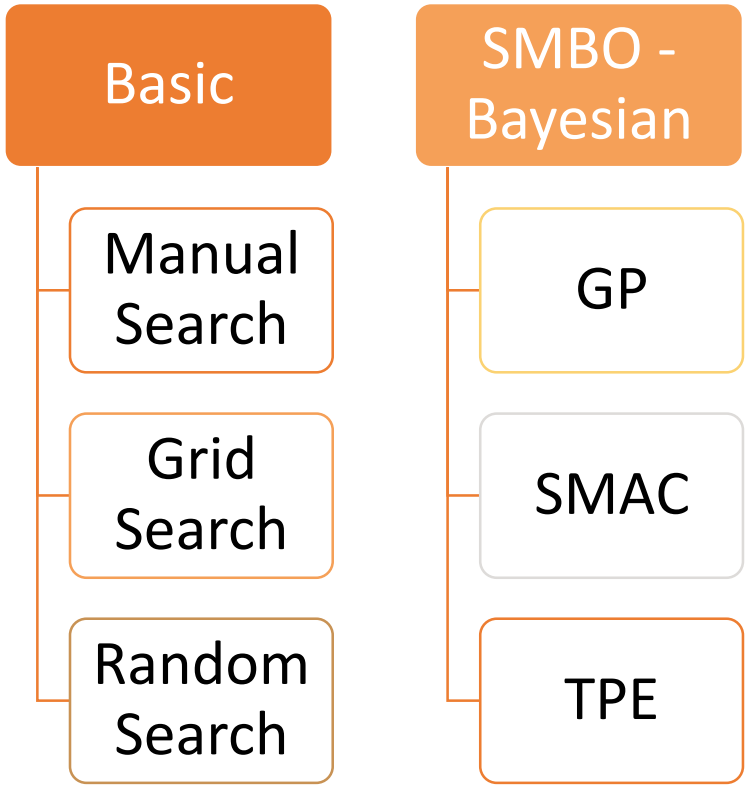
Train In Data

Basic Search vs Bayesian Optimization

Search strategies



Searches in a nutshell

	Grid Search	Random Search	Bayesian
Parallelize	✓	✓	ish
Effective in high dimension	✗	✓	✓
Effective in (very) low dimension	✓	-	-
Suited for continuous hyperparameters	✗	✓	✓
Complex Machine Learning models	✗	ish	✓
Resource intensive	-	✓	✗
Wall-clock time intensive	-	✗	✓
Hyperparameter values	Manually defined	Drawn from a distribution	Drawn from a distribution

Bayesian searches in a nutshell

	GP	SMAC	TPE
Approximate	f(x)	f(x)	hyperparam distribution
Surrogate	Gaussian processes	Random Forests	Density kernels (tree parzen estimators)
Hyperparam interaction	✓	✓	✗
Speed	Less fast	Fast	fast
Implementation	Scikit-Optimize	Scikit-Optimize	Hyperopt, Optuna
Nested spaces	✗	✗	✓
Performance	Second best	best	Third best

Bayesian searches in a nutshell

	GP	SMAC	TPE	Random
Approximate	f(x)	f(x)	hyperparam distribution	-
Surrogate	Gaussian processes	Random Forests	Density kernels (tree parzen estimators)	-
Hyperparam interaction	✓	✓	✗	✓
Speed	Less fast	Fast	fast	-
Implementation	Scikit-Optimize	Scikit-Optimize	Hyperopt, Optuna	Sklearn, **Hyperopt**, **Optuna**, others
Nested spaces	✗	✗	✓	✓
Performance	Second best	Best	Third best	Best*

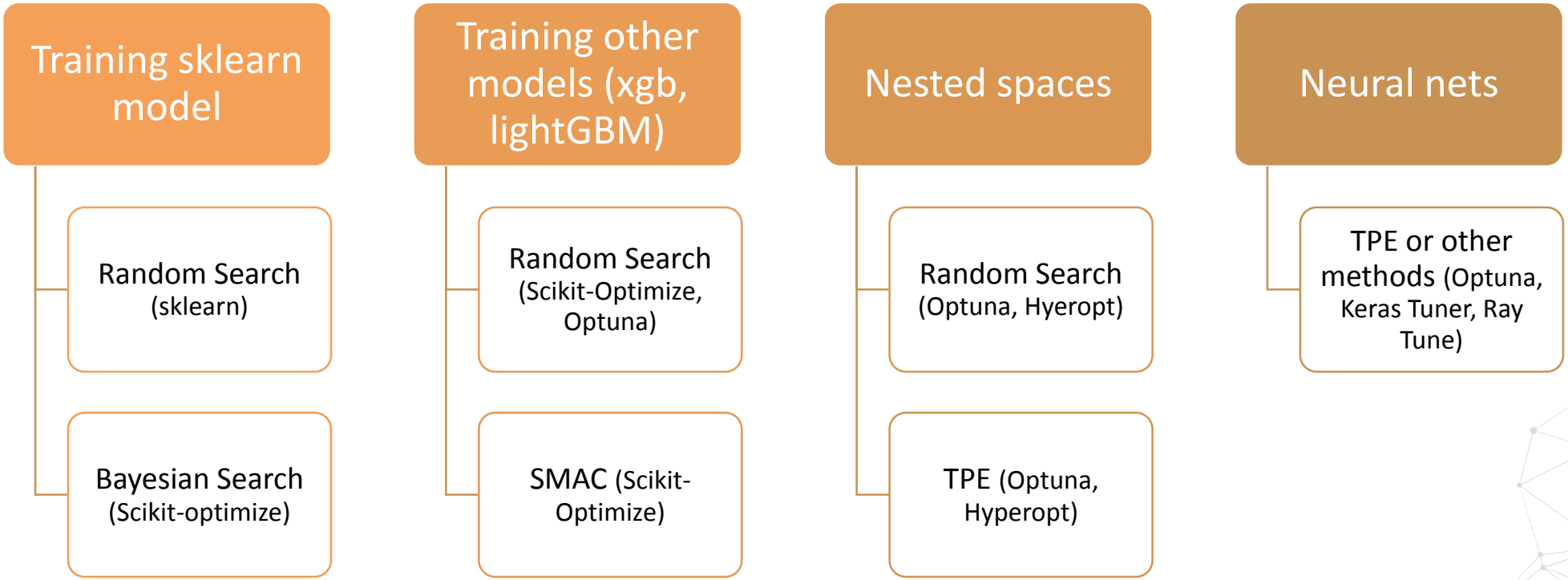
* If allowed sufficient iteractions

Bayesian searches in a nutshell

Experiment	#evals	SMAC Valid.Error	SMAC Best Value	Spearmint Valid.Error	Spearmint Best Value	TPE Valid.Error	TPE Best Value
branin (0.398)	200	0.427±0.019	0.400	0.398±0.000	**0.398**	0.510± 0.077	0.406
har6 (-3.32237)	200	-2.960±0.200	-3.165	-2.623±1.026	**-3.322**	-2.920±0.164	-3.115
Log.Regression	100	0.086±0.006	0.077	0.073±0.001	**0.070**	0.082±0.003	0.75
Log.Regression 5CV	100	0.081±0.003	**0.078**	0.083±0.001	0.080	0.089±0.007	0.081
LDA ongrid	50	1269.6±2.9	**1266.2**	1272.6±10.2	**1266.2**	1271.5±3.5	**1266.2**
SVM ongrid	100	0.241±0.001	**0.241**	0.246±0.009	**0.241**	0.241±0.000	**0.241**
HP-NNET convex	100	0.19±0.014 (8)	**0.175**	0.209±0.003 (6)	0.020	0.202±0.011 (9)	0.179
HP-NNET convex 5CV	50	0.208±0.010 (4)	0.198	0.230±0.015 (2)	0.215	0.210±0.012 (4)	**0.189**
HP-NNET MRBI	100	0.527±0.019 (8)	0.504	0.501±0.032 (5)	**0.467**	0.503±0.021	0.478
HP-NNET MRBI 5CV	50	0.512±0.032 (4)	0.478	0.497±0.020 (3)	**0.476**	0.518±0.000 (1)	0.518
HP-DBNET convex	200	0.159 (1)		0.138 (1)		0.135 (1)	
Auto-WEKA [2]	30h	**0.221**		N/A		0.255	

Taken from Eggensperger et al, 2013

Which method should I use?



Train In Data

THANK YOU

www.trainindata.com