# RADICAL-Learning

Radical.hpo

Shantenu Jha: Project Manager

Matteo Turilli: Technical Supervisor

### What is HPO?

- Any machine learning model has some values (hyperparameters) that need to be specified a priori before the training process.
- They help adapt the model to the data and they influence the quality of the prediction.
- Hyperparameter optimization deals with the search of the best combination of values for the given model, and there are already many methods that help us find them.

### **HPO Approaches**

- 1. Grid Search: try all combinations (brute force)
- 2. Random Search: try as many combinations as possible (better)
- 3. Informed Search (Bayesian Optimization): try the most promising combinations by reevaluating where to look next (even better)

...more

# Background

There already exists HyperSpace, a parallel Bayesian Model-Based
 Optimization (parallel Bayesian SMBO) library with one of the main goals of optimizing model performance with respect to hyperparameters. It supports Scikit-Optimize, RoBo and Hyperband

i.e, we can run a bag-of-tasks of optimizations where each task runs the Gaussian process that explores a search space (hyperspace)

### Goal

 Since our HPO module wants to integrate HyperSpace with EnTK, the logical path to follow is to treat each hyperspace optimization as an independent task, and set up a bag (stage) of tasks inside a single pipeline

 We take full advantage of EnTK and achieve concurrency at EnTK level, while still making use of HyperSpace's spaces creation and optimization through its supported HPO engines

### Requirements

#### Functional

- It must satisfy Bayesian Model-Based Optimization (SMBO) requirements:
  - Define a machine learning model (or objective function)
  - Provide with train/test datasets (if ML model selected)
  - Define a validation protocol: cross-validation (if ML model selected)
  - Define the parameter search space: upper and lower bounds for each hyperparameter
  - Define the optimization function: Gaussian process with guided sampling
- Must use RADICAL-EnTK

#### Non-Functional

- The code must be simple
- The code must be easy to maintain

#### Scalability

 It must use the maximum number of cores available on the largest XSEDE machine, without significant overhead

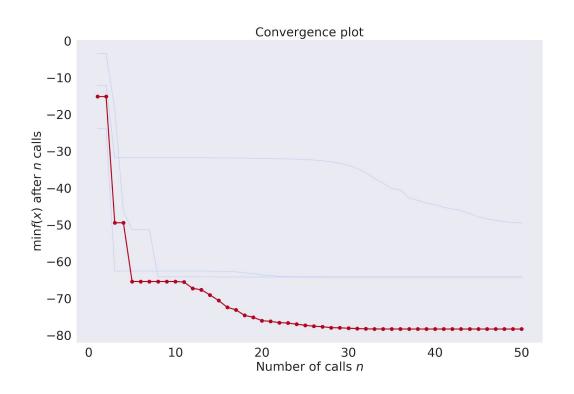
### Execution of HyperSpace

- HyperSpaces use Scikit-Optimize
- Creates combinations of hyperparameters (hyperspaces) using overlapping boundaries between hyperparameters
- Bag-of-tasks are executed with mpi4py
- Number of tasks depends on the number of hyperparameters for the model:
  - HyperSpaces = 2<sup>h</sup> where H is the number of hyperparameters
  - Avg. num of hyperparameters ~ 7-8 but depending on model can go up to 12
  - Each optimization runs for N-iterations, where N is ~100
- Tasks (Bayesian optimization step) are independent on HyperSpace, requiring each 1 MPI Rank at minimum and using 1 core each

### Minimal Example: Styblinski-Tang

Time taken: 0.8317						
Function value obtained: -46.8635	1 [ 6	0.0%]	8 [     100	0.0%] 15 [	100.0%]	22 [     100.0%]
Current minimum: -46.8635	2 [] 6	7%]		0.0%] 16 [	0.0%]	23 [ 0.0%]
Iteration No: 48 started. Searching for the next optimal point.	3 [ 6	0.0%]	10 [   0	9.7%] 17 [	0.0%	24 [ 0.0%]
Iteration No: 48 ended. Search finished for the next optimal point.	4 [ 6	0.0%]	11 [ e	9.0%] 18 [	0.0%	25 [ 0.0%]
Time taken: 0.9449		0.0%]	12 [ 0	0.0%] 19 [	0.0%	26 [ 0.0%]
Function value obtained: -47.2895	6 [ 6	0.0%]	13 [ 6	9.0%] 20 [	0.0%	27 [ 0.0%]
Current minimum: -47.2895	7 [     100	0.0%]	14 [ 6	0.0%] 21 [	0.0%	28 [ 0.0%]
Iteration No: 49 started. Searching for the next optimal point.	Mem[		3.65G/1	L26G] Tasks	: 69, 97 thr;	5 running
Iteration No: 49 ended. Search finished for the next optimal point.	Swp[		60.5M/16	.8G] Load	average: 2.55	3.80 11.29
Time taken: 0.9206				Uptim	e: 21 days, 1	3:42:11
Function value obtained: -47.7904						
Current minimum: -47.7904	PID USER	PRI	NI VIRT RES	SHR S CPU%	MEM% TIME+	Command
Iteration No: 50 started. Searching for the next optimal point.	21310 karahbit	20	0 33.0G 212M	104M R 100.	0.2 0:17.5	2 python3 benchmark.pyn
Iteration No: 50 ended. Search finished for the next optimal point.	1309 karahbit	20	0 33.0G 207M	104M R 99.6	0.2 0:17.5	5 python3 benchmark.pyn
Time taken: 1.1440 Rotate Le	ft 1308 karahbit	20				5 python3 benchmark.pyn
Function value obtained: -48.0842	21311 karahbit		0 33.0G 208M	104M R 99.6	0.2 0:17.5	0 python3 benchmark.pyn
Current minimum: -48.0842	20713 karahbit	20				
/home/karahbit/miniconda3/envs/radical.hpo/lib/python3.6/site-packages/scikit_learn-0.2	24776 root	20	0 624M 503M	2328 S 0.0	0.4 1h35:2	9 /var/lib/pcp/pmdas/proc/
1.3-py3.6-linux-x86_64.egg/sklearn/externals/joblib/initpy:15: DeprecationWarning:	7747 pcp	20	0 106M 4884	3032 S 0.0	0.0 8:44.5	1 /var/lib/pcp/pmdas/perfe
sklearn.externals.joblib is deprecated in 0.21 and will be removed in 0.23. Please imp	7750 pcp	20	0 106M 4884	3032 S 0.0	0.0 5:30.3	<pre>0 /var/lib/pcp/pmdas/perfe</pre>
ort this functionality directly from joblib, which can be installed with: pip install j	20188 root	20	0 334M 3524	2568 S 0.0	0.0 0:00.0	7 slurmstepd: [6821394.0]
oblib. If this warning is raised when loading pickled models, you may need to re-serial	2741 root	20				5 /opt/puppetlabs/puppet/b
ize those models with scikit-learn 0.21+.	21296 root	20	0 204M 3436	2516 S 0.0	0.0 0:00.0	2 slurmstepd: [6821394.4]
warnings.warn(msg, category=DeprecationWarning)	2078 root	20	0 3388M 56820	10080 S 0.0	0.0 0:05.2	5 /opt/packages/slurm/defa
Command being timed: "mpirun -n 4 python3 benchmark.pyndims 2results /hom	21049 nscd	20	0 1502M 2408	1404 S 0.0	0.0 0:41.3	8 /usr/sbin/nscd
e/karahbit"	21287 karahbit					1 mpiexec.hydra -n 4 pytho
User time (seconds): 0.02	21303 karahbit	20				1 /opt/intel/compilers_and
System time (seconds): 0.02	5820 root	20				9 /usr/bin/python2 -Es /us
Percent of CPU this job got: 0%	1 root	20				3 /usr/lib/systemd/systemd
Elapsed (wall clock) time (h:mm:ss or m:ss): 0:30.16	5996 root	20				5 /usr/bin/python2 -Es /us
Average shared text size (kbytes): 0	5341 polkitd					2 /usr/lib/polkit-1/polkit
Average unshared data size (kbytes): 0	18764 root	20				4 /usr/sbin/irqbalancef
Average stack size (kbytes): 0	6276 root	20				2 hpasmlited -f /dev/hpilo
Average total size (kbytes): 0	5874 root	20				3 /usr/sbin/rsyslogd -n
Maximum resident set size (kbytes): 5240	6754 root	20				8 /opt/hp/hp-health/bin/hp
Average resident set size (kbytes): 0	5760 munge	20				3 /usr/sbin/mungednum-t
Major (requiring I/O) page faults: 0	20177 root	20				3 slurmstepd: [6821394.ext
Minor (reclaiming a frame) page faults: 4484	6278 root	20				2 hpasmlited -f /dev/hpilo
Voluntary context switches: 127	6260 root	20				5 /var/lib/pcp/pmdas/linux
Involuntary context switches: 64	6137 root	20				3 /usr/sbin/automountfo
Swaps: 0	5863 root	20				8 /usr/sbin/rsyslogd -n
File system inputs: 0	20197 karahbit					
File system outputs: 0	20196 root	20				2 slurmstepd: [6821394.0]
Socket messages sent: 0	11915 root	20				8 /opt/puppetlabs/puppet/b
Socket messages received: 0	21050 nscd	20				0 /usr/sbin/nscd
Signals delivered: 0	6149 pcp	20				5 /usr/libexec/pcp/bin/pmc
Page size (bytes): 4096	21223 postfix	20				4 qmgr -l -t unix -u
Exit status: 0	7751 root	20				3 perl /var/lib/pcp/pmdas/
<pre>(radical.hpo) [karahbit@r508 hyperdrive]\$ /usr/bin/time -v mpirun -n 4 python3 benchmar k.pyndims 2results /home/karahbit</pre>	2743 root 2744 root					<pre>0 /opt/puppetlabs/puppet/b 9 /opt/puppetlabs/puppet/b</pre>
K.pyndims 2lesdits /nome/karanbit	2744 root 2762 root					8 /opt/puppetlabs/puppet/b
[0] <bit@br018:~ "two"="" 07:28="" 14-nov-19<="" benchmarks="" hyperdrive*="" hyperspace="" styblinskitang="" td=""><td></td><td></td><td></td><td></td><td></td><td></td></bit@br018:~>						
to Street of Annual Control of the C	Haneap Pasecul	36	arone i iliteri i oli i	SUI LBY	I NILGO - PONIC	o i Mill Hoult

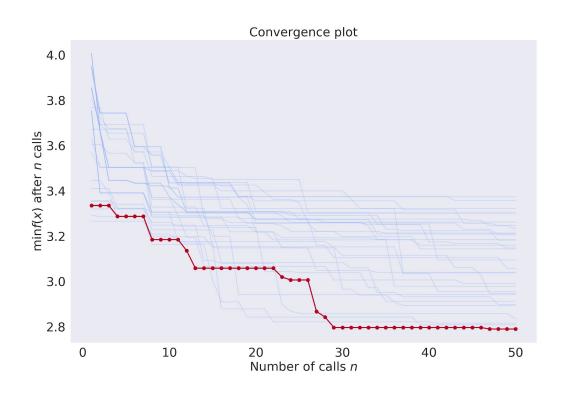
### Minimal Example: Styblinski-Tang



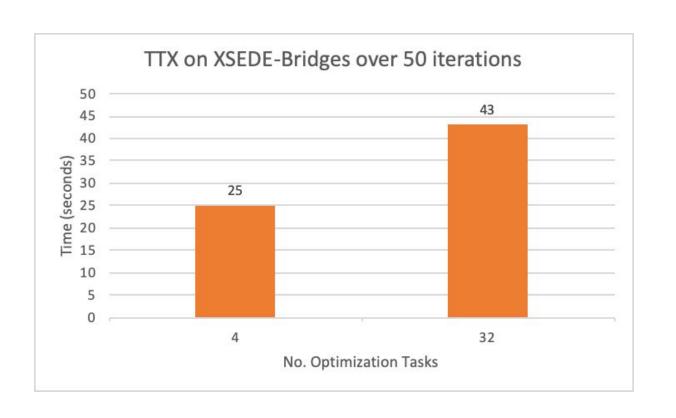
### **ML Example: Gradient Boosted Trees**

ite-packages/scikit_optimize-0.5.2-py3.6.egg/skopt/optimier/optimizer.py:399: UserWarning: The objective has been evaluated at this point before.  warnings.warn("The objective has been evaluated " //nome/karahbit/miniconde3/envs/radical.hpo/lib/python3.6/i ite-packages/scikit_optimize-0.5.2-py3.6.egg/skopt/optimier/optimizer.py:399: UserWarning: The objective has been evaluated at this point before.  warnings.warn("The objective has been evaluated " //nome/karahbit/miniconda3/envs/radical.hpo/lib/python3.6/site-packages/scikit_learn-0.21.3-py3.6-linux-x86.64.egg/silearn/externals/jobilb/ init.pv:15: Deprecation/warning	1 [100.0] 2 [100.0] 3 [100.0] 5 4 [199.3] 5 5 [100.0] 6 6 [100.0] 7 [100.0] Mem[    5 Swp[	8 [10] 9 [10] 10 [11] 11 [10] 12 [10] 13 [10] 14 [10] 5.126/1	00.0] 00.0] 99.3] 00.0] 00.0] 00.0] 126G] 6.8G]	Load av	3] 23 .0] 24 .0] 25 .0] 26	6 [ 0. 7 [ 0. 8 [ 0. thr; 17	.0] .0] .0] .0] .0] .0] 7 runnir 34 10.71		9 [ 10 [ 11 [ 12 [ 13 [ 14 [	100.0]  99.3]  99.3]  99.3] 100.0] 100.0] /126G] 16.8G]	Load ave	5] 23 [ 0] 24 [  0] 25 [ 0] 26 [ 0] 27 [	9.04 11.07
sklearn.externals.joblib is deprecated in 0.21 and will be		PRI	NI VIRT	RES	SHR S	CPU% ME	M% T1	PID USER	PRI	NI VIRT	RES	SHR S CPU%	MEM% TI
e removed in 0.23. Please import this functionality direct			0 33.40		41760 R		0:0			0 33.40		1760 R 100.	
ly from joblib, which can be installed with: pip install ;	21008 karahbit	20	0 33.40	153M	41772 R 3	100. 6	0:0	27870 karahbi	t 20	0 33.40	146M 4	1724 R 100.	0.1 0:0
oblib. If this warning is raised when loading pickled mode			0 33.40	149M	41772 R :	100. (	0:6	27866 karahbi	t 20	0 33.46	148M 4	1716 R 100.	0.1 0:0
ls, you may need to re-serialize those models with scikit-					41760 R :			27861 karahbi		0 33.40		1760 R 100.	
learn 0.21+.	21009 karahbit		0 33.40		41764 R :			27865 karahbi		0 33.40		1760 R 100.	
warnings.warn(msg, category=DeprecationWarning)	21000 karahbit				41760 R :			27874 karahbi				1716 R 100.	
/home/karahbit/miniconda3/envs/radical.hpo/lib/python3.6/s			0 33.40					27862 karahbi		0 33.40		1728 R 100.	
ite-packages/hyperspaces-0.3.0-py3.6.egg/hyperspace/space/								27869 karahbi				1760 R 100.	
skopt/space.py:173: UserWarning: Each hyperspace contains	21011 karahbit 20996 karahbit							27871 karahbi		0 33.40		1620 R 100. 1716 R 99.E	
a single value.  Command being timed: "mpirun -n 32 python3 gbm.py	20996 karandit 20997 karahbit		0 33.40					9 27864 karahbi 9 27863 karahbi		0 33.46		1716 R 99.5	
results /home/karahbit/results/gbm"	21002 karahbit							27867 karahbi		0 33.46		1692 R 99.5	
User time (seconds): 0.01	21002 karahbit							27872 karahbi		0 33.46		1620 R 99.5	
System time (seconds): 0.03	21003 karahbit							27868 karahbi		0 33.40		1764 R 99.5	
Percent of CPU this job got: 0%	21004 karahbit		0 33.40					27873 karahbi		0 33.40		1716 R 99.5	
Elapsed (wall clock) time (h:mm:ss or m:ss): 0:42								27860 karahbi		0 33.46		1748 R 98.8	
49	14714 karahbit		0 120N		1464 R			22111 karahbi		0 120M		1460 R 2.6	
Average shared text size (kbytes): 0	20984 root	20	0 203N					7750 pcp	20	0 106M	4884	3032 S 0.7	0.0 5:3
Average unshared data size (kbytes): 0	20991 karahbit	20	0 16716	1688	1336 S	0.0	0.0 0:6	27847 root	20	0 204M	3440	2516 S 0.6	0.0 0:0
Average stack size (kbytes): 0	5799 munge	20	0 734N	1 1876	1524 S	0.0	0.0 0:1	7747 pcp	20	0 106M	4884	3032 S 0.6	0.0 8:4
Average total size (kbytes): 0	20974 karahbit	20	0 14964	1736	1388 S	0.0	0.0 0:6	2078 root	20	0 3388M	56860 1	0080 S 0.6	0.0 0:0
Maximum resident set size (kbytes): 5260	20975 karahbit		0 308N					27853 karahbi		0 16716		1284 S 0.6	
Average resident set size (kbytes): 0	6139 root	20	0 1198N					24776 root	20	0 624M		2328 S 0.6	
Major (requiring I/O) page faults: 0	14429 root	20	9 266N					2741 root	20	0 379M		8644 S 0.6	
Minor (reclaiming a frame) page faults: 4487	28231 root	20	0 378N					2 21839 root	20	0 267M		2428 S 0.6	
Voluntary context switches: 248	26260 pcp	20	0 94916				9.0 0:6		20	0 187M		2616 S 0.6	
Involuntary context switches: 64	14430 root	20	0 2661					18764 root	20	0 21768		996 S 0.6	
Swaps: 0 File system inputs: 0	6225 root 6223 root	20 20	0 413N 0 413N					5820 root 7751 root	20 20	0 561M 0 203M		6316 S 0.6 3452 S 0.6	
File system outputs: 0	8553 root	20	0 21768				3.0 4:1		20	0 413M		776 S 0.6	
Socket messages sent: 0	6146 root	20	0 93532				0.0 1:4		20	0 561M		6316 S 0.6	
Socket messages received: 0	14436 root	20	0 333N				0.0 0:0		20	0 27656		1404 S 0.6	
Signals delivered: 0	1 root	20	0 187N				3.0 7:4		20	0 735M		1932 S 0.6	
Page size (bytes): 4096	32413 root	20	0 295N		9984 S			2 24019 root	16	-4 51972		2956 S 0.6	
Exit status: 0	5749 root	20		12028			3.0 2:4		20			5356 S 0.6	
(radical.hpo) [karahbit@r507 examples]\$ /usr/bin/time -v r		20	0 3379N					5863 root	20	0 405M		4848 S 0.6	
pirun -n 32 python3 gbm.pyresults /home/karahbit/resu		20	0 734N	1876	1524 S	0.0	0.0	16228 nslcd	20	0 449N	6080	3492 S 0.6	0.0 0:3
ts/gbm	5811 munge							6754 root	20	0 10640	616	488 S 0.6	0.0 0:4
	2830 root							21841 root	20			2428 S 0.6	
[0] <it@br018:~ "two"="" 08:14="" 14-nov-19<="" examples*="" hyperspace="" td=""><td>F1Help F2Setu</td><td>p F3Sea</td><td>rch<mark>F4</mark>Fil</td><td>lter F5 Tr</td><td>ee F6So</td><td>rtBy<b>F7</b>N</td><td>Nice -<mark>F</mark>8</td><td><b>F1</b>Help <b>F2</b>Set</td><td>up F3Se</td><td>arch<mark>F4</mark>Fil</td><td>ter<mark>F5</mark>Tre</td><td>e F6SortBy</td><td>F7Nice -F8</td></it@br018:~>	F1Help F2Setu	p F3Sea	rch <mark>F4</mark> Fil	lter F5 Tr	ee F6So	rtBy <b>F7</b> N	Nice - <mark>F</mark> 8	<b>F1</b> Help <b>F2</b> Set	up F3Se	arch <mark>F4</mark> Fil	ter <mark>F5</mark> Tre	e F6SortBy	F7Nice -F8

### **ML Example: Gradient Boosted Trees**



## TTX on XSEDE-Bridges over 50 iterations



### References

https://scikit-optimize.github.io

https://examples.dask.org/machine-learning/hyperparam-opt.html

https://towardsdatascience.com/hyperparameter-optimization-in-python-part-0-introduction-c4b66791614b

https://github.com/jdakka/hyperspace-RCT

https://github.com/yngtodd/hyperspace

https://buildmedia.readthedocs.org/media/pdf/hyperspace/latest/hyperspace.pdf

https://towardsdatascience.com/an-introductory-example-of-bayesian-optimization-in-python-with-hyperopt-aae40fff4ff0

https://medium.com/criteo-labs/hyper-parameter-optimization-algorithms-2fe447525903