HITS
Heidelberg Institute for
Theoretical Studies

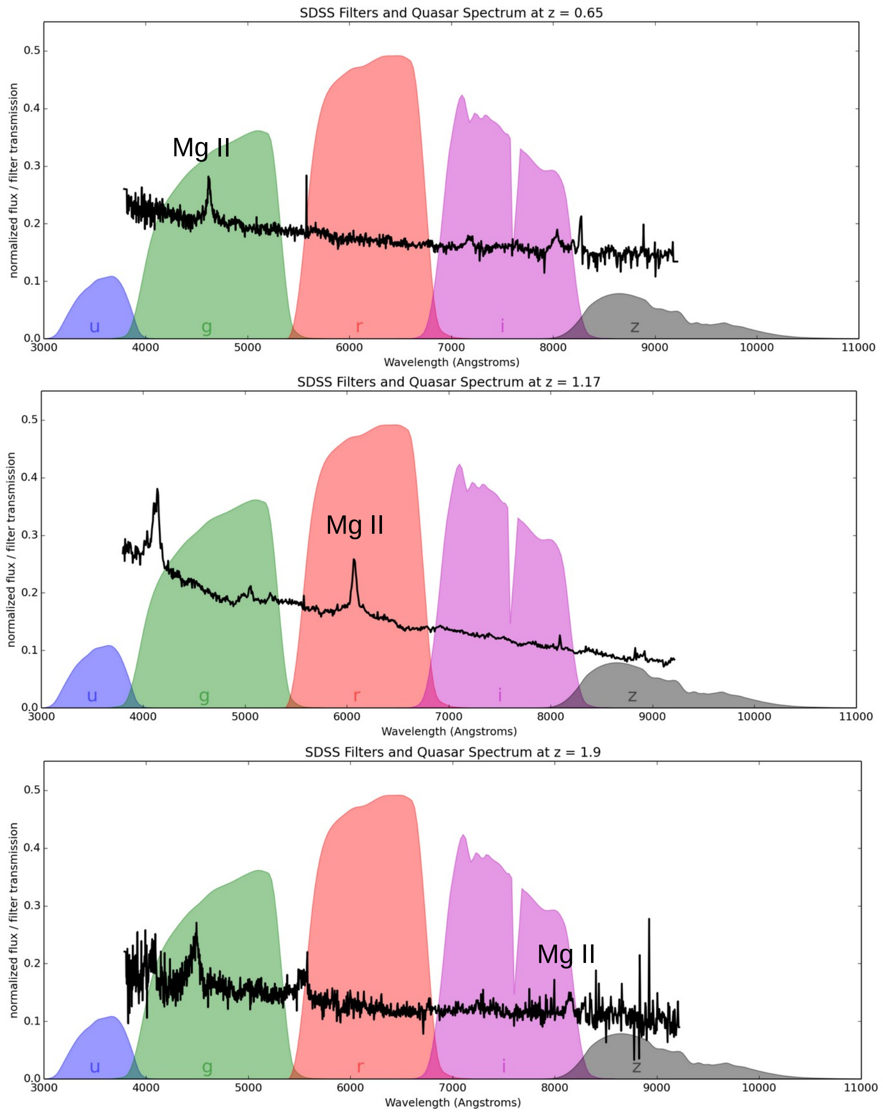
AiN

The two worlds of photometric redshift estimation via machine learning: fully automatic vs feature based

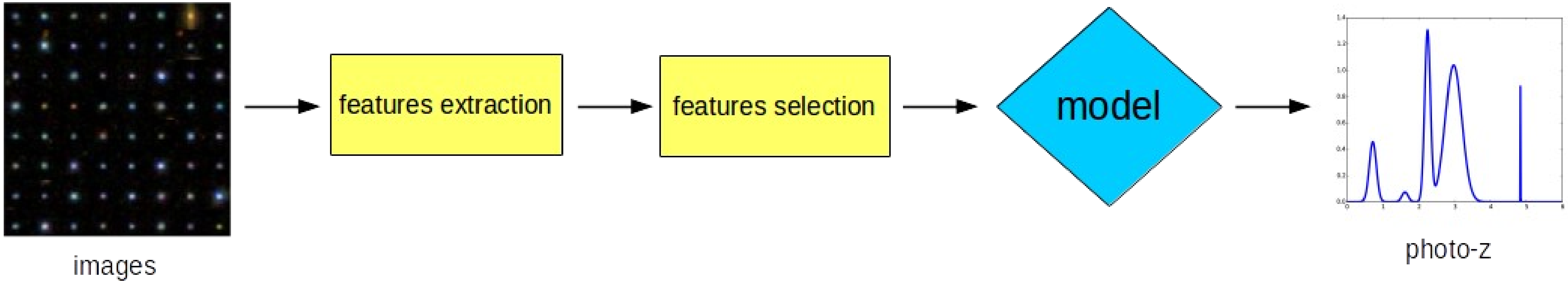
Antonio D’Isanto – Astroinformatics group

Redshift

- measures of distance are fundamental in Astronomy
- redshift (shift of spectral lines to the red) → measured through spectroscopy, but expensive and time consuming task
- photometric redshifts are estimated based on broadband imaging, allowing to obtain distances for many objects
- → template fitting and empirical methods (machine learning)



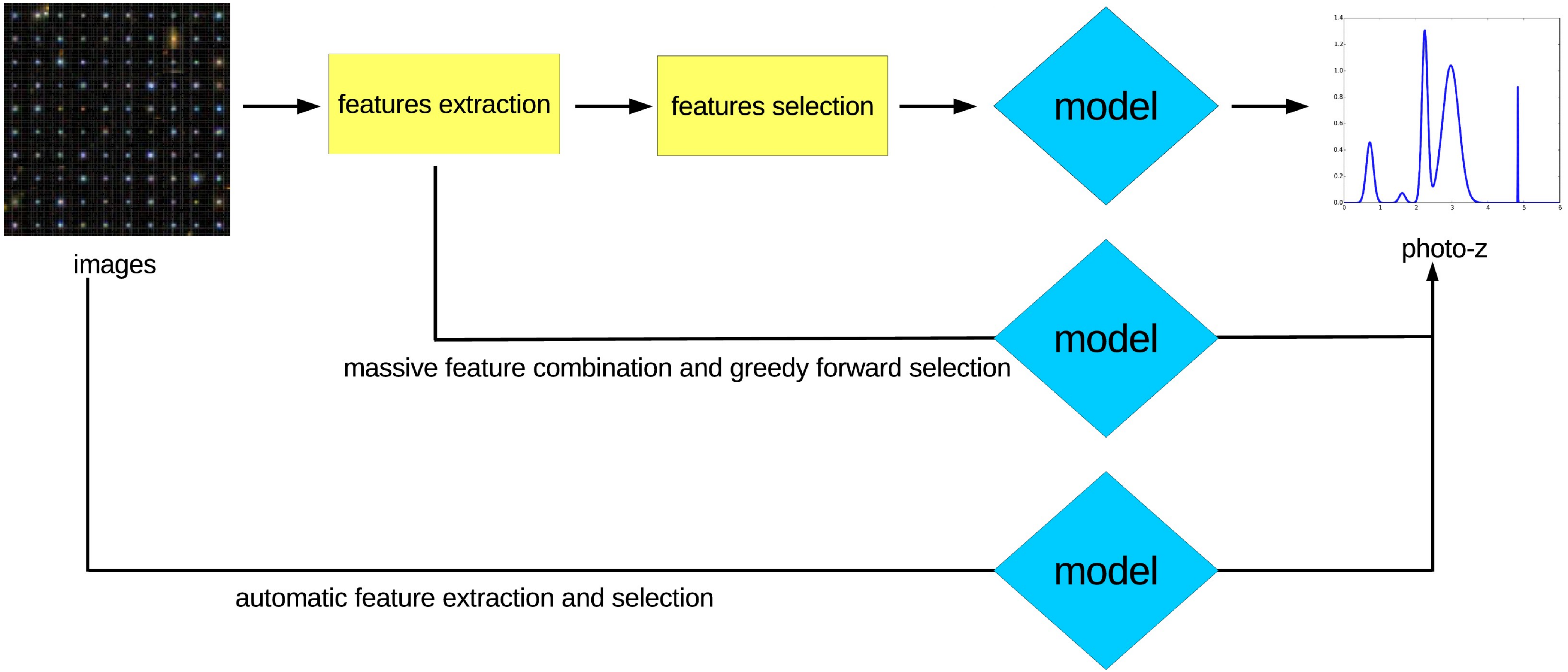
Photometric redshift pipeline



Focuses

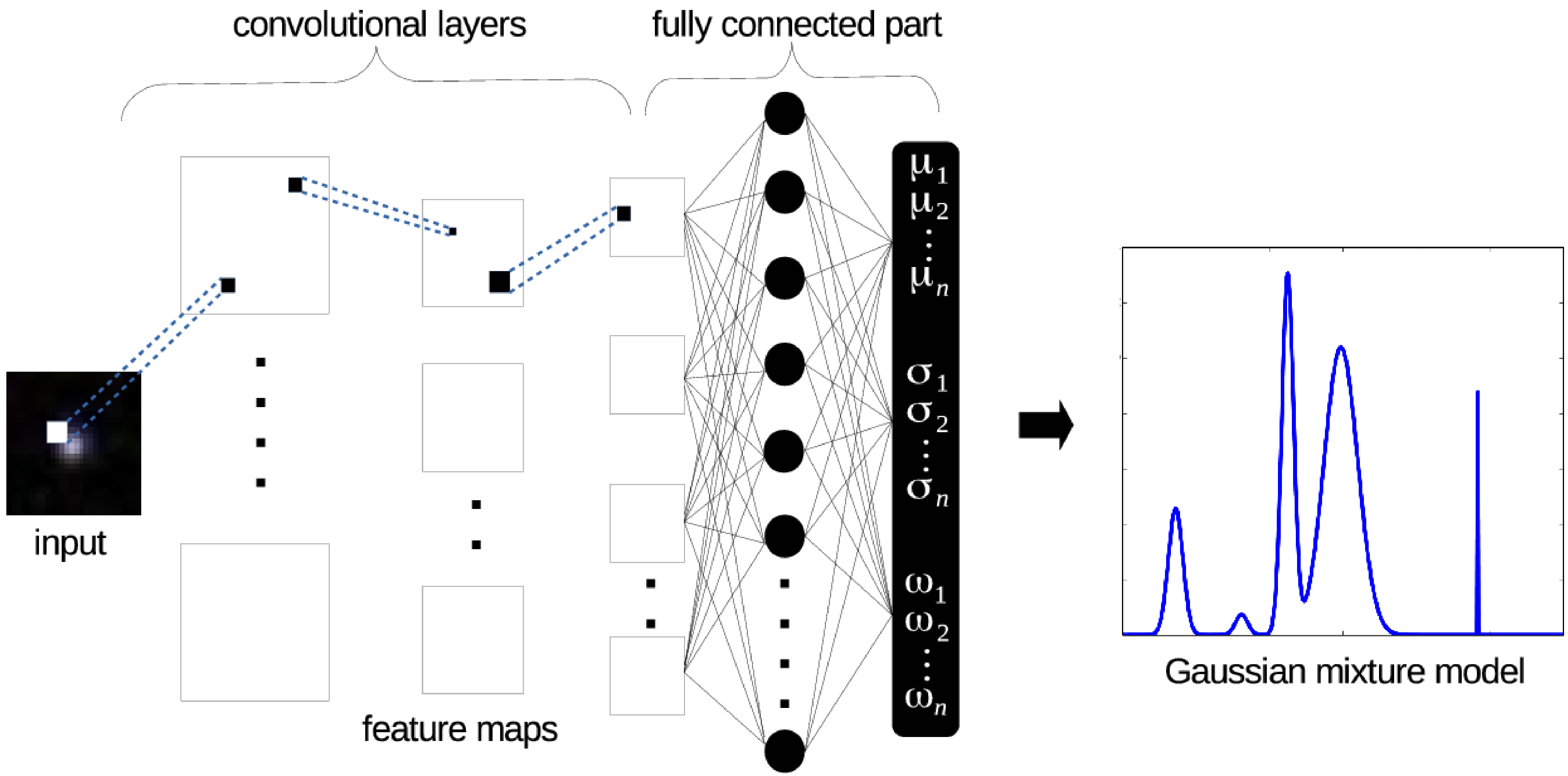
- Feature extraction and selection
- Point estimates vs PDFs
  → correct error evaluation
- Multimodalities
- Use of available information
- Performance

Two alternative models



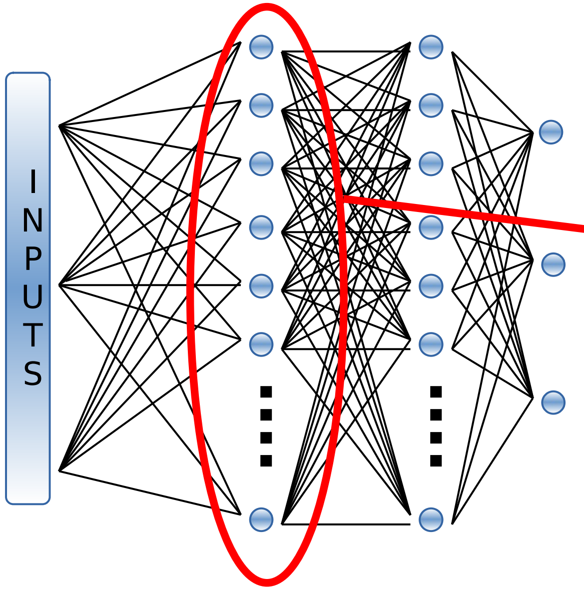
Model 1

Deep Convolutional Mixture Density Network (DCMDN) → CNN + MDN



$$f(g(h(j(X))))=GMM(\mu_n,\sigma_n,\omega_n)$$

Feed forward neural network

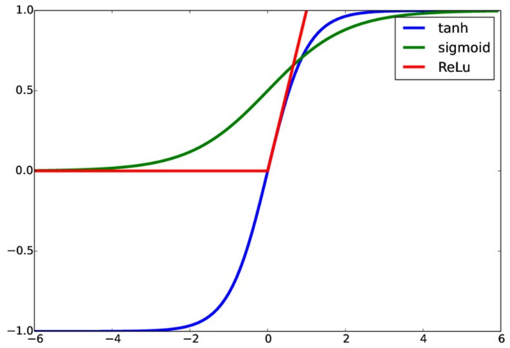


$$z_j = f\left(\sum w_{ji}^{(1)} x_i + b_j^{(1)}\right)$$

$$y_k(x,w) = g\left(\sum w_{kj}^{(2)} f\left(\sum w_{ji}^{(1)} x_i + b_j^{(1)}\right) + b_k^{(2)}\right)$$

f,g = activation function:

$$\tanh = \frac{e^z - e^{-z}}{e^z + e^{-z}} \qquad sigmoid = \frac{1}{1+e^{-z}} \qquad ReLu = max(0,x)$$

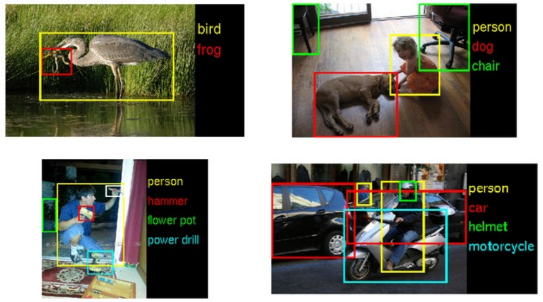


Deep learning

A class of machine learning algorithms, based on artificial neural networks:

- cascade of non linear processing units;
- automatic learning of features;
- different levels of abstraction.

image recognition

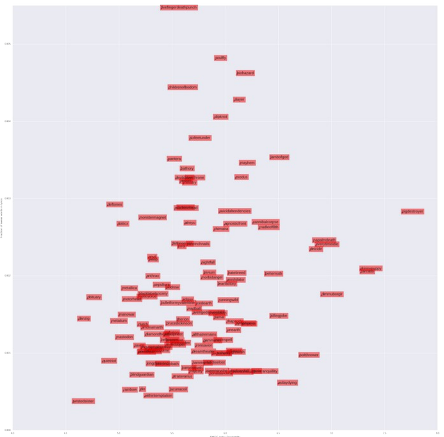


Credits: https://blogs.nvidia.com/blog/2014/09/07/imagenet/

speech recognition

Credits: https://www.ntid.rit.edu/sites/default/files/pd/Symposium_2017/LBerke_MHuenerfauth_1300_1140.pdf

natural language processing



Credits: http://www.degeneratestate.org/posts/2016/Apr/20/heavy-metal-and-natural-language-processing-part-1/

Convolutional neural network

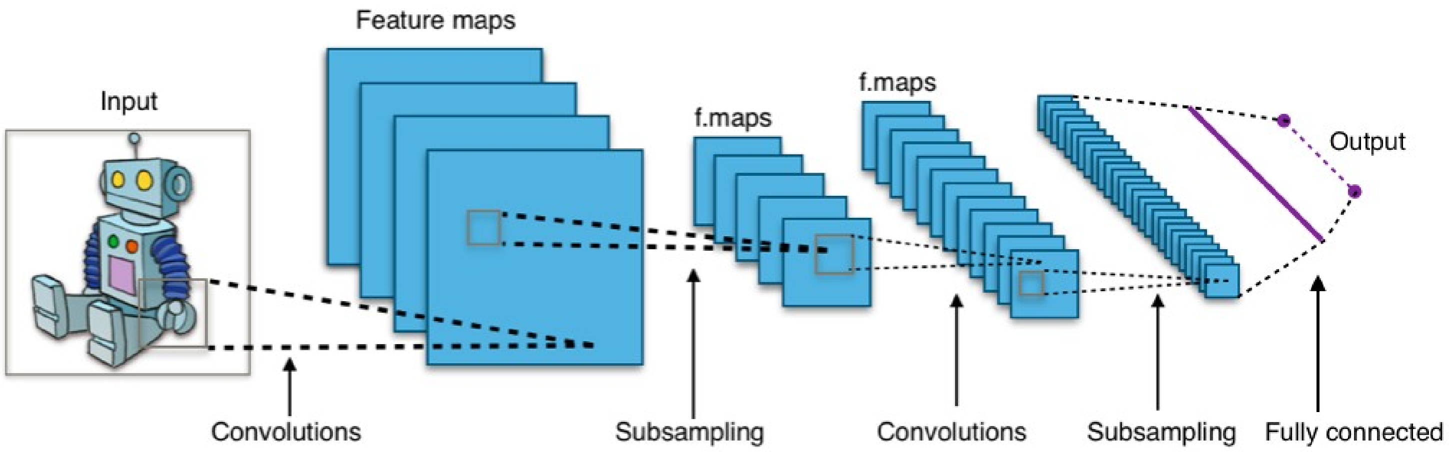
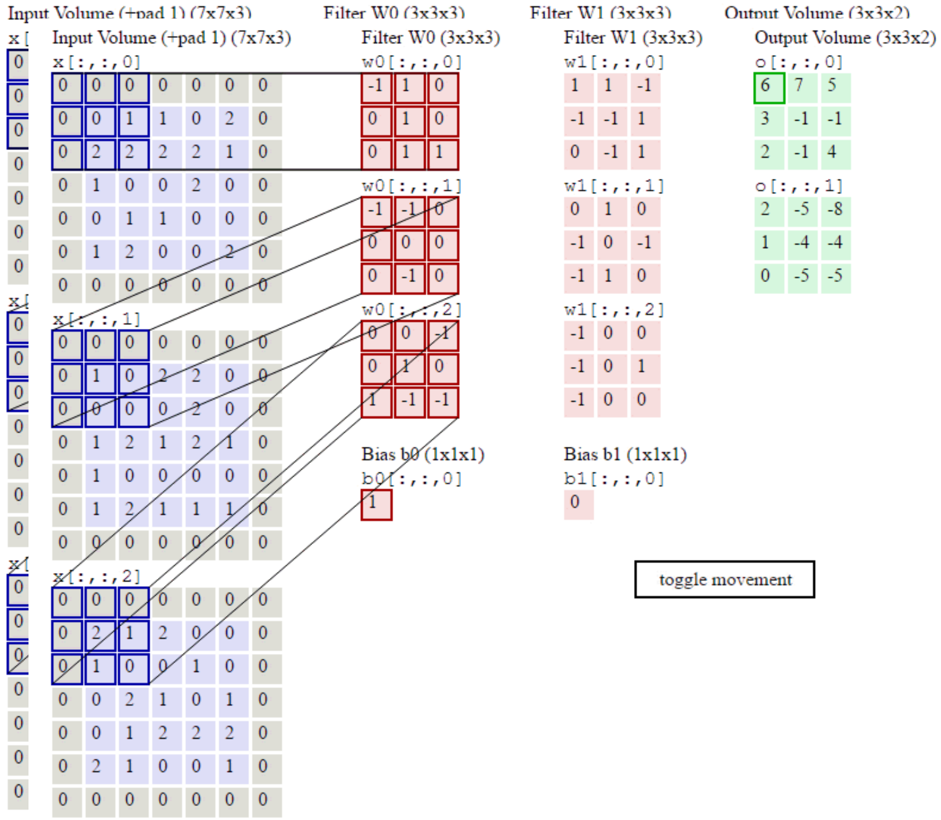


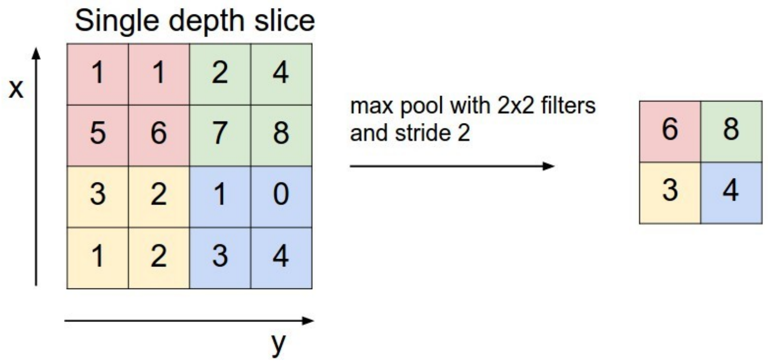
Image from Wikipedia

Neural network characterized by two main parts:

- Convolutional part --> only local connection
- Fully connected part --> full connection between neurons, typically a deep MLP

Convolutional part

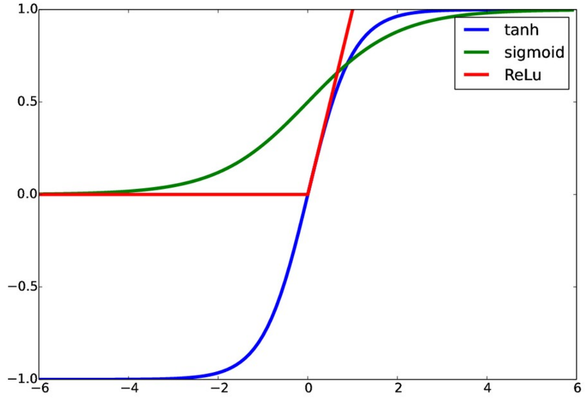




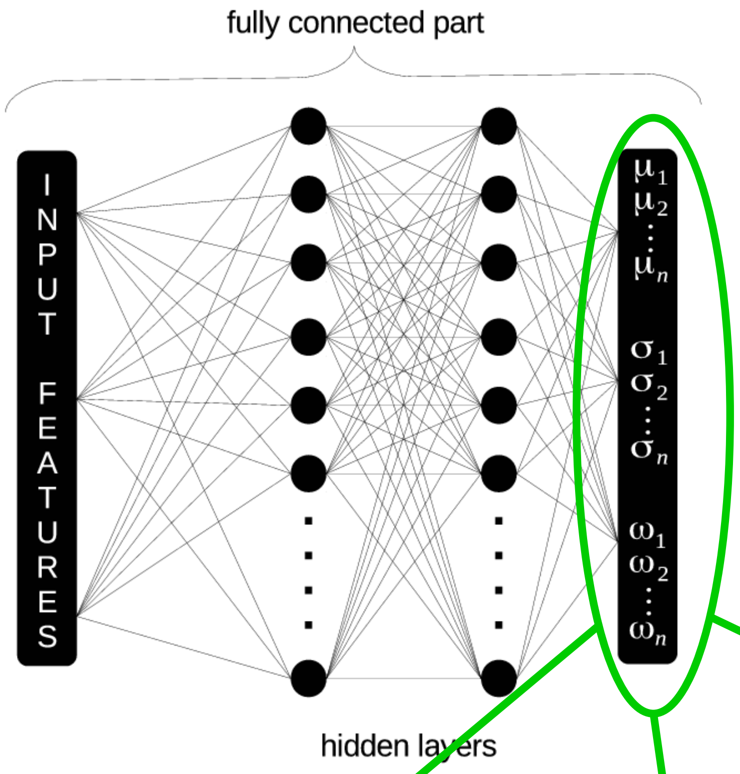
$$h_{ij}^{k} = f\left(\left(W^{k} * x\right)_{ij} + b_k\right)$$

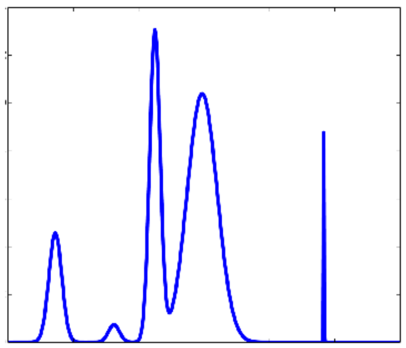
f = activation function:

- tanh
- sigmoid
- Rectified Linear Unit: $f(x) = max(0, x)$



Mixture Density Network





Gaussian mixture model

$$p(x)=\sum_{j=1}^{n}\omega_j N\left(x|\mu_j,\sigma_j\right)$$

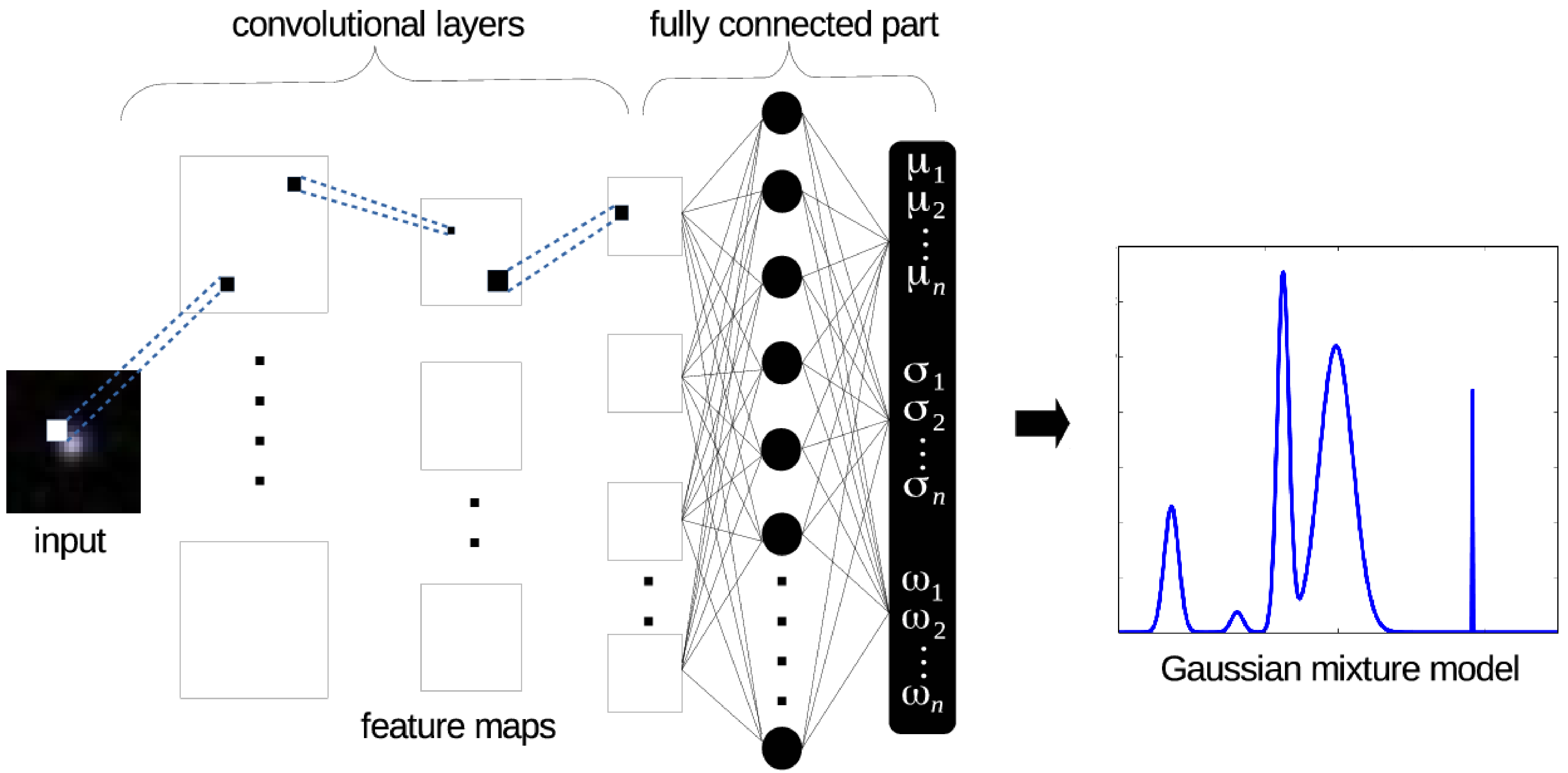
$$\mu_j = z_{output(\mu_j)}$$

$$\sigma_j = \exp\left(z_{output(\mu_j)}\right)$$

$$\omega_j = \frac{\exp\left(z_{output(\mu_j)}\right)}{\sum_{j=1}^{n}\exp\left(z_{output(\mu_j)}\right)}$$

DCMDN

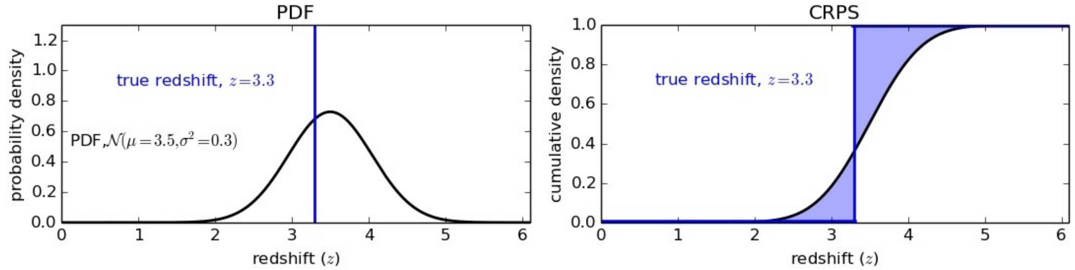
Deep Convolutional Mixture Density Network (DCMDN) → CNN + MDN



$$f(g(h(j(X)))) = GMM(\mu_n, \sigma_n, \omega_n)$$

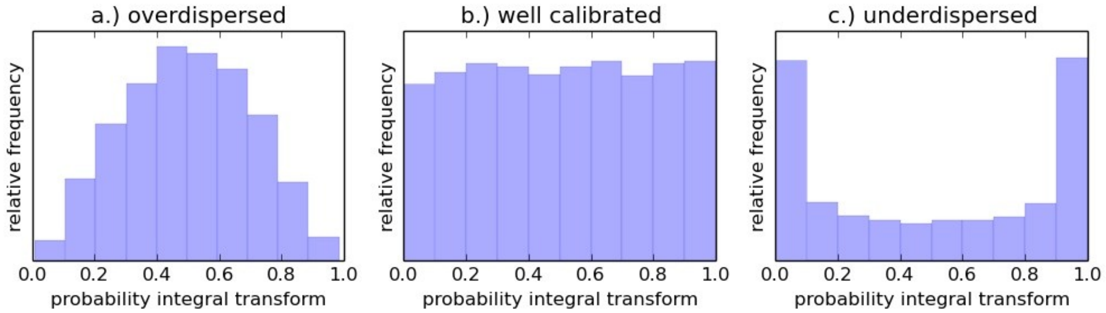
The importance of proper scores: CRPS and PIT

CRPS: continuous rank probability score



$$CRPS = \int \left[CDF(x) - CDF_a(x) \right]^2 dx$$

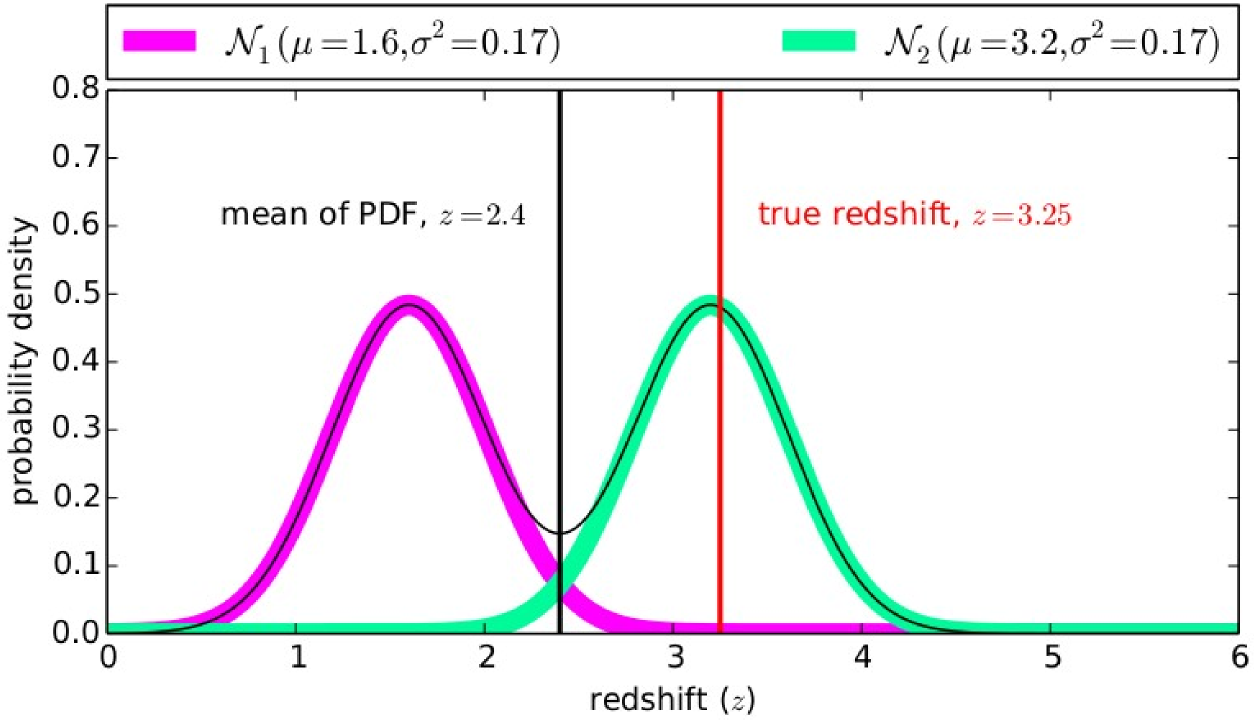
PIT: probability integral transform

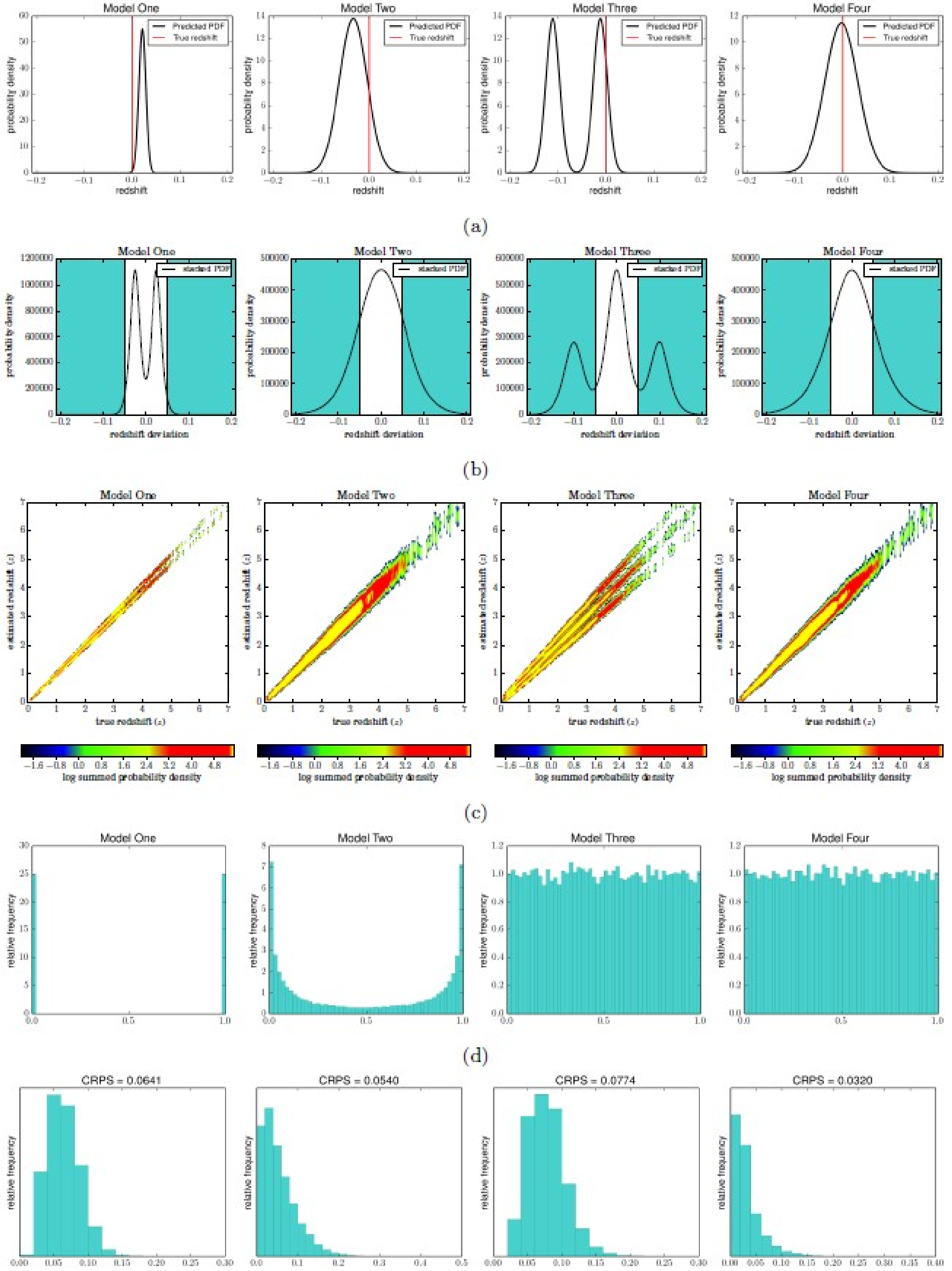


$$p_t = F_t(x_t)$$

See: Tilmann Gneiting, Fadoua Balabdaoui, Adrian Raftery. Probabilistic forecasts, calibration and sharpness.. 2007.

Why proper scores?





Python libraries for Deep Learning

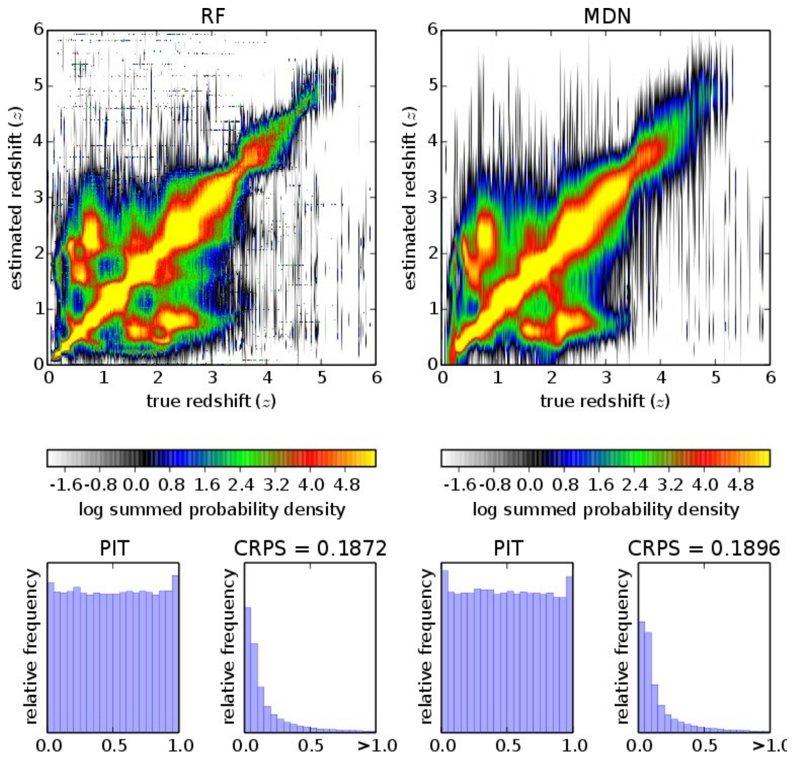
TensorFlow

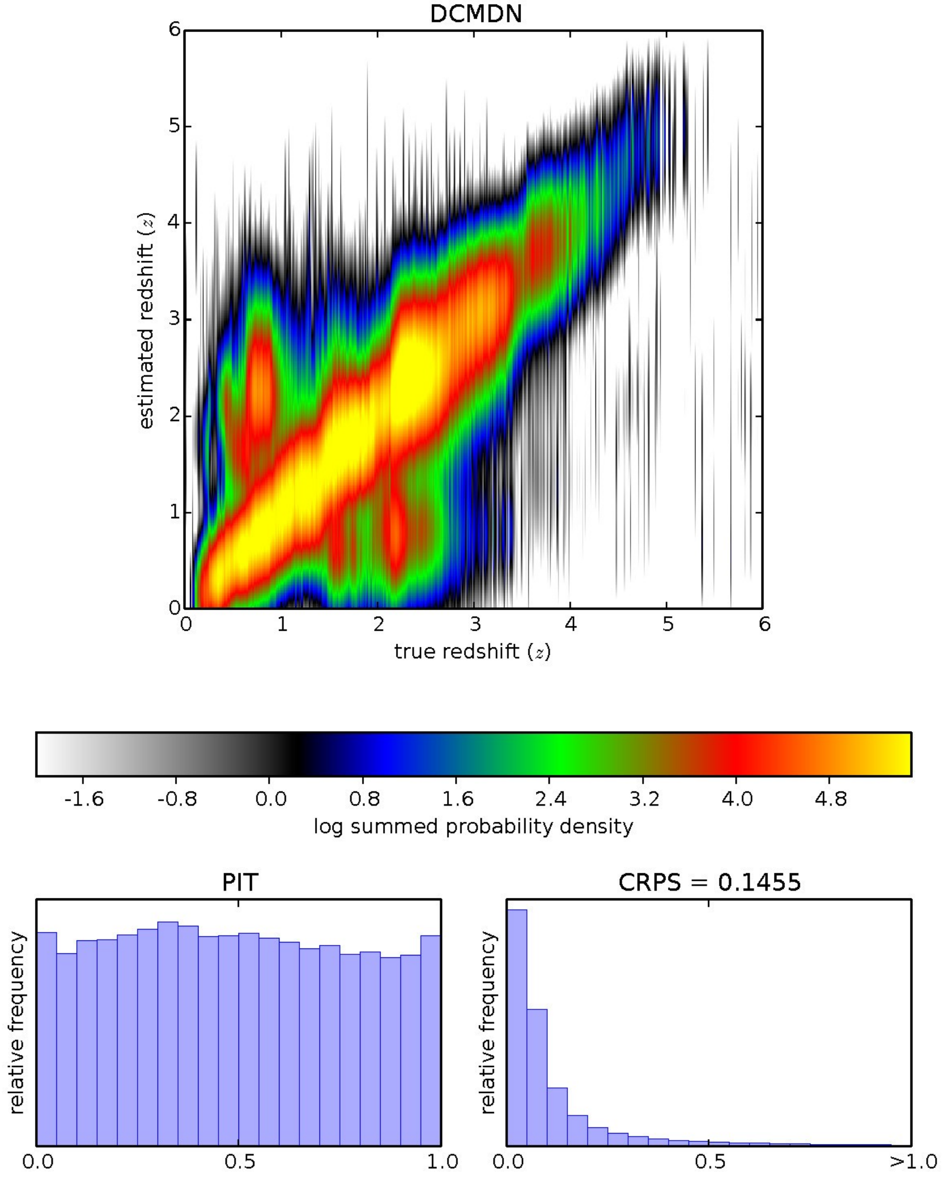
- Developed by Google
- Based on symbolic structure, like Theano
- High level API, including Keras support
- Several cool features (Tensorboard, autokeras, etc.)

The importance of GPUs

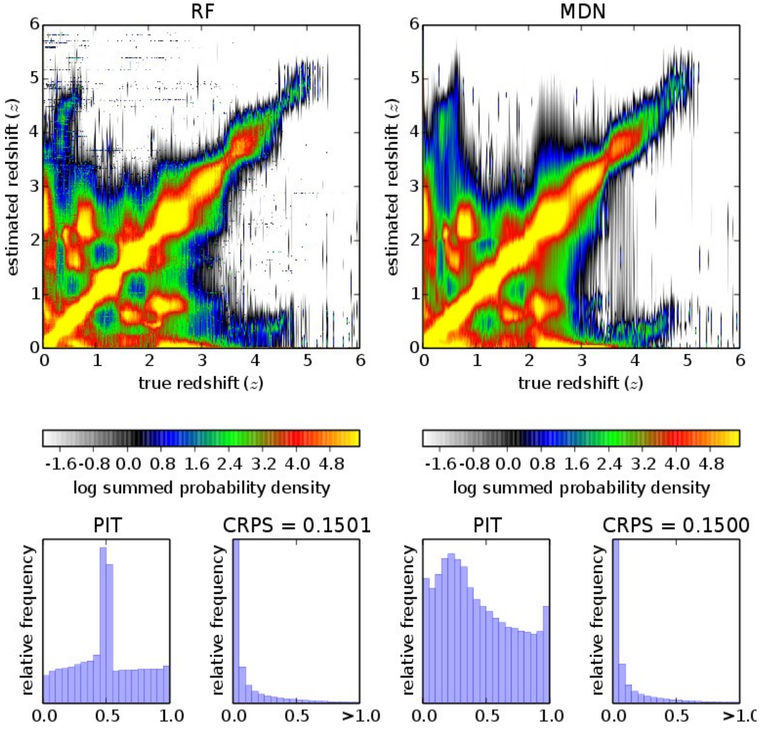
Machine	Running time/epoch
Single CPU	4 h 8 m
Multicore CPU (8x)	82 m
GPU Nvidia Titan X	90 s
GPU Nvidia P100	83 s
GPU Nvidia P40	81 s

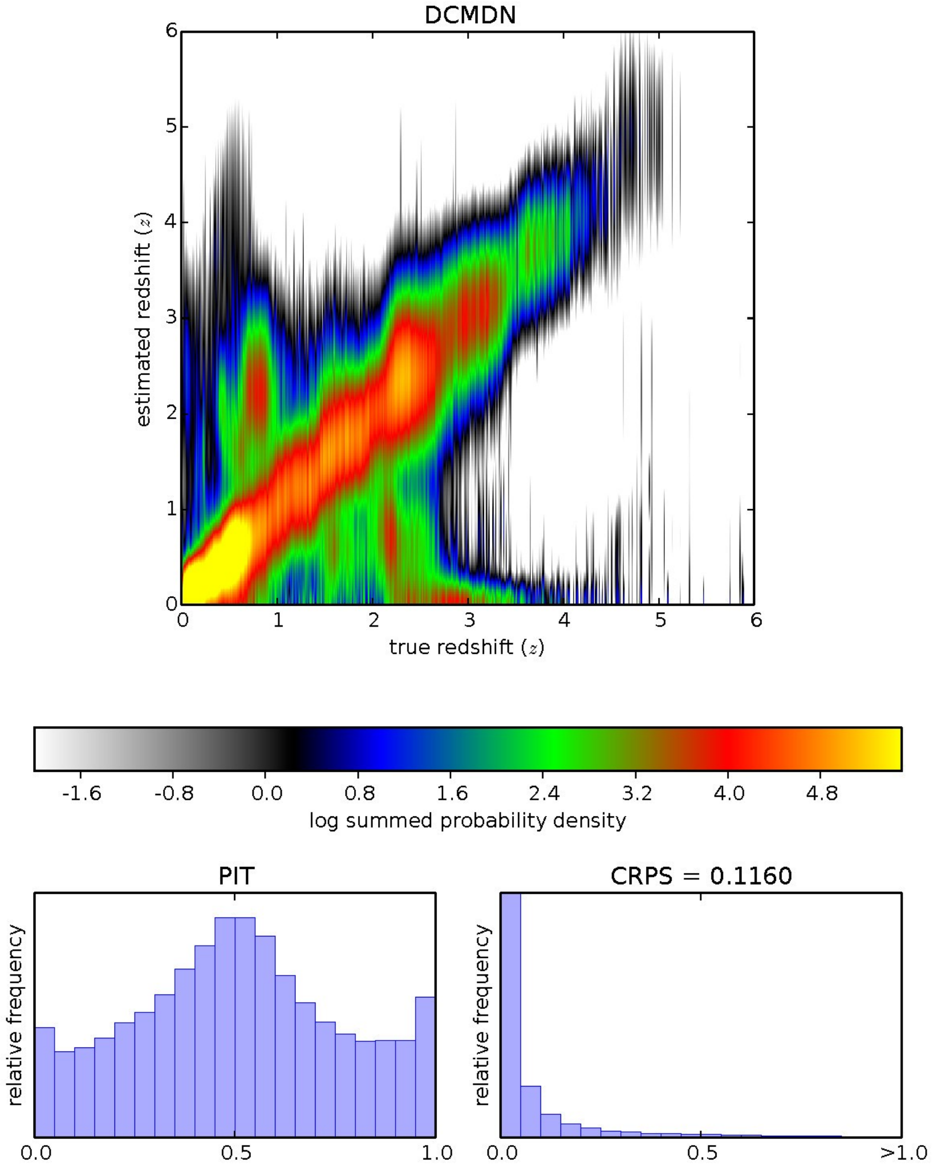
Results - Quasar



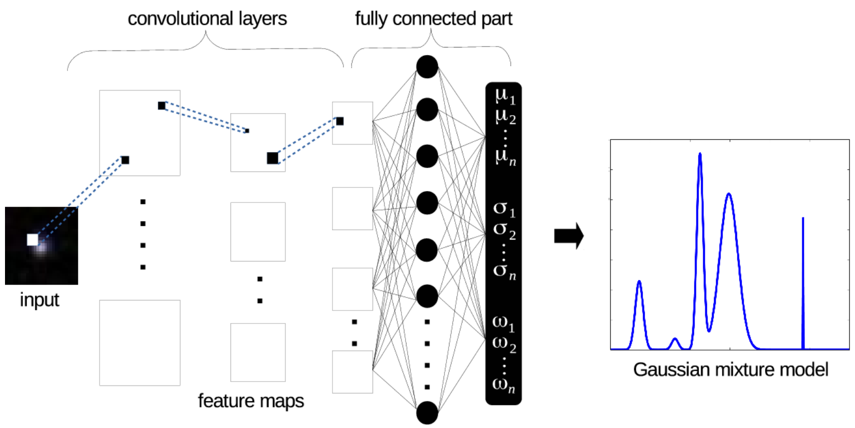


Results - Mixed





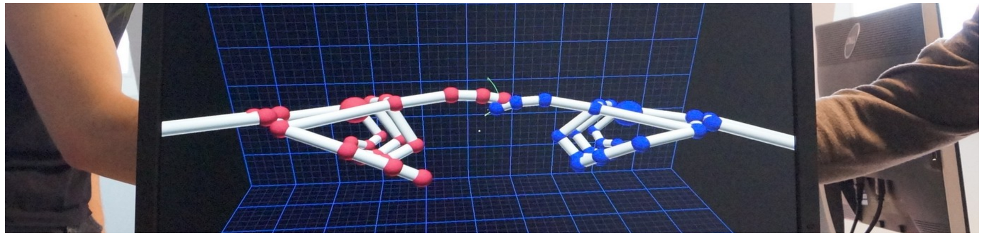
Model 2



black box?

physical meaning?

Classic$_{10}$
r_{psf}
r_{model}
$u_{psf} - g_{psf}$
$g_{psf} - r_{psf}$
$r_{psf} - i_{psf}$
$i_{psf} - z_{psf}$
$u_{model} - g_{model}$
$g_{model} - r_{model}$
$r_{model} - i_{model}$
$i_{model} - z_{model}$



- 90 parameters from SDSS database: magnitudes, errors, radii, ellipticities
- Massive combinations in differences, pairs, ratios, plus dereddening
- Ending with r = 4,520 features

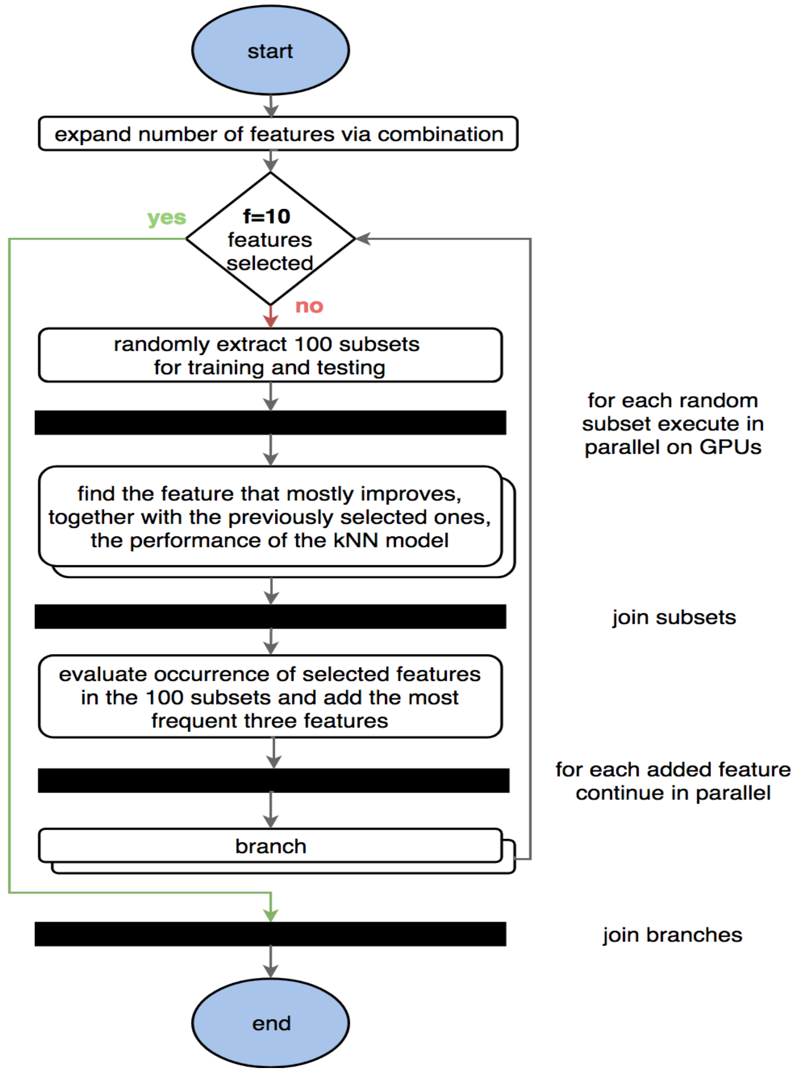
All possible combinations of f = 10 features:

$$n = \frac{r!}{f! * (r-f)!} = 9.7 \times 10^{29}$$

$$S \subset F : |S| = 10 \rightarrow \sim 10^{30}$$

Forward selection

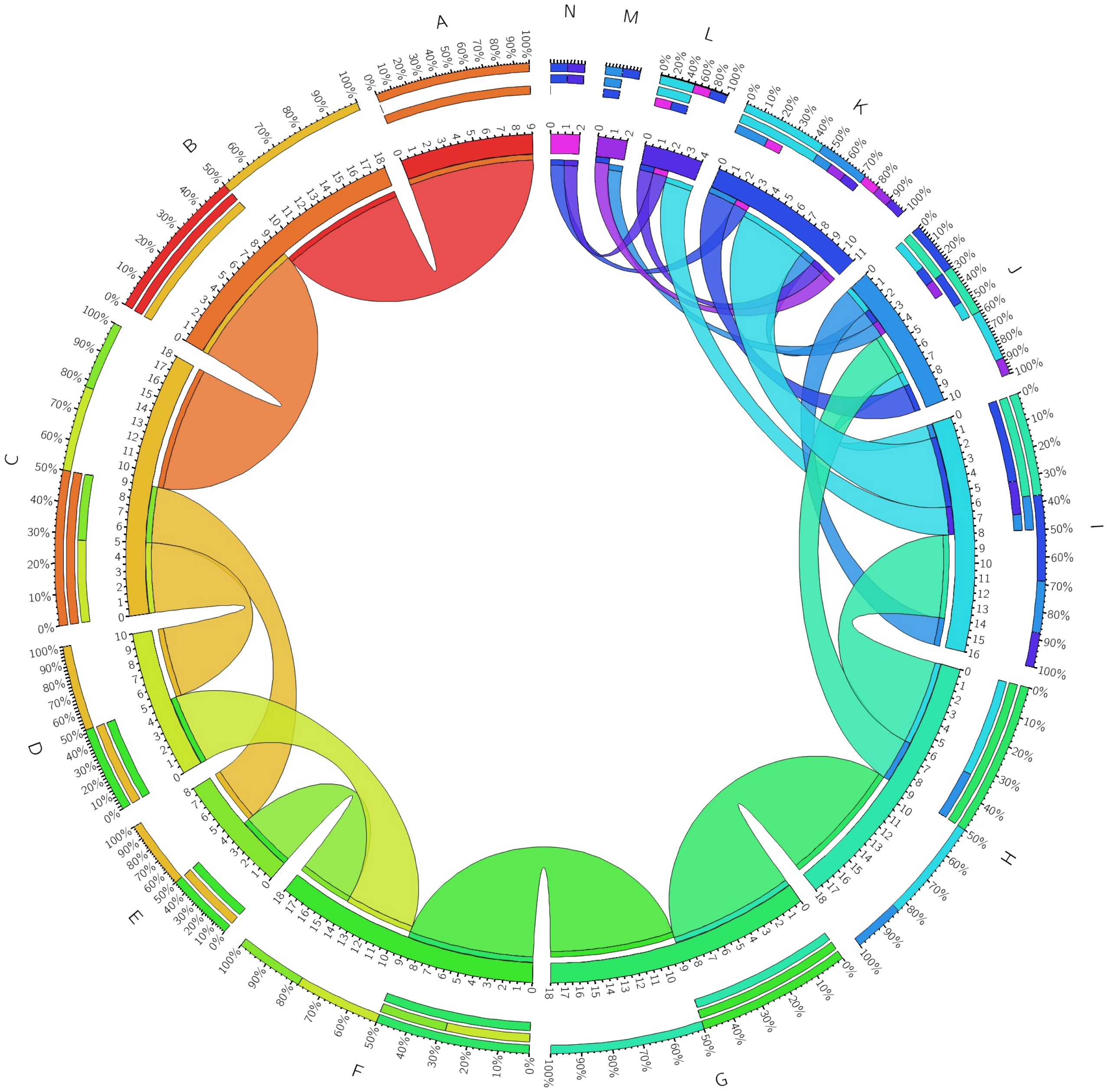
- Building a tree of most efficient features
- Every branch of the tree is a set of features
- Intense use of GPU computing : 100 samples X 4520 features X cross validation
- Evaluate the best set through random forest (RF) experiments for every branch



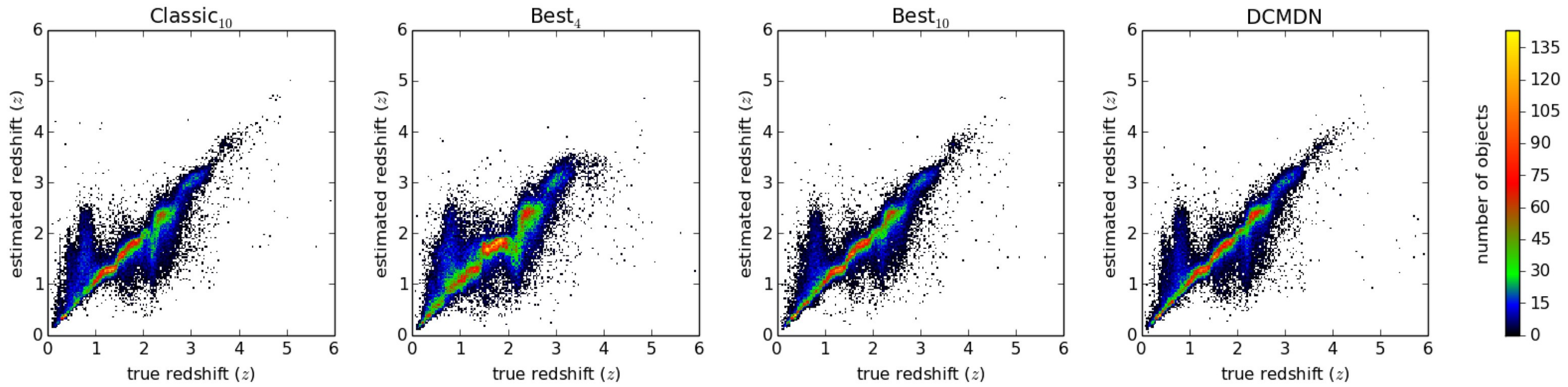
$$for\ i \in \{1,\ldots,10\} : for\ s \in \{S \subset F : |S| = 10, S^{1,\ldots,i-1} = S_{best}^{1,\ldots,i-1}\} : f(s) = s_{best}^{i}$$

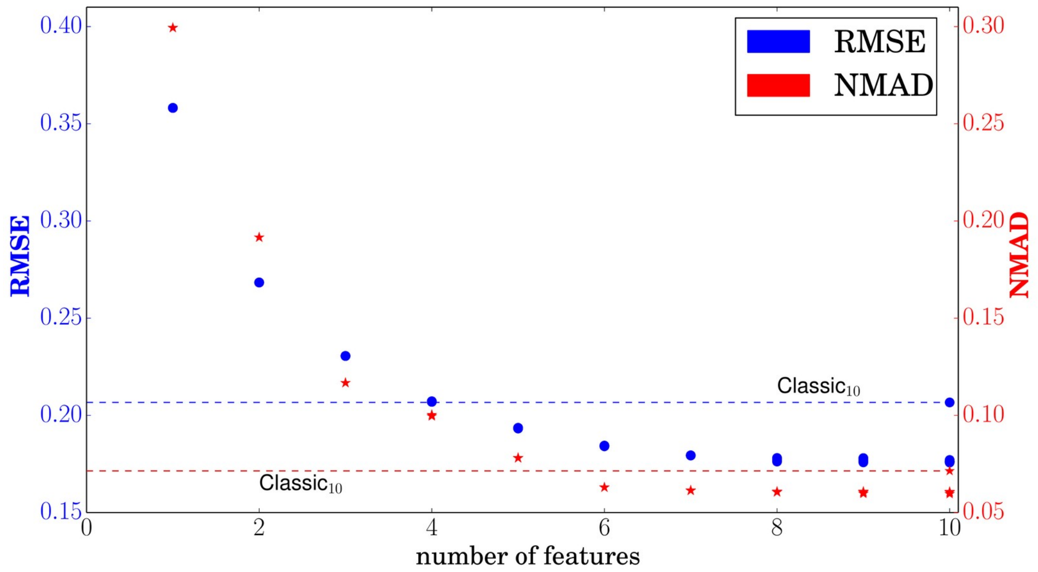
Chord diagram: how to visualize a complex feature structure

color	name	feature
	A	i_{petro}/i_{psf}
	B	g_{psf}-u_{model}
	C	i_{exp}/r_{psf}
	D	$\sqrt{\sigma^2_{r_{model}} + \sigma^2_{r_{dev}}}$
	E	$\sqrt{\sigma^2_{g_{model}} + \sigma^2_{r_{dev}}}$
	F	r_{psf}/g_{exp}
	G	i_{psf}/z_{model}
	H	$i_{psf} - i_{dev}$
	I	r_{petro}/r_{psf}
	J	$z_{psf} - z_{model}$
	K	i_{psf}-r_{model}
	L	$r_{psf} - r_{petro}$
	M	z_{exp}/z_{psf}
	N	i_{psf}-i_{dev}



Results





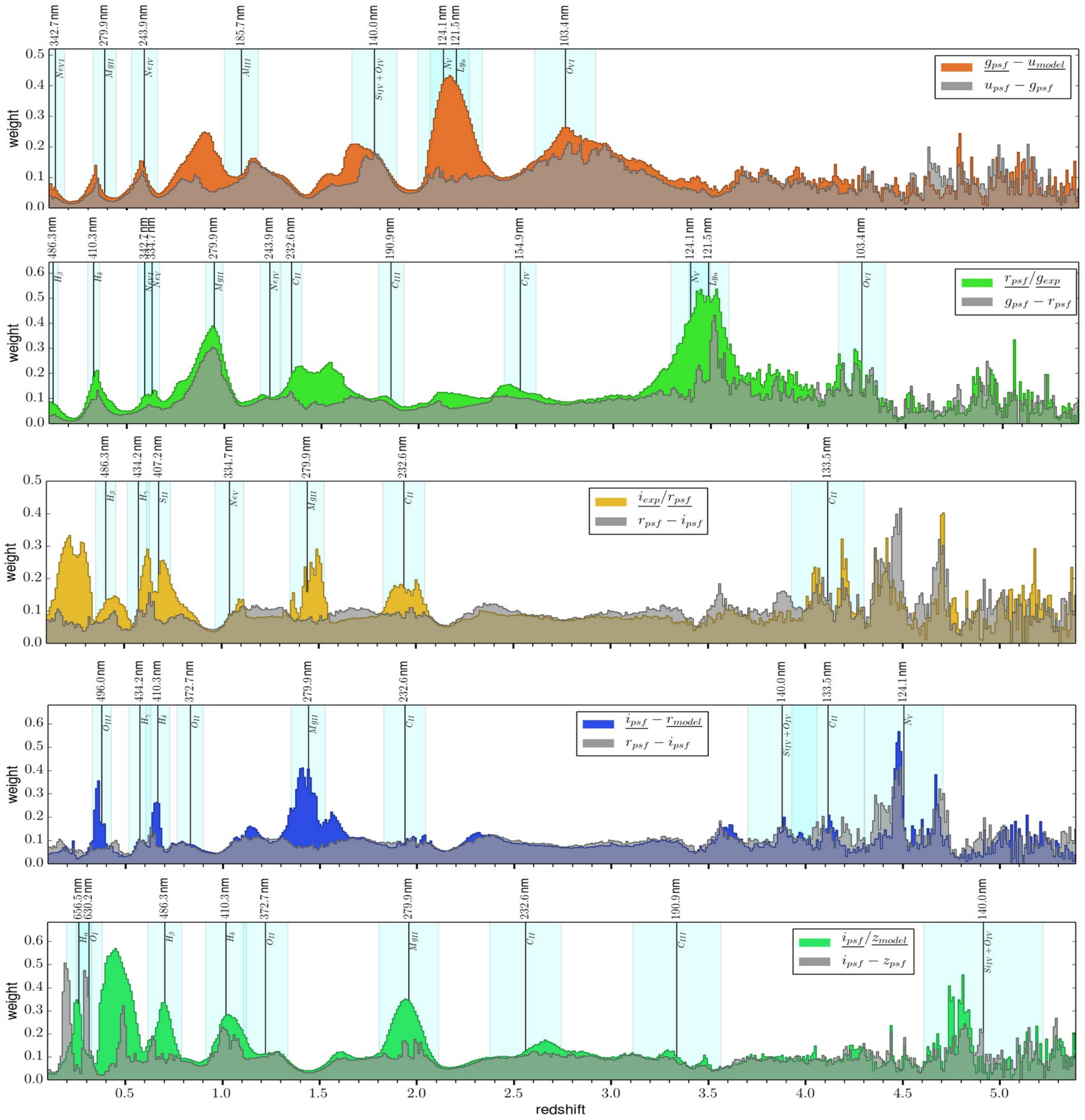
Set	Features	RMSE	CRPS
Classic	10	0.207	0.167
Best 4	4	0.206	0.203
Best 10	10	0.174	0.140
DCMDN	65,536	0.184	0.146

Finding a meaning I

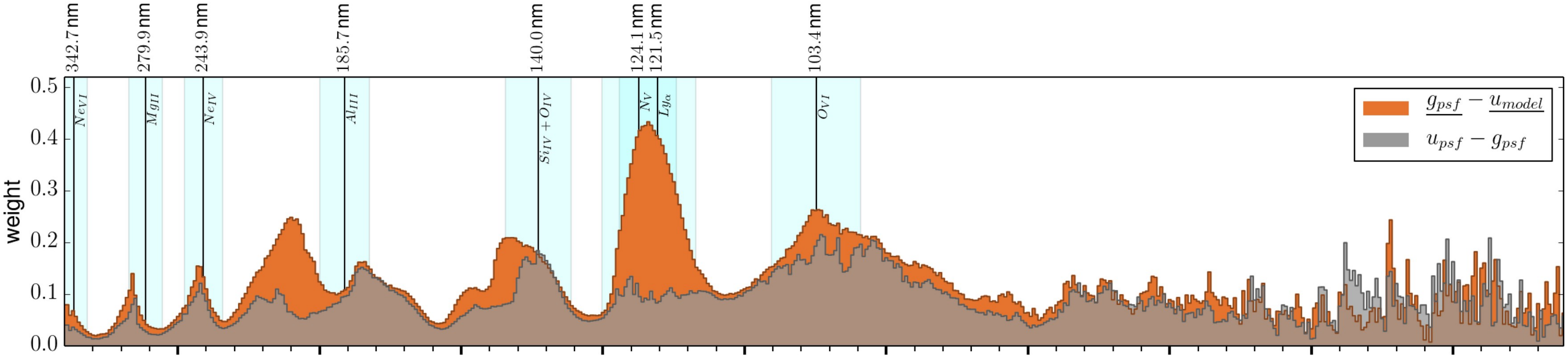
$$z=\frac{\lambda_{observed}}{\lambda_{emitted}}-1=\frac{\lambda_{filter\ intersect}}{\lambda_{qso\ emission\ line}}-1$$

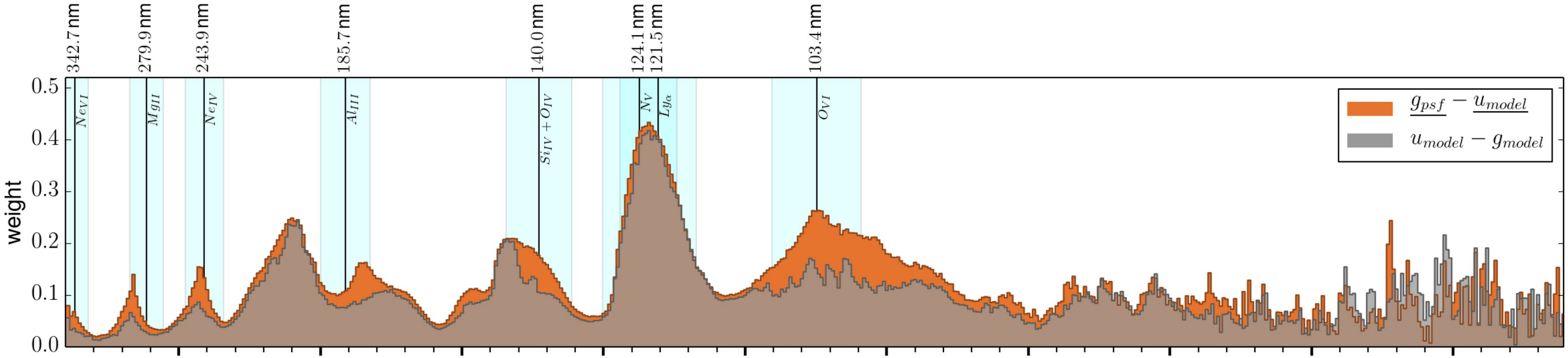
Quasar emission lines

Feature importance per bin calculated as given by RF



Finding a meaning II





Model comparison

Model 1

- Fully automatic
- Better performance
- Calibration, sharpness
- Probabilistic, multimodal
- No pre-classification
- Black-box
- Feature interpretation

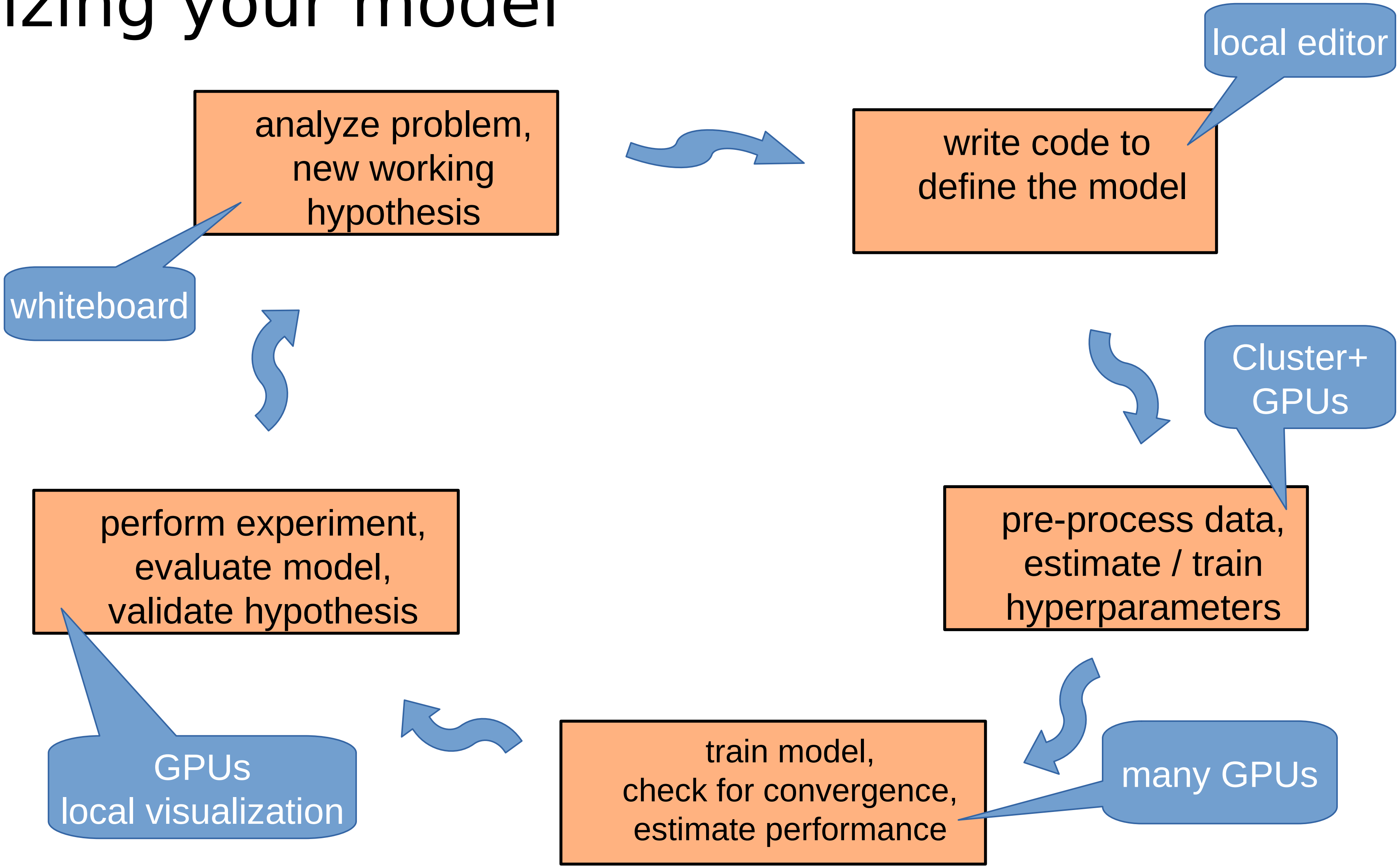
Model 2

- Better performance n >= 4
- Better performance DCMDN
- Physical interpretation
- Probabilistic calibration
- No absolute best set

ML in astronomy / Common mistakes

- Not enough training data
- Abuse of data augmentation
- Architectures too big
- Not representative training data
- Bad choice of loss function and hyperparameters

Optimizing your model



Conclusions

- Very general models
- Possibility of use in different fields → weather forecasts, medical applications
- Introduction of new evaluation tools in astronomy → CRPS, PIT
- Working in preparation of new missions → Euclid

Bibliography:

Probabilistic photometric redshift estimation in massive digital sky surveys via machine learning, A. D’Isanto, PhD thesis, https://doi.org/10.11588/heidok.00026000

A. D’Isanto and K. L. Polsterer, 2018, A&A, 609, A111

A. D’Isanto, S. Cavuoti, F. Gieseke and K. L. Polsterer, 2018, A&A, 616, A97

Rasp S., Lerch S., Monthly Weather Review, vol. 146, issue 11, pp. 3885-3900

Avati A. et al, 2019, arXiv:1806.08324

HITS

www.h-its.org

Visit us online! (and stay healthy…)

@HITStudies

/TheHITSters

/HITStudies

/the_hitsters