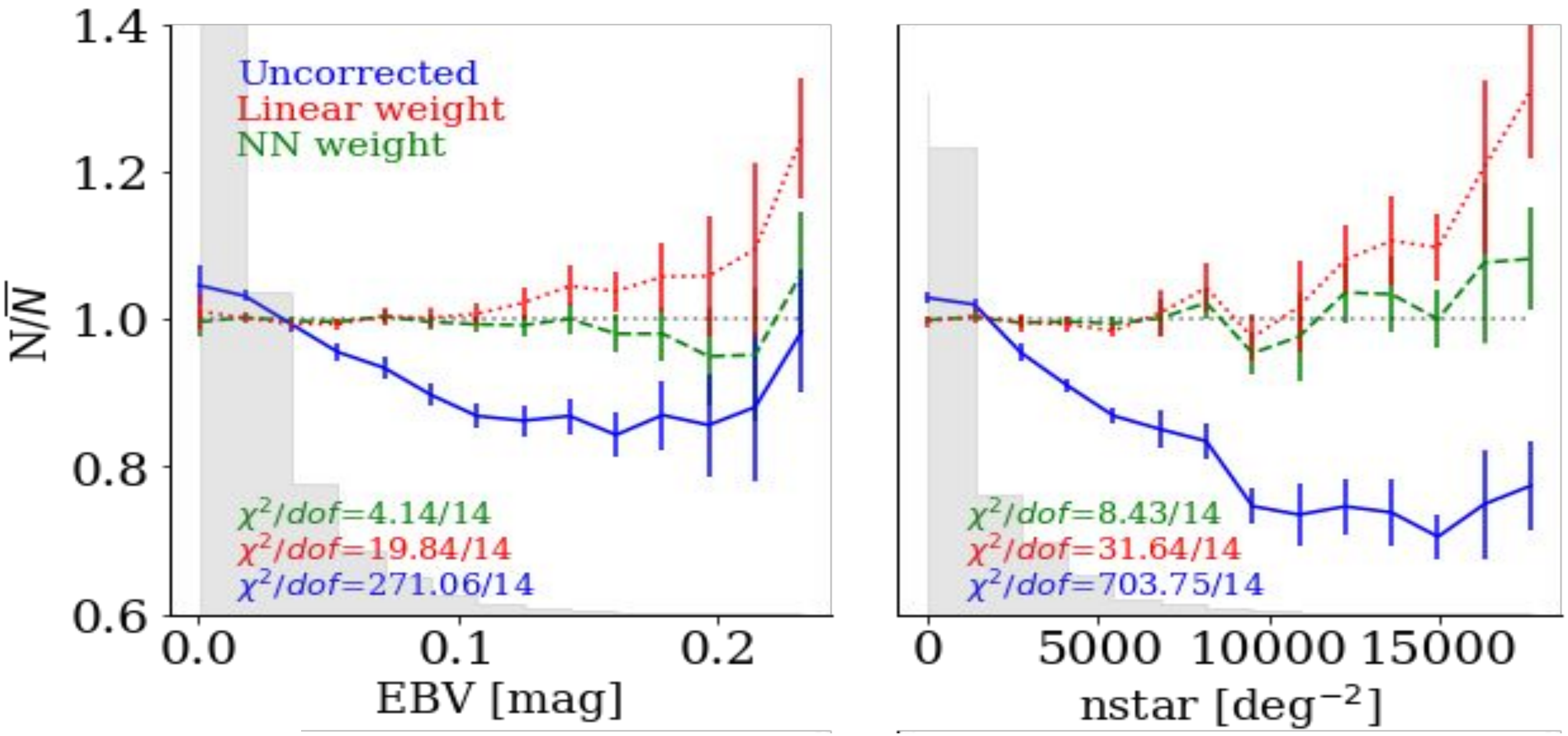


Average density in each bin of a particular systematic

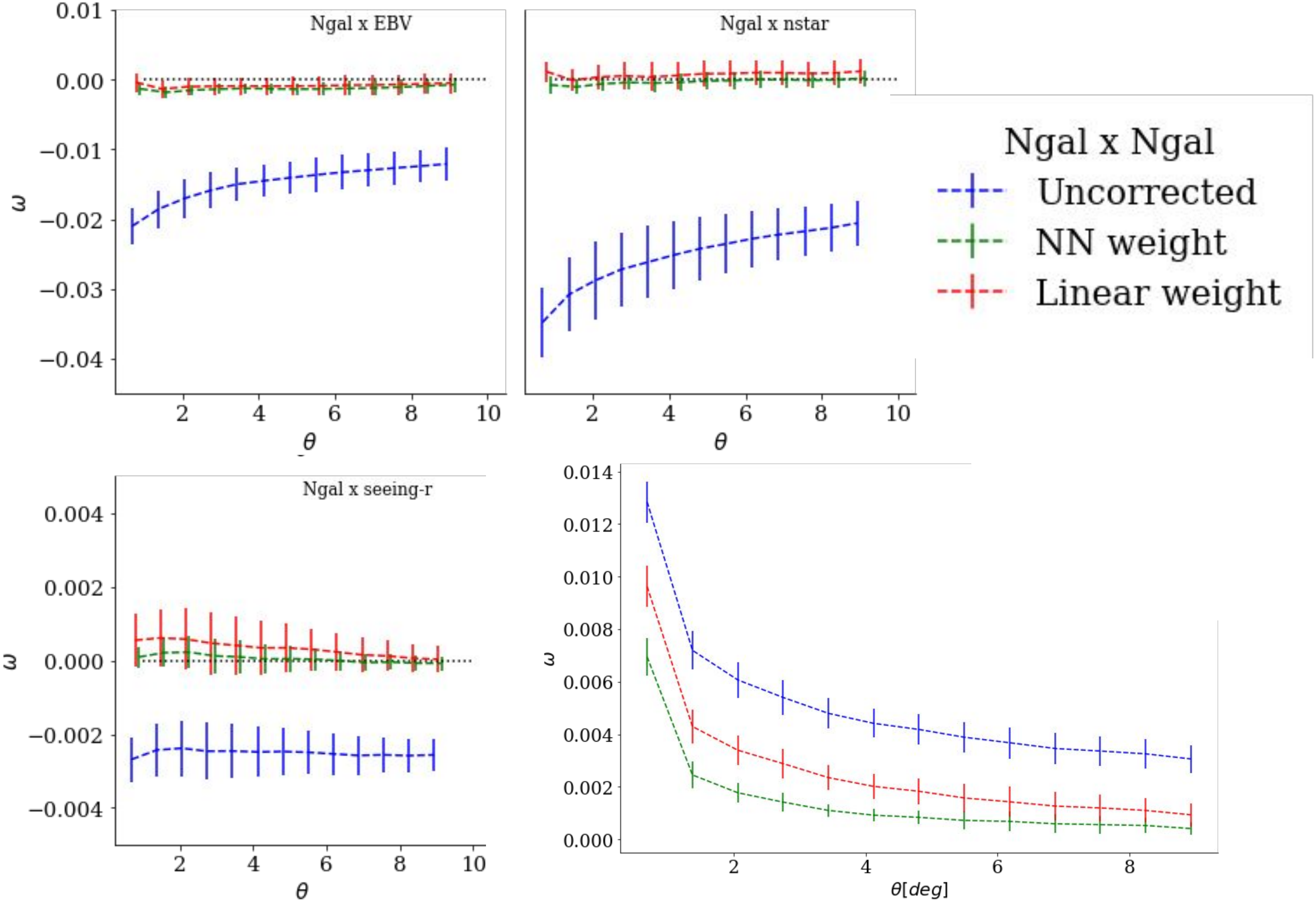


Fluctuation in galaxy counts

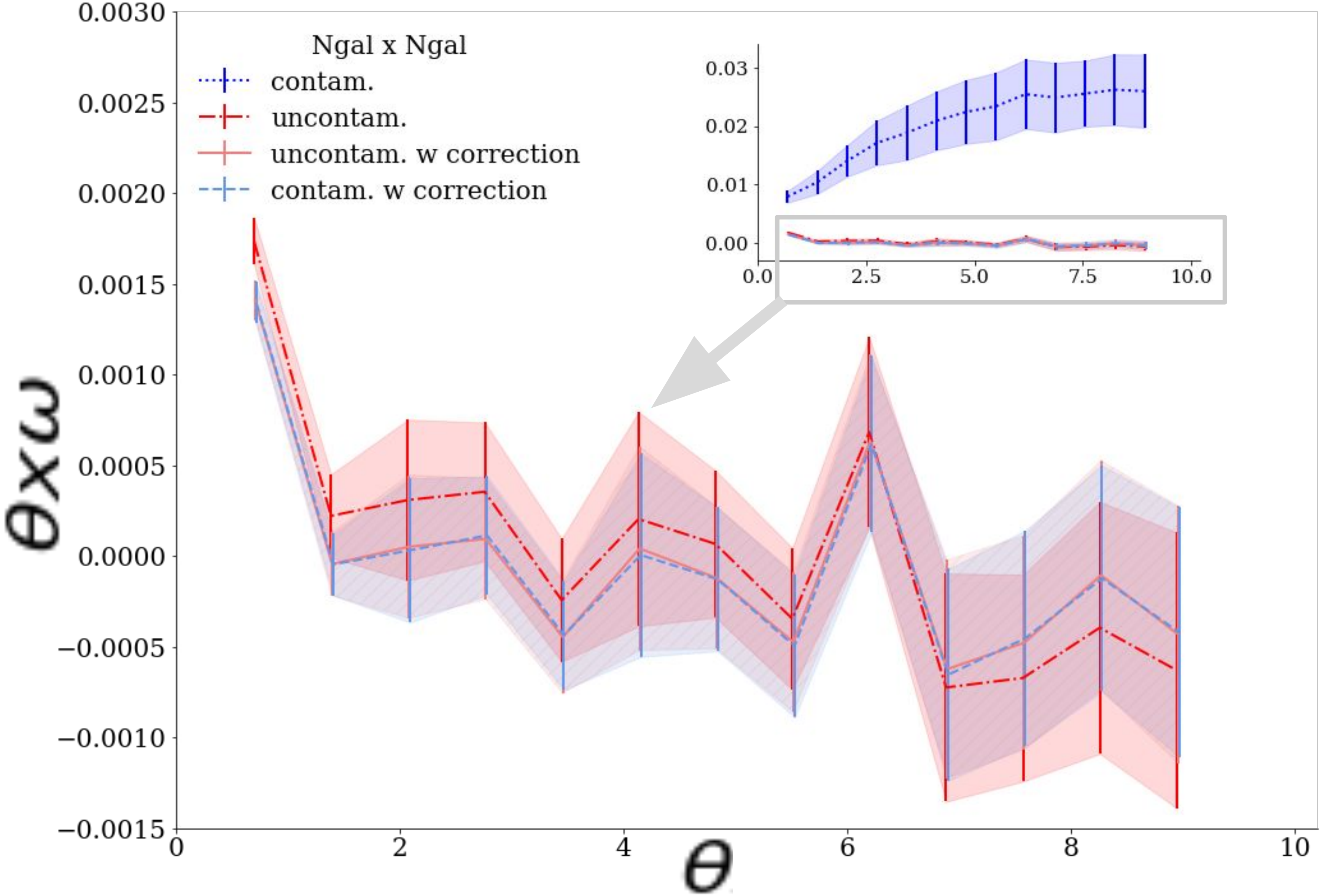
$$\delta + 1 \equiv \frac{N}{\overline{N}} = \frac{N}{sf \times R}$$



Clustering statistics

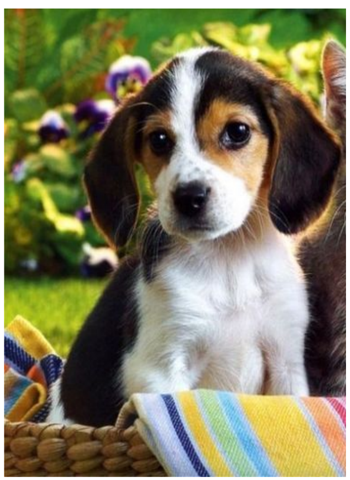


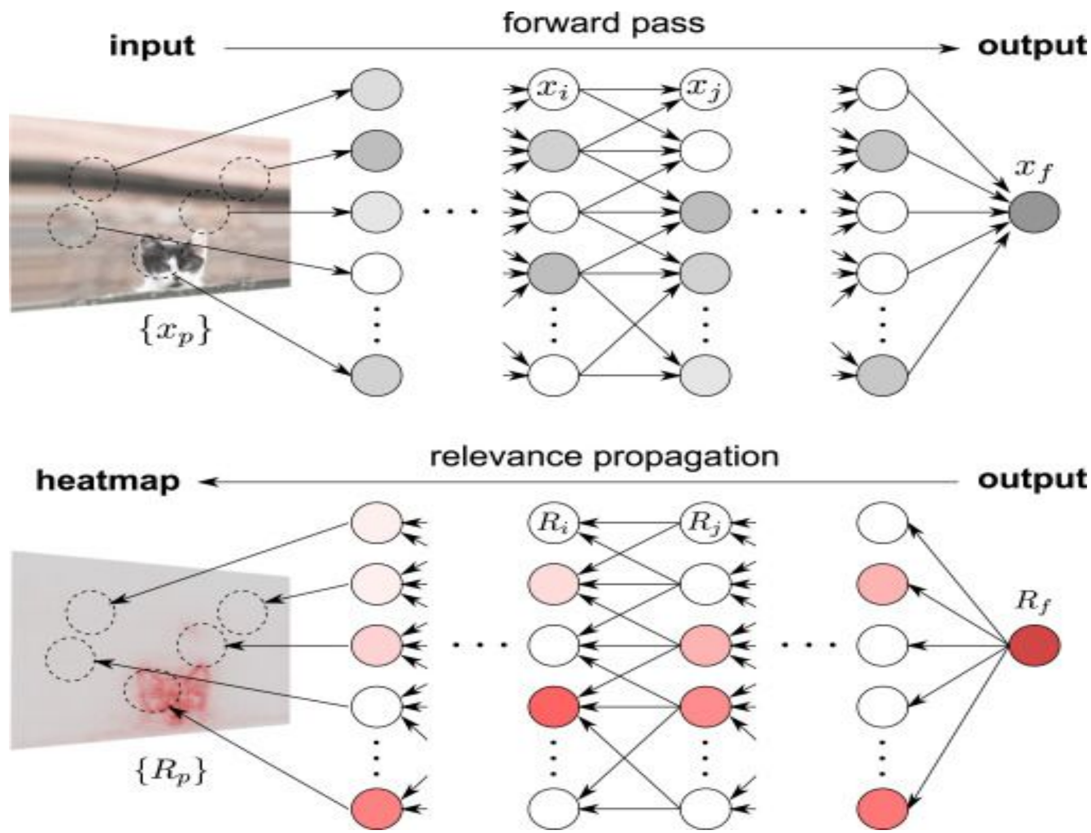
Validation via mocks [1 realization so far]



NEXT:

- Finish the 2x100 mock test
- Experiment with different healpix resolutions
- Feature importance with Layerwise Relevance Propagation
  And or Ablation





An example of using DECam data to build a new method for correcting systematic effects on target density and clustering measurements

Beneficial to galaxy surveys eg. DESI, eBOSS

Thank You!

Harvard Business Review

TECHNOLOGY

Using AI to Invent New Medical Tests

by John J. Dillon and Paul A. Friedman

MARCH 26, 2018

Science Home News Journals Topics Careers

Self-taught artificial intelligence beats doctors at predicting heart attacks

Technology BBC

Google's 'superhuman' DeepMind AI claims chess crown

6 December 2017

Share

Google says its AlphaGo Zero artificial intelligence program has triumphed at chess against world-leading specialist software within hours of teaching itself the game from scratch.

Why does deep and cheap learning work so well?*

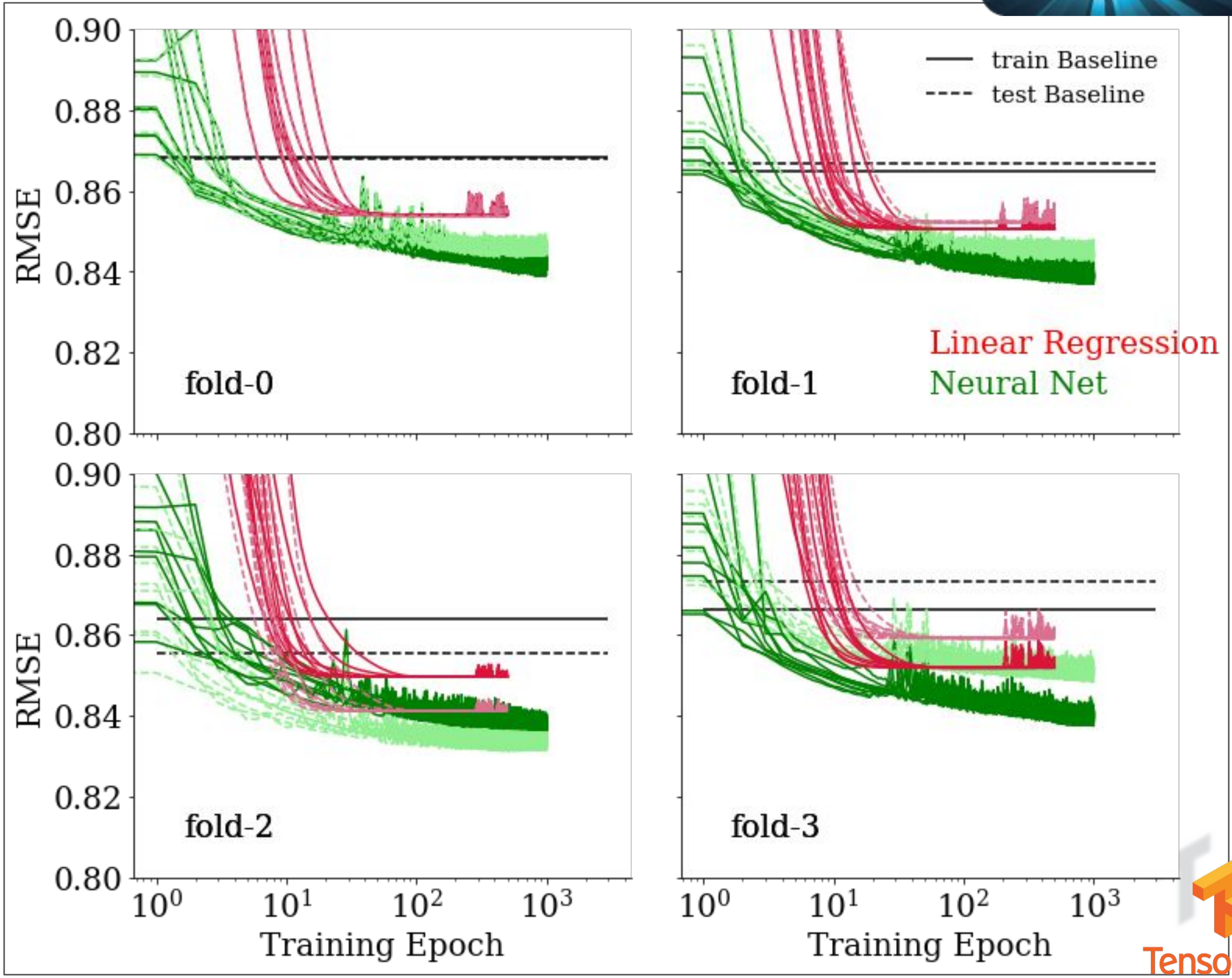
Henry W. Lin, Max Tegmark, and David Rolnick
Dept. of Physics, Harvard University, Cambridge, MA 02138
Dept. of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139 and
Dept. of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139

(Dated: July 21 2017)

When investigating the quality of a neural net, there are several important factors to consider:

- **Expressibility:** What class of functions can the neural network express?
- **Efficiency:** How many resources (neurons, parameters, *etc.*) does the neural network require to approximate a given function?
- **Learnability:** How rapidly can the neural network learn good parameters for approximating a function?

This paper is focused on expressibility and efficiency, and more specifically on the following well-known [4–6] problem: *How can neural networks approximate functions well in practice, when the set of possible functions is exponentially larger than the set of practically possible networks?* For example, suppose that we wish to classify


0.90
0.88
0.86
0.84
0.82
0.80
RMSE
fold-0
fold-1
fold-2
fold-3
train Baseline
test Baseline
Linear Regression
Neural Net
10^0
10^1
10^2
10^3
Training Epoch

Linear relationship between systematic and observed galaxy density

$$\delta_g^{\mathrm{o}} = \delta_g^t + \sum_i \epsilon_i \delta_i$$

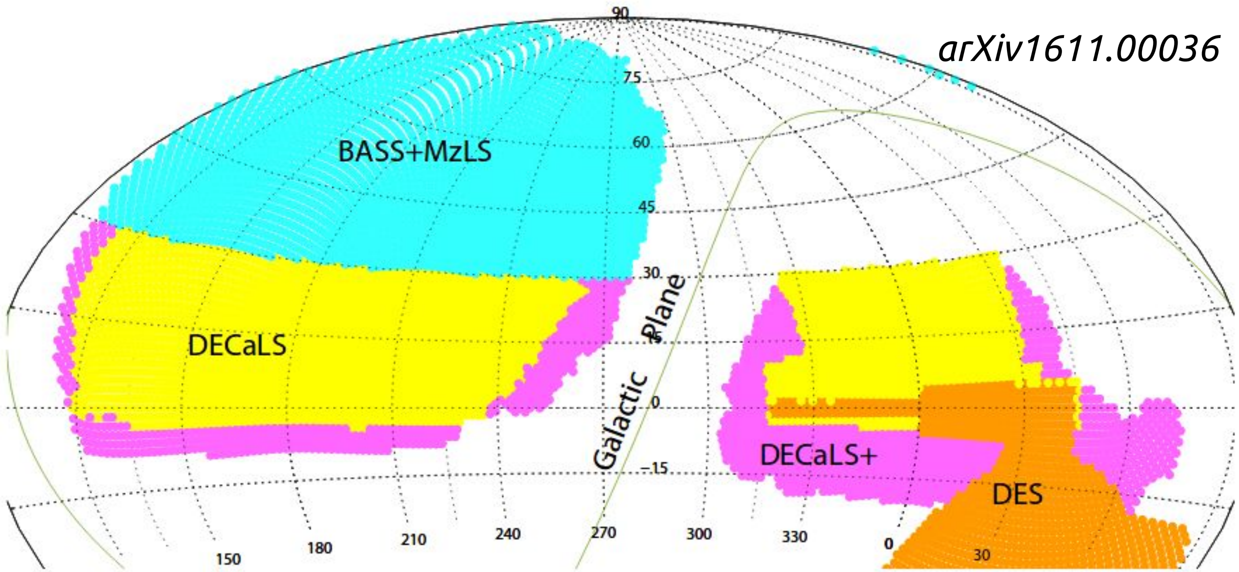
$$w_g^t(\theta) = w_g^{\mathrm{o}}(\theta) - \sum_i \epsilon_i^2 w_i(\theta) - \sum_{i,j>i} 2\epsilon_i \epsilon_j w_{i,j}(\theta)$$

and

$$w_{g,i}^o = \sum_j \epsilon_j w_{i,j}(\theta),$$

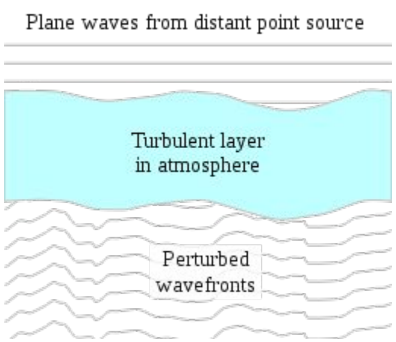
Data: the Legacy Surveys DR5

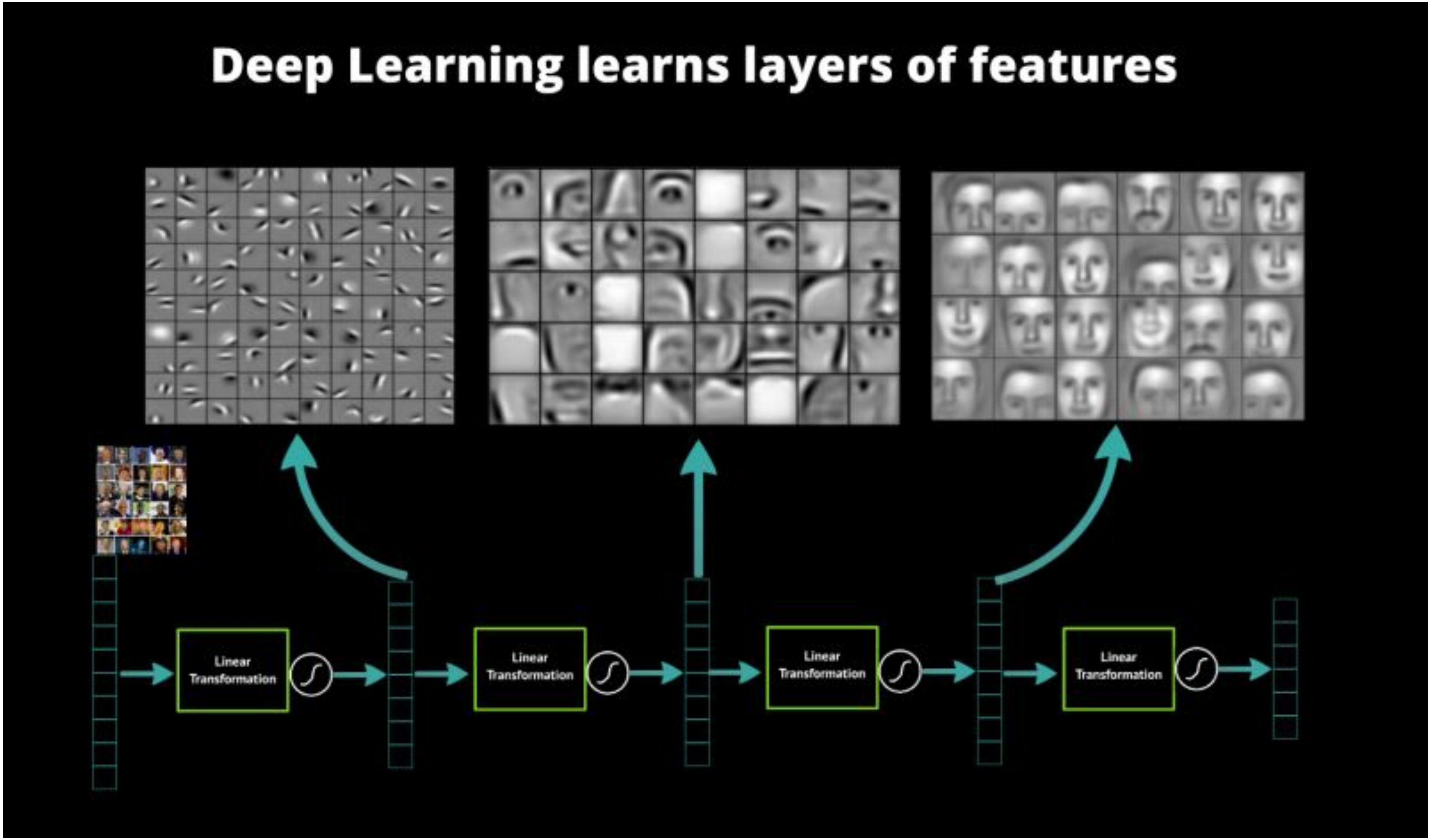
Telescope	Bands	Area deg^2	Location
Blanco DECam	g,r,z	9k	NGC+SGC (Dec $\leq +34$ deg)
Bok 90Prime	g,r	5k	NGC (Dec $\geq +34$ deg)
Mayall MOSAIC-3	z	5k	NGC (Dec $\geq +34$ deg)
WISE-W1	3.4 μm	all sky	all-sky
WISE-W2	4.6 μm	all sky	all-sky



arXiv1611.00036

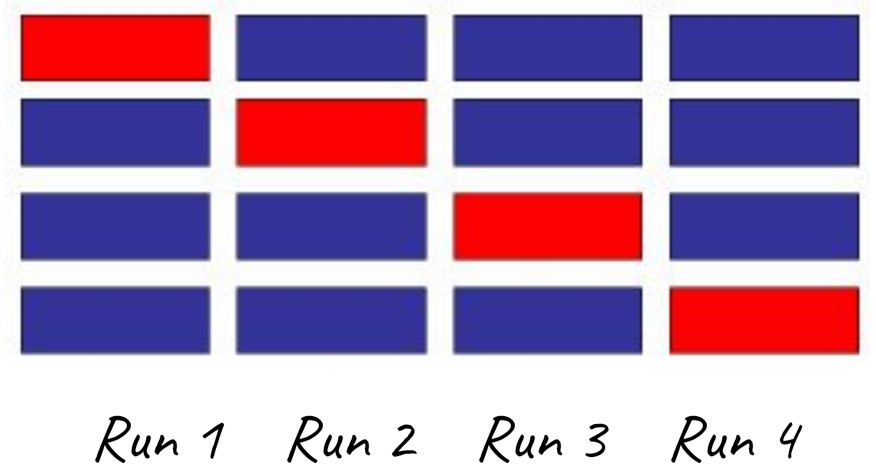
Imaging Systematics that affect galaxy densities





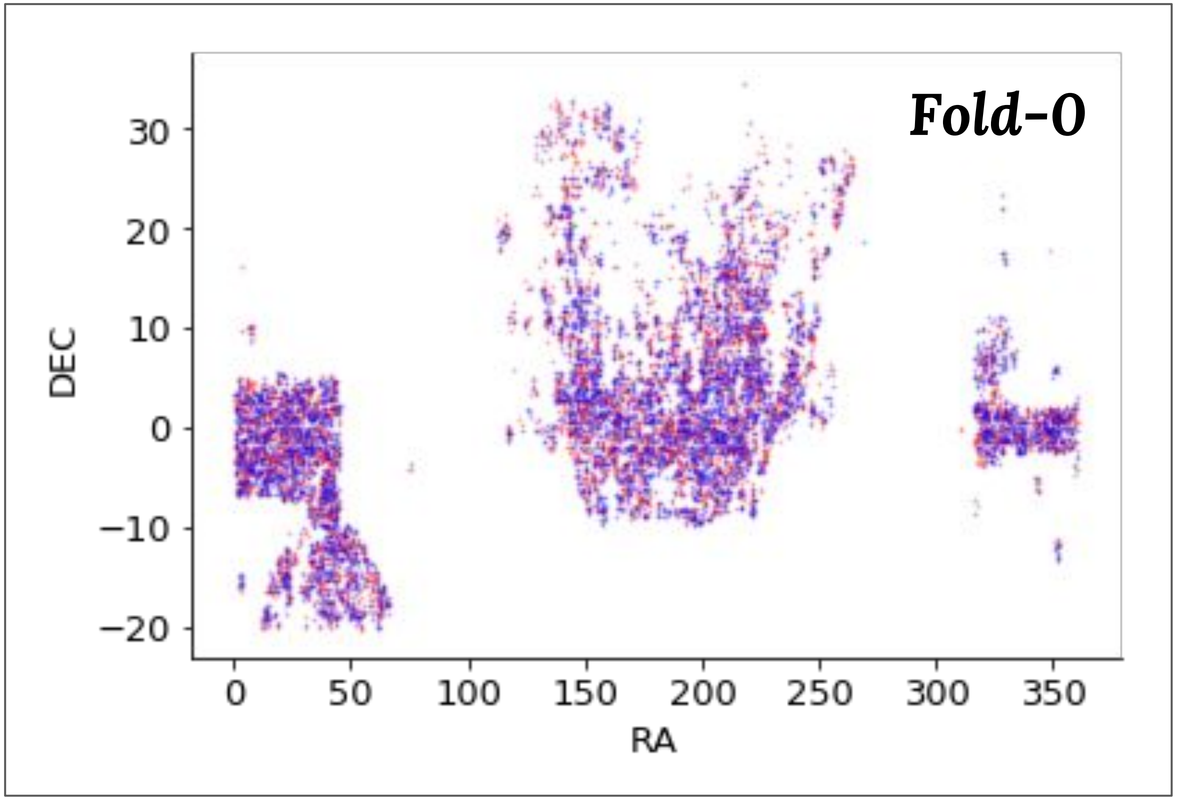
http://blog.datarobot.com/a-primer-on-deep-learning

4-Fold Validation



training

testing



Number of Linear Regions of Shallow vs. Deep Networks [Montufar et a., NIPS'14]

Conjecture

A deep network has significantly greater representational power than a shallow one.

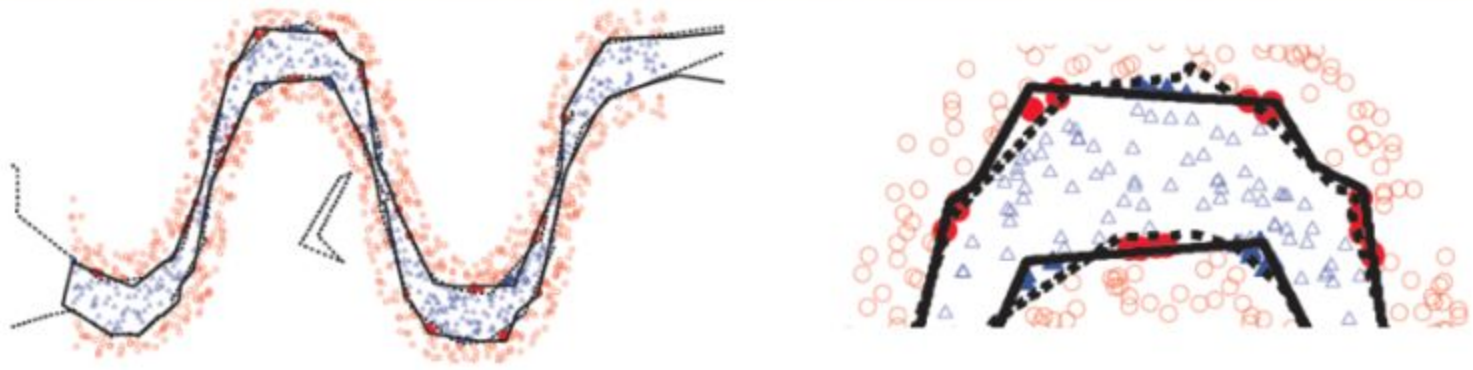


Figure 1: Binary classification using a shallow model with 20 hidden units (solid line) and a deep model with two layers of 10 units each (dashed line). The right panel shows a close-up of the left panel. Filled markers indicate errors made by the shallow model.