AUG 07, 2020

Argonne NATIONAL LABORATORY

Hyperparameter Optimization and DeepHyper

Prasanna Balaprakash and Misha Salim

Argonne National Laboratory

Team

Prasanna Balaprakash, MCS & ALCF

Romain Egele, MCS

Misha Salim, ALCF

Stefan Wild, MCS

Venkat Vishwanath, ALCF

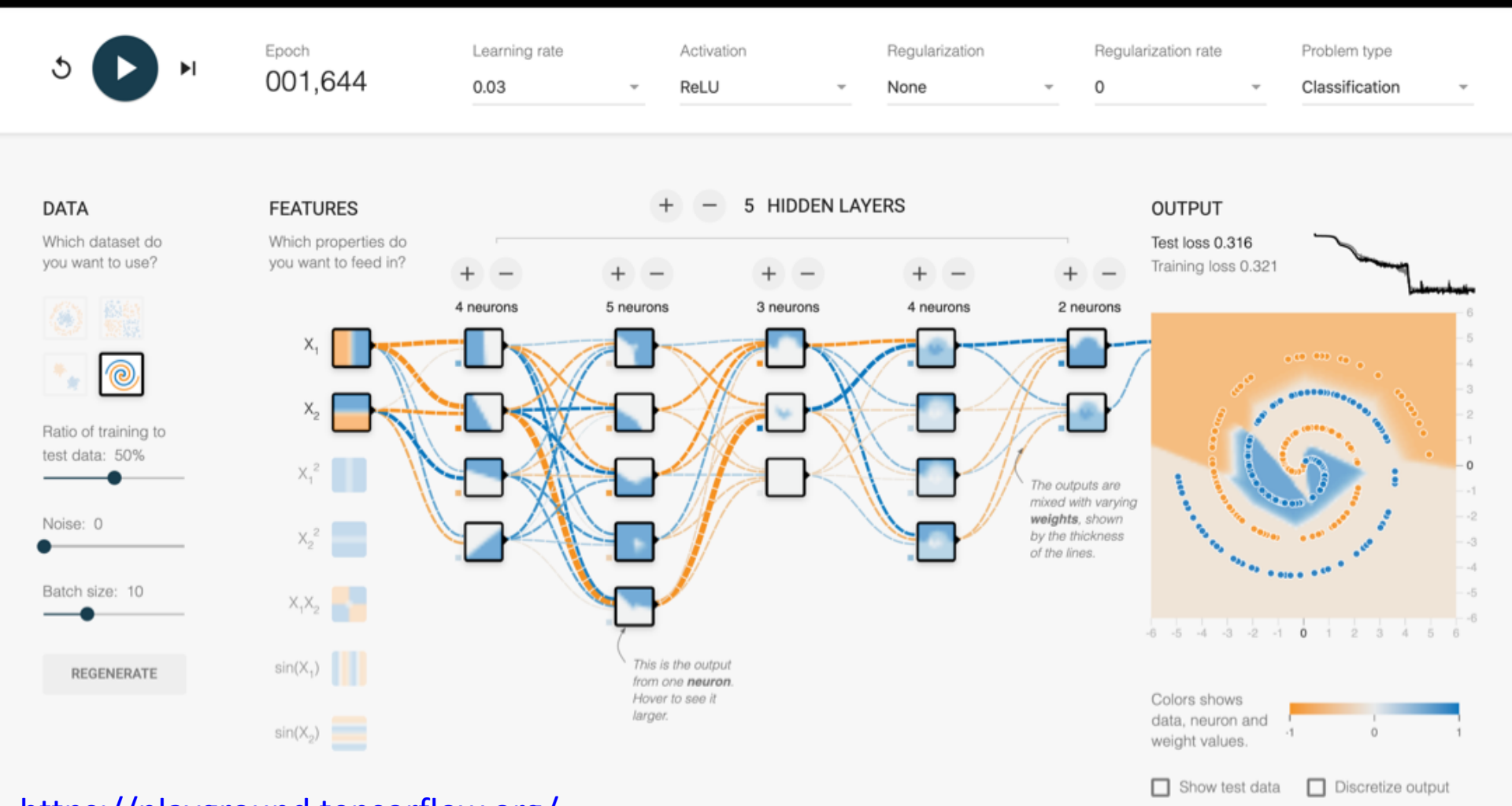
Taylor Childers, ALCF

Tom Uram, ALCF

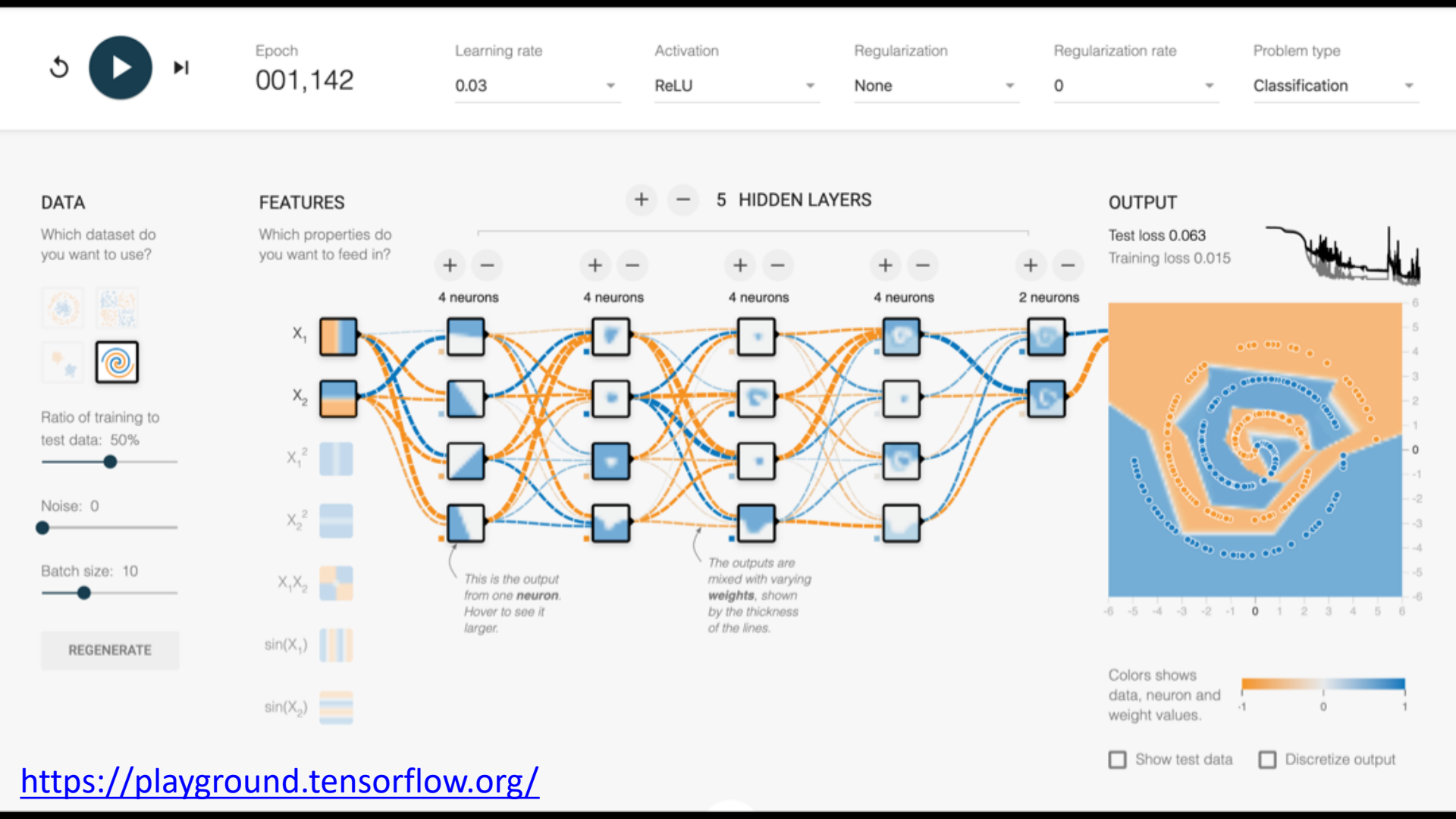
Elise Jennings, ALCF-Alumni

Romit Maulick, ALCF

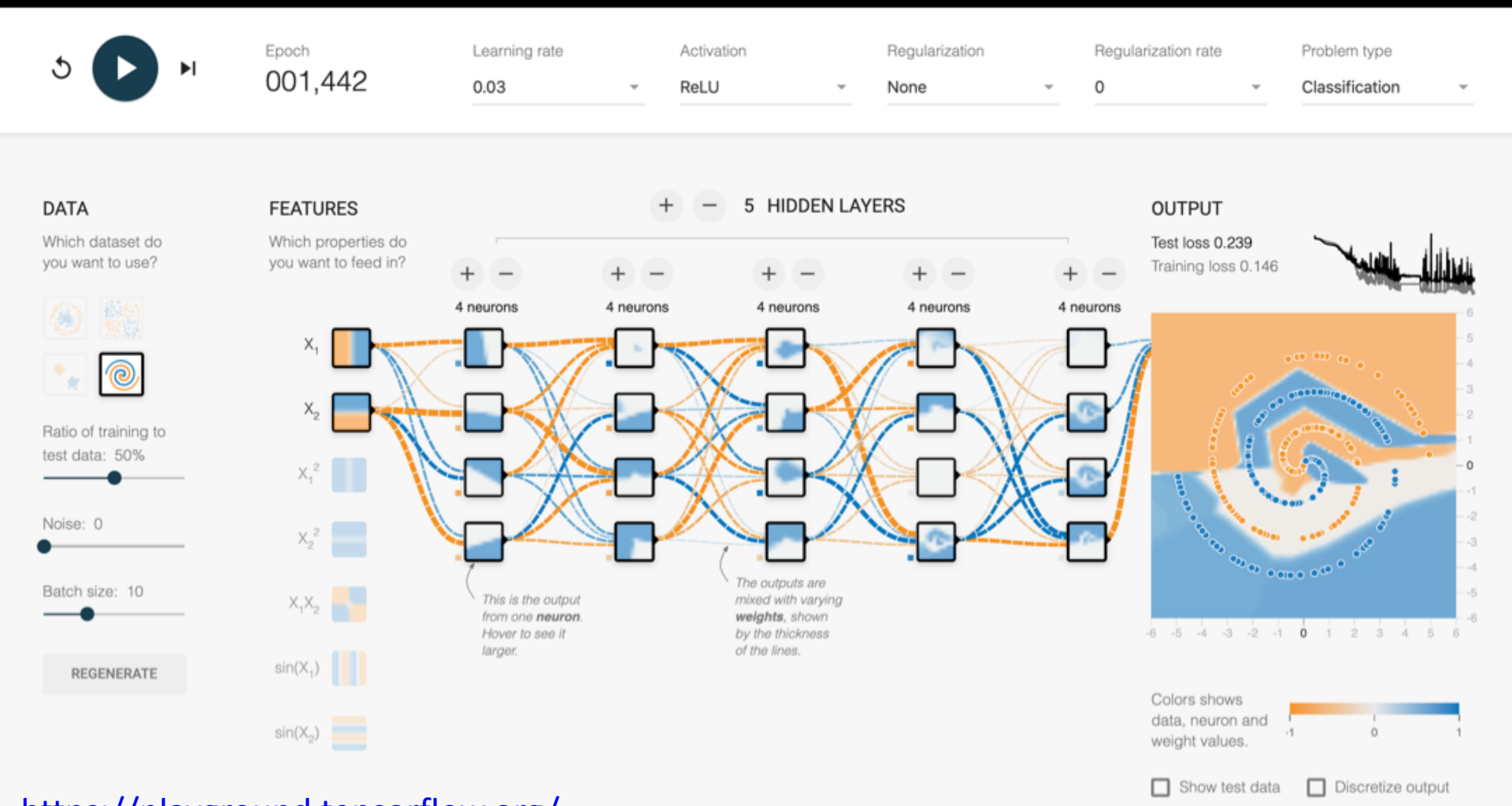
Bethany Lusch, ALCF



https://playground.tensorflow.org/

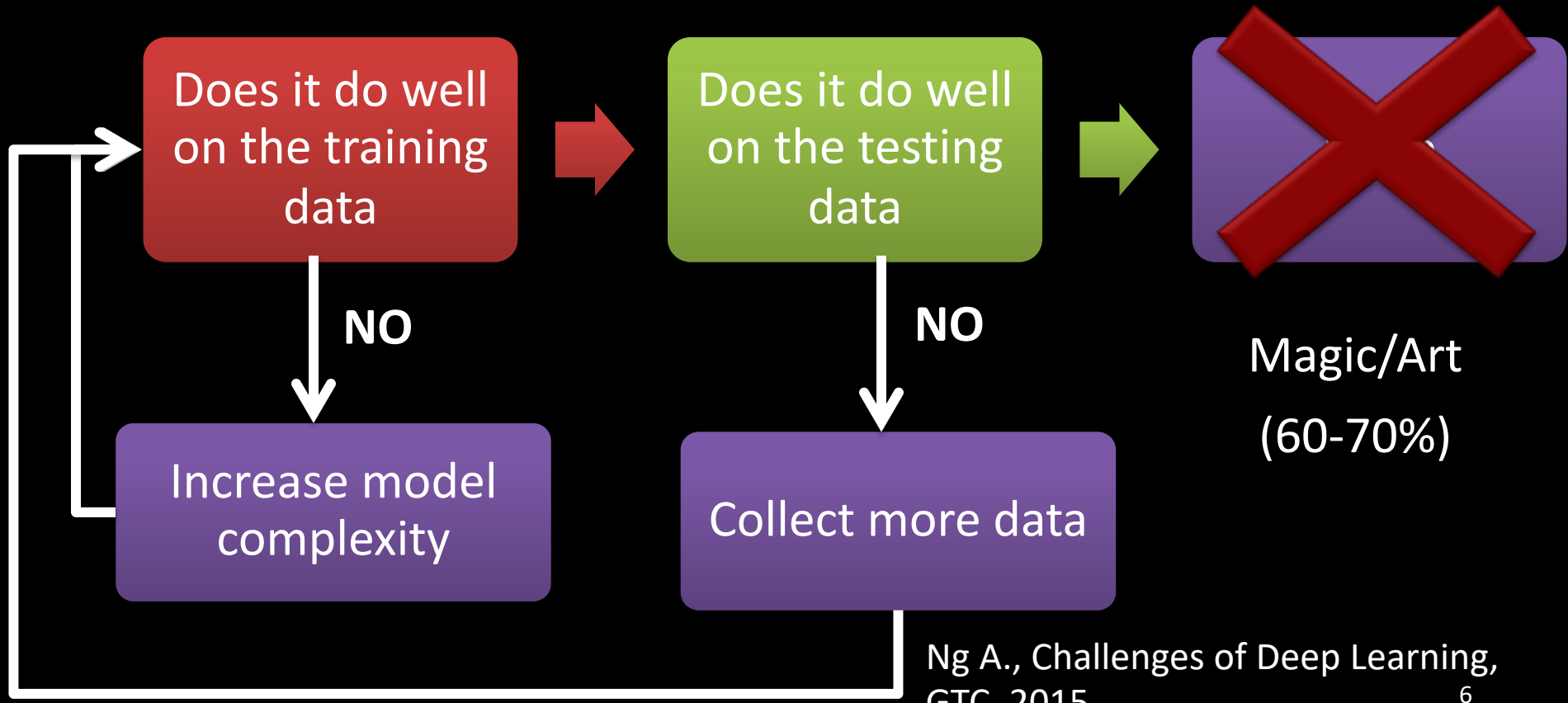


https://playground.tensorflow.org/



https://playground.tensorflow.org/

Supervised (deep) learning



Ng A., Challenges of Deep Learning, GTC, 2015

Automated machine learning for deep learning

Lower-level problem:

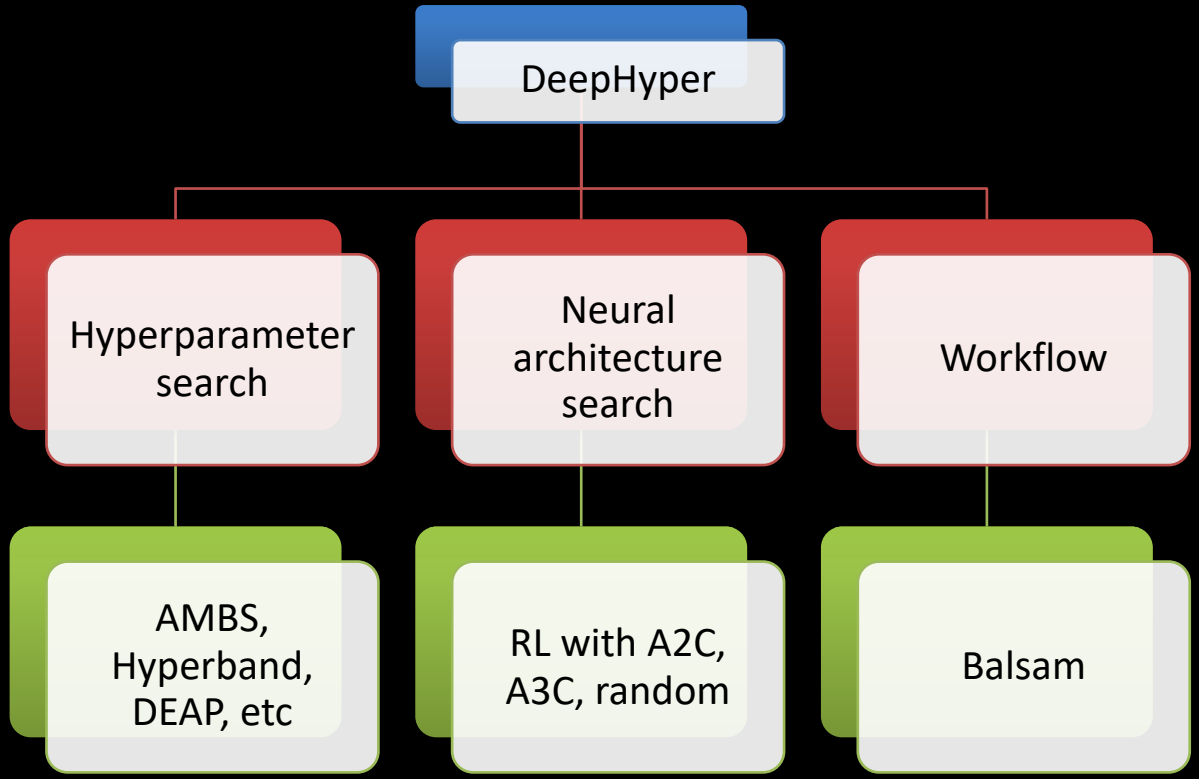
$$\text{solve} \quad \underset{w}{\text{minimize}} \ \text{err}_T\left(\left[\mathcal{X}_\mathcal{A}, \mathcal{X}_\mathcal{P}\right]; \mathcal{T}; w\right)$$

Upper-level problem:

$$\text{solve} \quad \underset{\mathcal{X}_\mathcal{A}, \mathcal{X}_\mathcal{P}}{\text{minimize}} \ \text{err}_V\left(\left[\mathcal{X}_\mathcal{A}, \mathcal{X}_\mathcal{P}\right]; \mathcal{V}; w^*\left[\mathcal{X}_\mathcal{A}, \mathcal{X}_\mathcal{P}\right]\right)$$

Architecture space Hyperparameter space

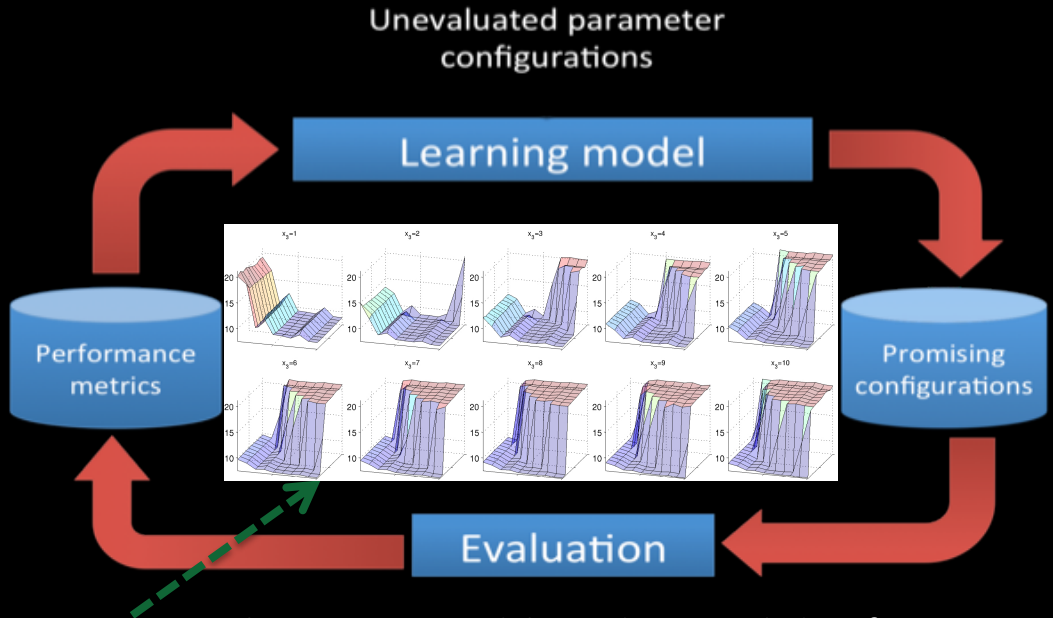
DeepHyper: Scalable AutoML package



https://github.com/deephyper/deephyper

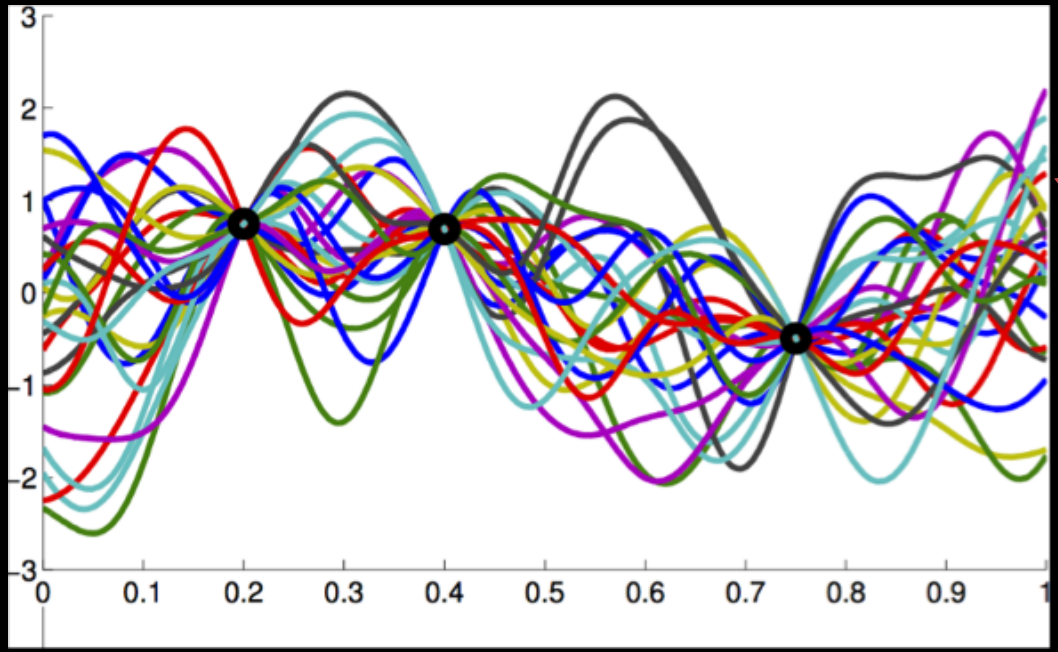
AMBS: Asynchronous model-based search

- Framework:
  - Initialization phase
    - Random or Latin hypercube sampling
  - Iterative phase
    - Fit model
    - Sample using the model



Example Surrogate Model Fitted to Sampled Performance (iterative refinement improves the learning model)

Bayesian optimization

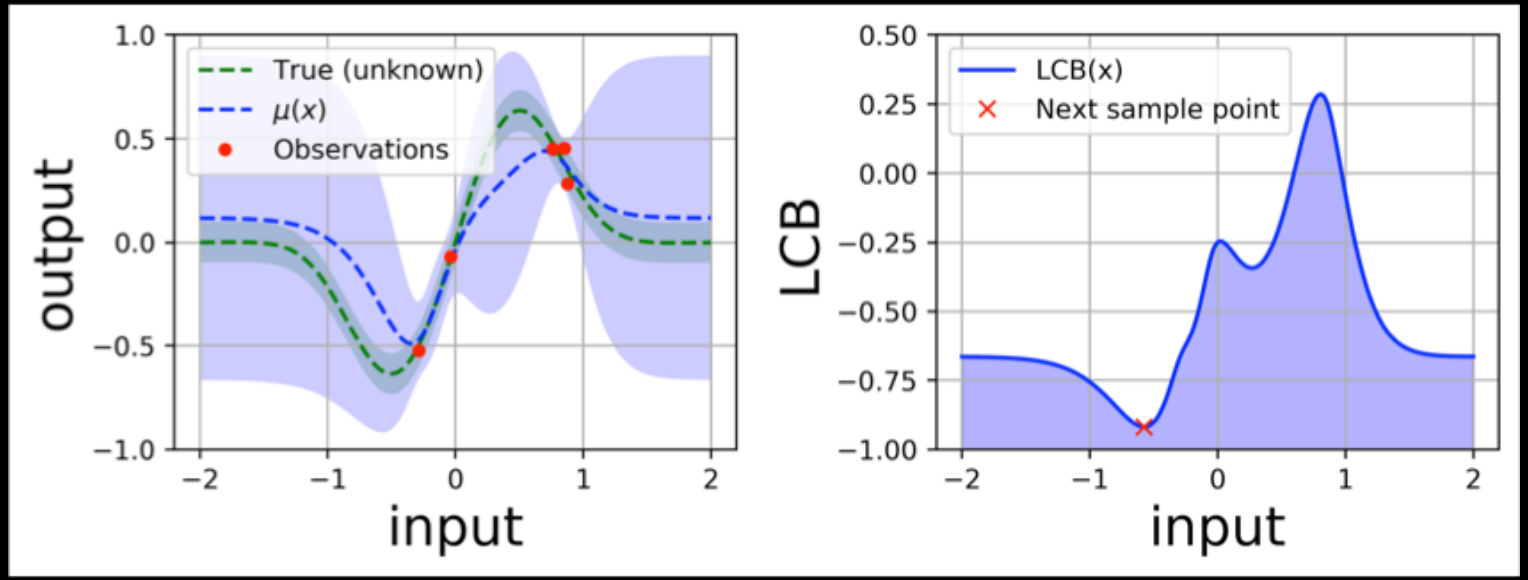


- Usual Gaussian process regression cannot handle nonordinal space natively
- Appropriate methods: random forest, extra tree regressor, Bayesian NN
- We use **Random Forest**

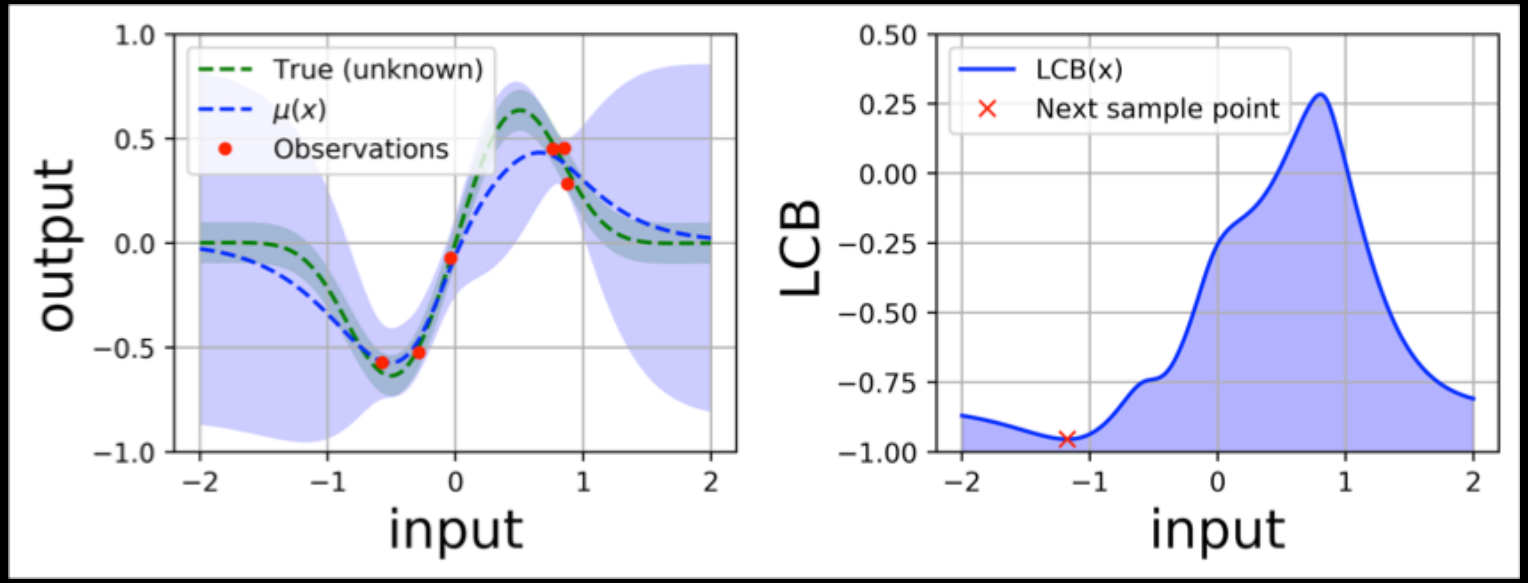
Bayesian optimization

$$LCB(x, \beta) = \mu(x) - \beta \times \sigma(x)$$

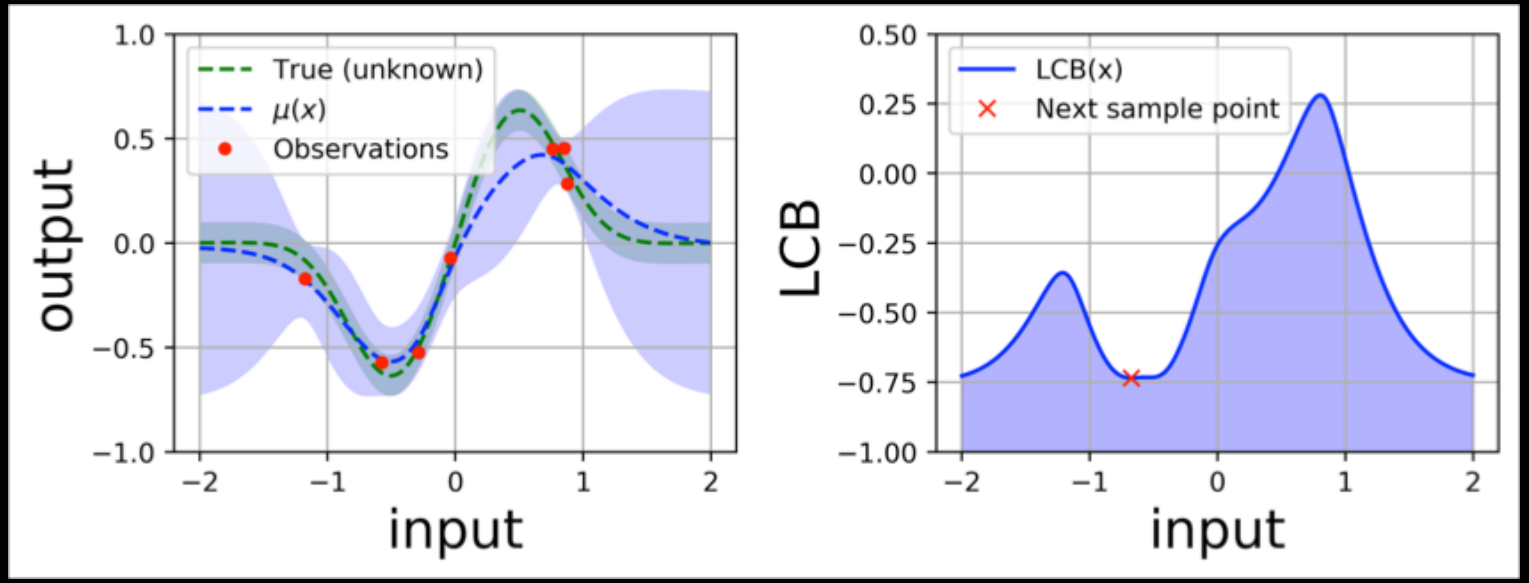
1.96



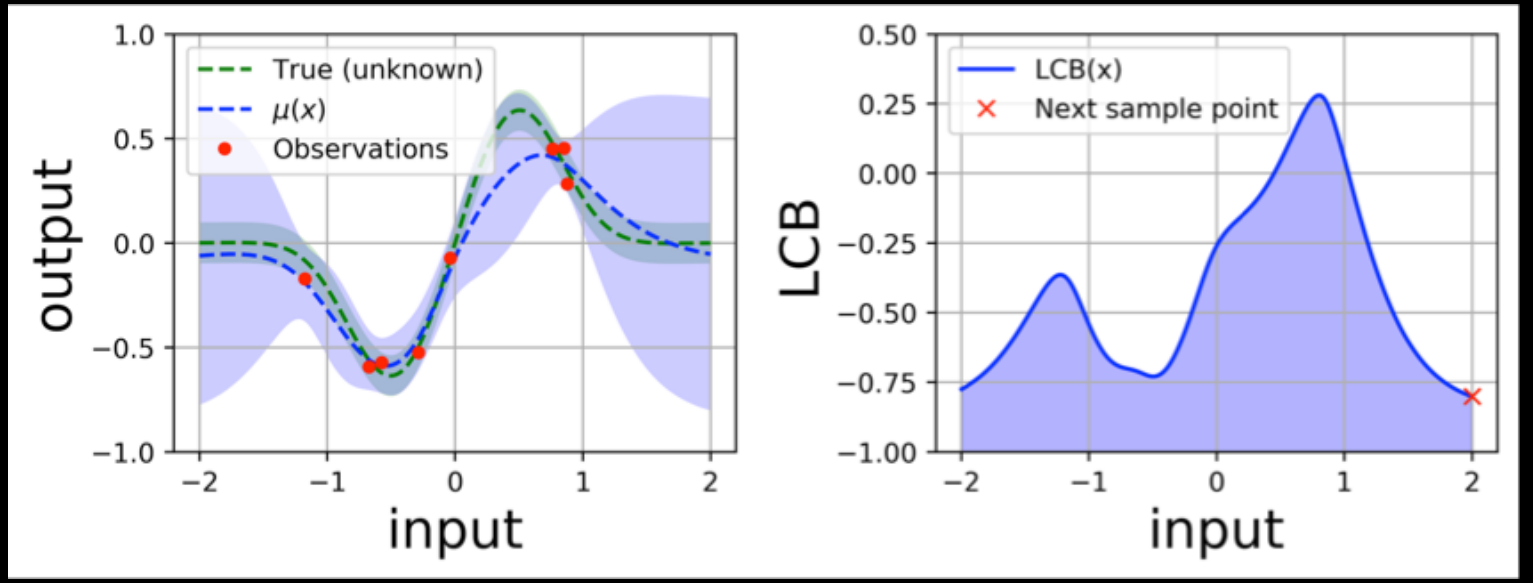
Bayesian optimization



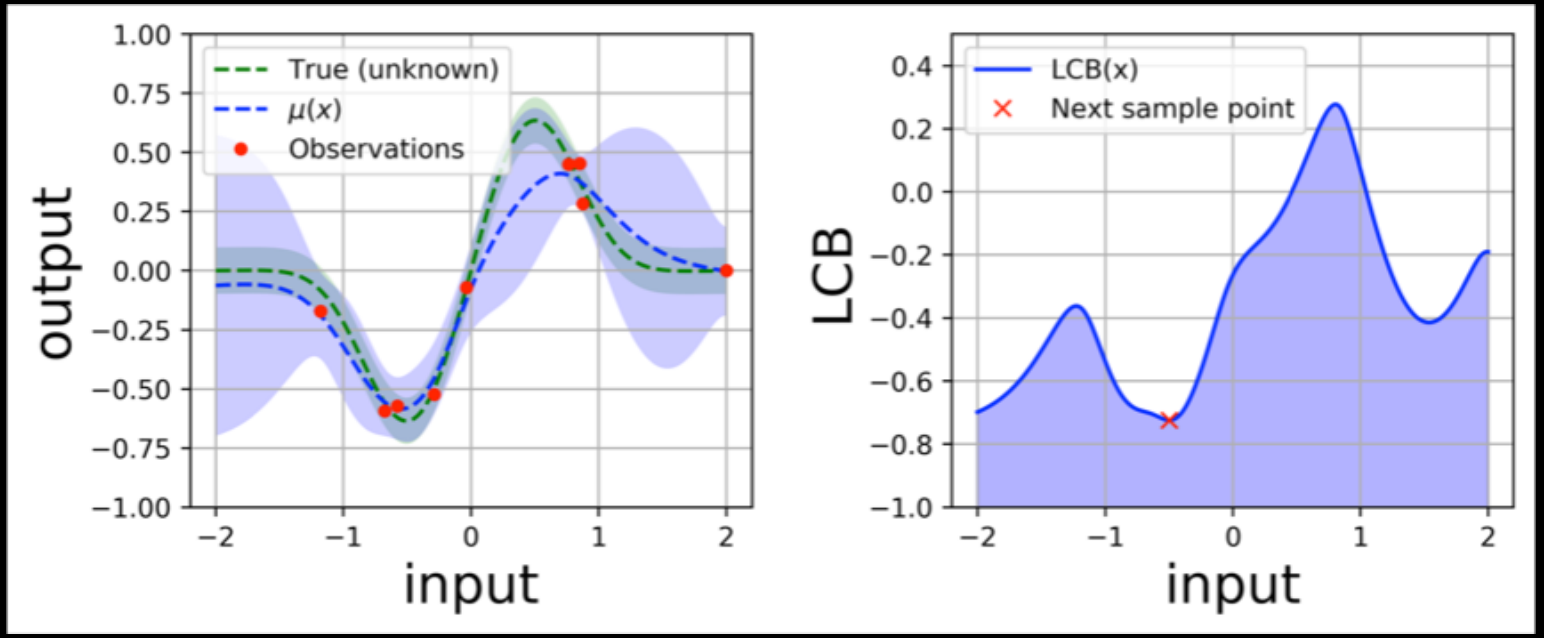
Bayesian optimization



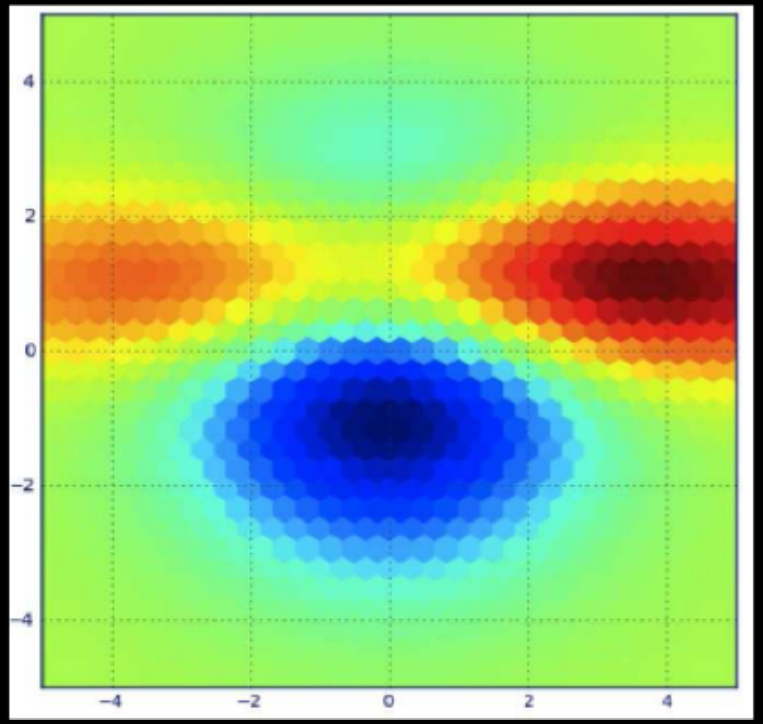
Bayesian optimization

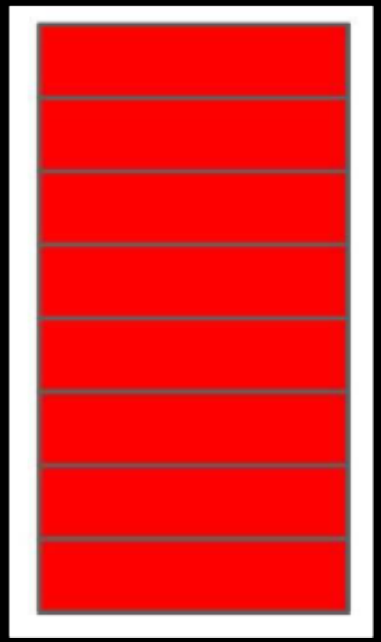


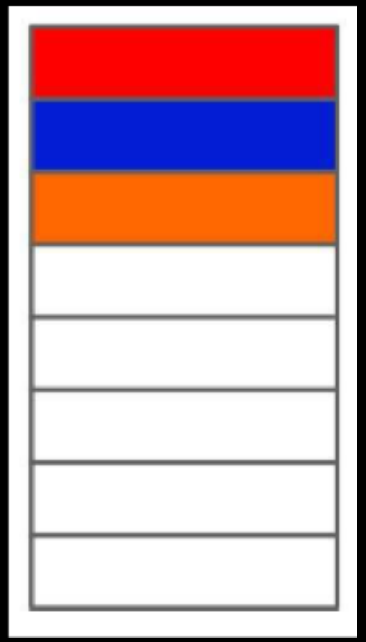
Bayesian optimization



Multipoint asynchronous sampling







Constant liar scheme for asynchronous update

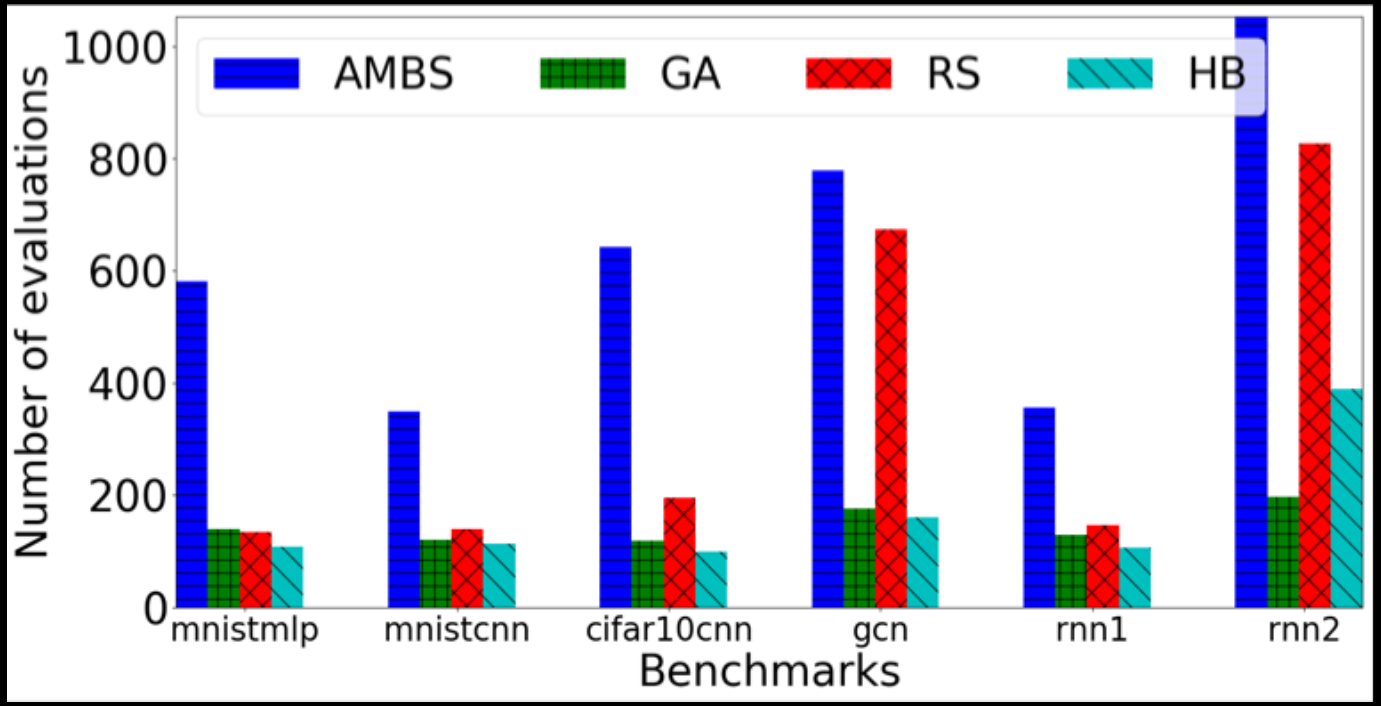
$$LCB_k(x, \beta) = \mu(x) - \beta \times \sigma(x)$$

$$f(\hat{x}_{LCB_k}) = \mu(\hat{x}_{LCB_k})$$

$$LCB_{k+1}(x, \beta) = \mu'(x) - \beta \times \sigma'(x)$$

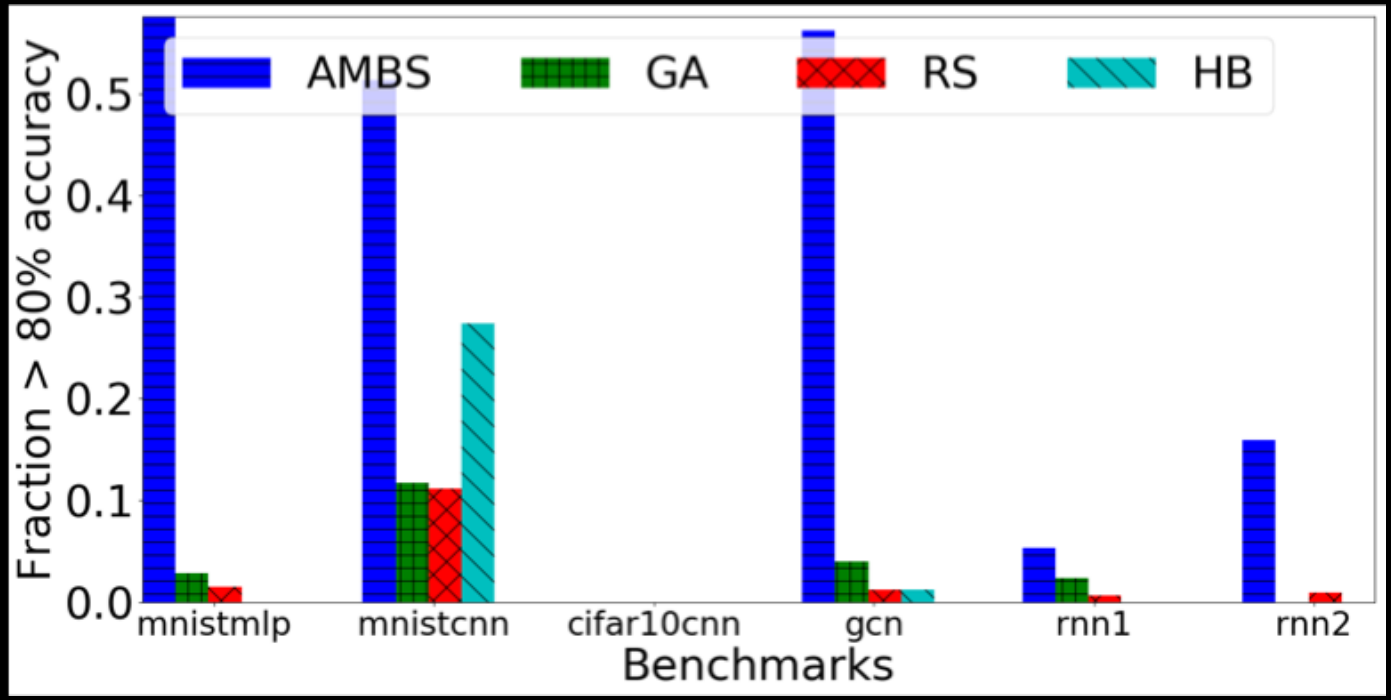
$$f(\hat{x}_{LCB_{k+1}}) = \mu'(\hat{x}_{LCB_{k+1}})$$

Comparison of search methods



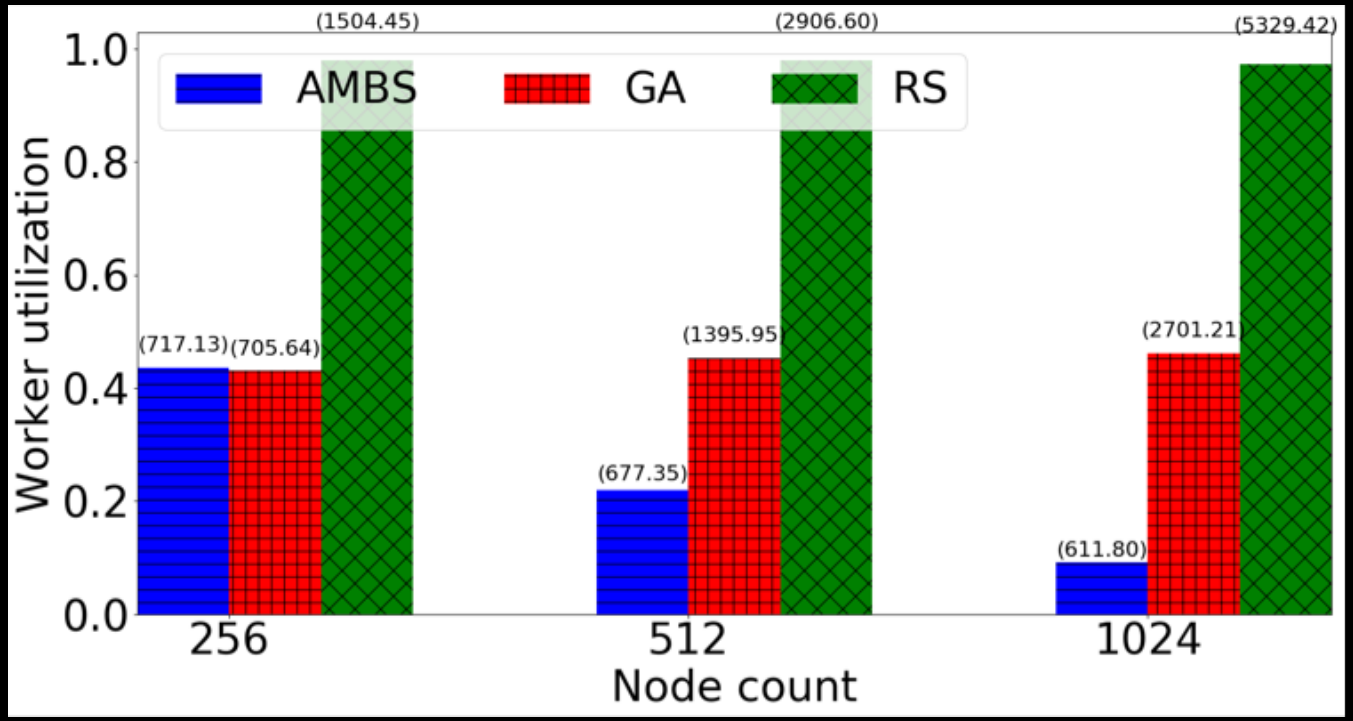
Target platform: Theta@ALCF (128 KNL nodes)
Stopping criterion: 2 hours

Comparison of search methods



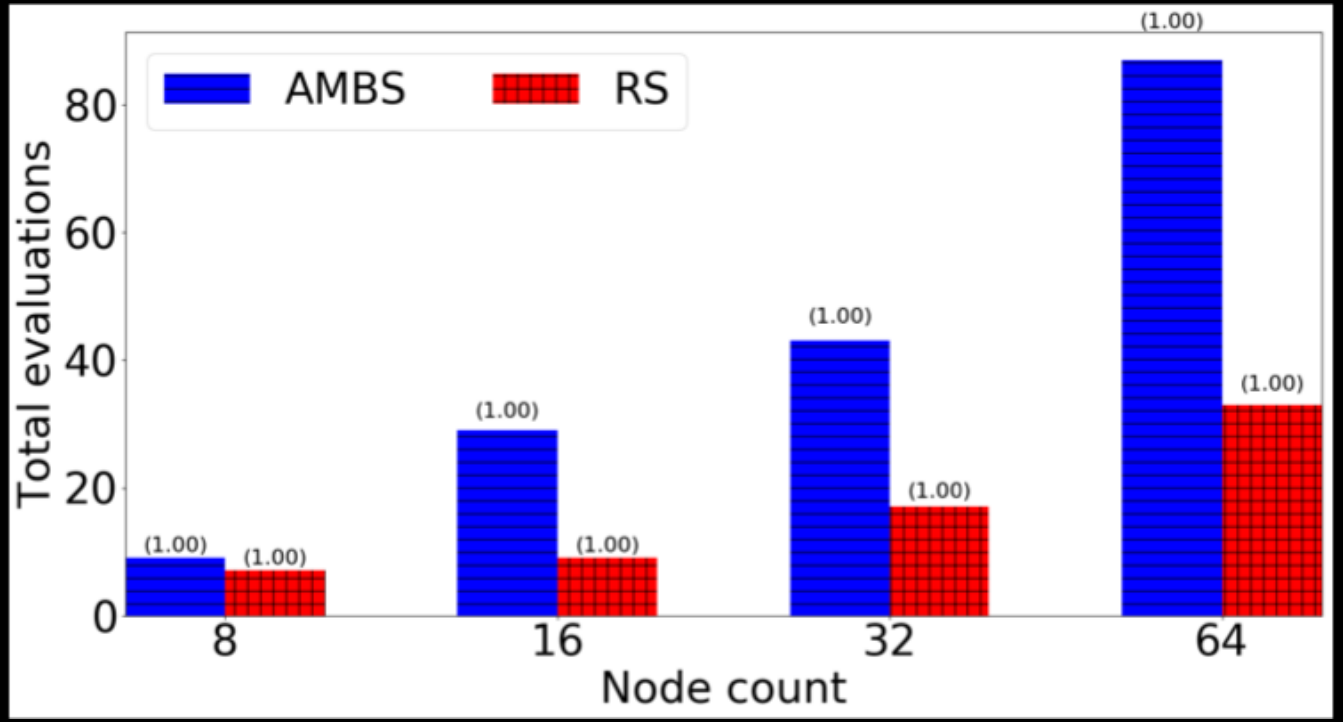
Target platform: Theta@ALCF (128 KNL nodes)
Stopping criterion: 2 hours

Scaling search methods



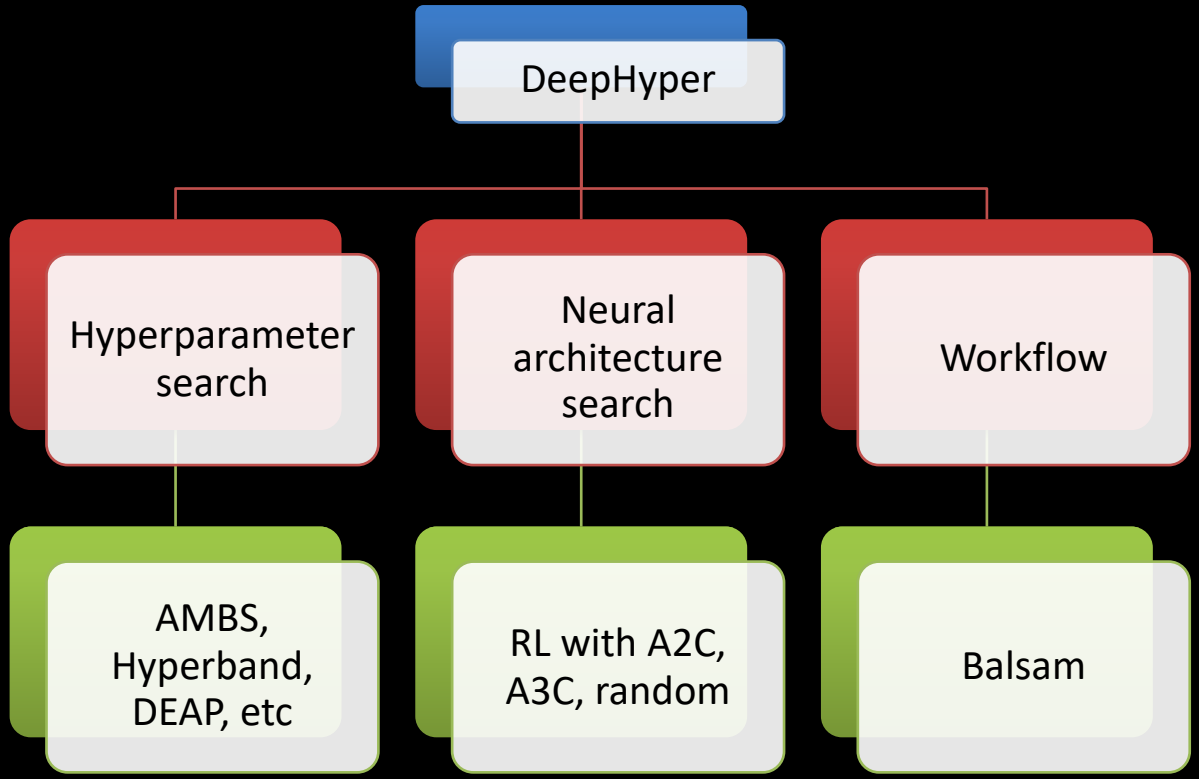
Target platform: Theta@ALCF (128 KNL nodes)
Benchmark: rnn2; Stopping criterion: 2 hours

AMBS vs RS



Target platform: Cooley (64 nodes Haswell + NVIDIA Tesla K80)
Benchmark: cifar10cnn; Stopping criterion: 1 hour

DeepHyper: Scalable AutoML package



https://github.com/deephyper/deephyper

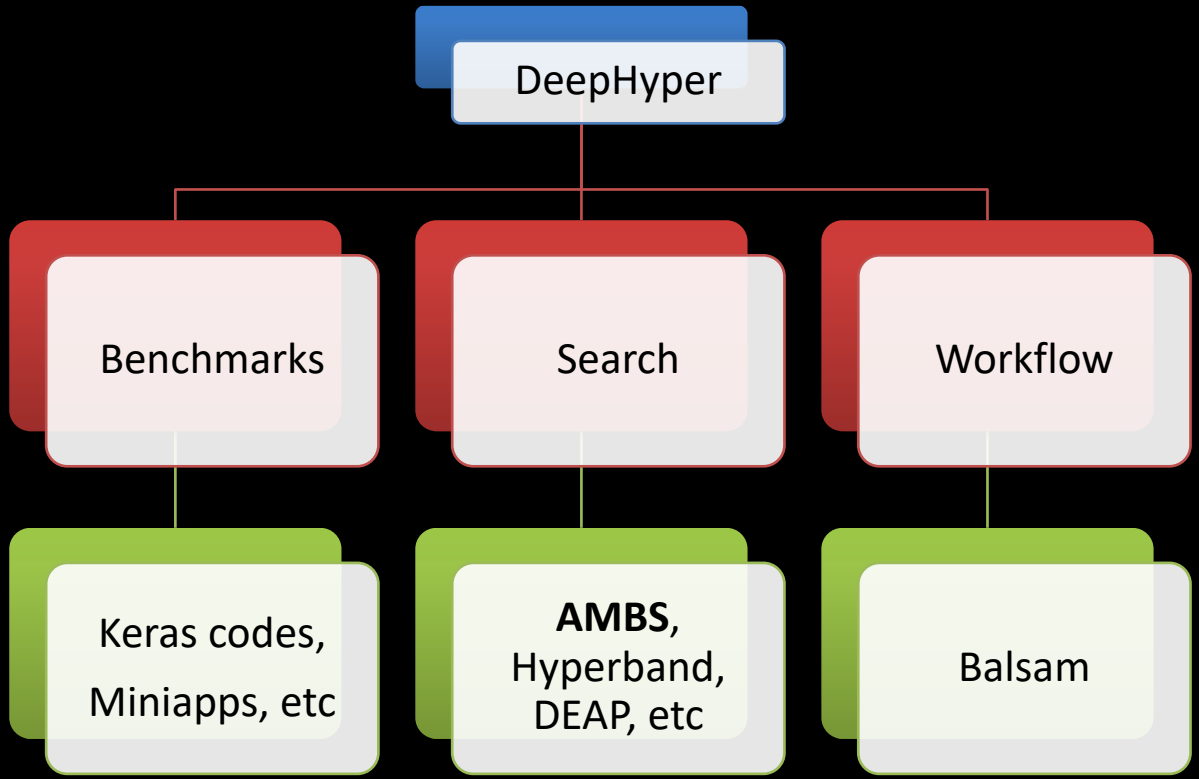
Acknowledgements

DOE Early Career Research Program, ASCR

Argonne Leadership Computing Facility

Laboratory Directed Research and Development (LDRD)

DeepHyper



https://github.com/deephyper/deephyper