Better than Deep Learning: Gradient Boosting Machines (GBM)

Szilárd Pafka, PhD
Chief Scientist, Epoch (USA)

½ Day Workshop, Budapest Data Forum Conference
June 2018

Edit profile

Szilard

@DataScienceLA

physics PhD, chief (data) scientist, meetup organizer, datascience.la, (visiting) professor, machine learning benchmarks 🇺🇸🇭🇺🇪🇺

Santa Monica, California linkedin.com/in/szilard

198 Following **2,067** Followers

At a Glance...

ML: sup.L: y = f(**x**) “learn” f from data (**y**, X)

training, testing/prediction, algos (LR,DT,NN…), optimization, overfitting, regularization...

GBM: ensemble of decision trees

GBM libs: R/Python

Experience with machine learning (supervised learning)

1 ○ heard about it

2 ○ basic concepts (training, overfitting, prediction/test), played with demos

3 ○ main concepts, understand basic algos (linear, trees, neural nets), wrote some code using ML libs

4 ○ trained and deployed ML models for real-world applications

Experience with R/Python

1 ○ none

2 ○ understand basic code, managed to run several demos

3 ○ can write scripts, use basic packages (esp. for data analysis/ML)

4 ○ proficient in writing R/Python code

Experience with machine learning (supervised learning)

1 ○ heard about it

2 ○ basic concepts (training, overfitting, prediction/test), played with demos

3 ○ main concepts, understand basic algos (linear, trees, neural nets), wrote some code using ML libs

4 ○ trained and deployed ML models for real-world applications **other than GBMs**

Experience with R/Python

1 ○ none

2 ○ understand basic code, managed to run several demos

3 ○ can write scripts, use basic packages (esp. for data analysis/ML)

4 ○ proficient in writing R/Python code

Experience with GBM libraries (not prerequisite, just asking)

	none	basic	intermmediate	advanced
R gbm package	○	○	○	○
Python sklearn GBM	○	○	○	○
xgboost (R/py)	○	○	○	○
h2o (R/py)	○	○	○	○
lightgbm (R/py)	○	○	○	○
Other	○	○	○	○

Disclaimer:

✔ I understand this is an **intermediate/advanced** workshop

Prerequisites:

basic ML concepts
R/Python experience

Schedule:

1. Intro talk (slides)
2. Demo main features (me running code)
3. Hands-on (you install/run code)

Laptop OS	R/Python	If Python	If Python, packages
Mac	Python	Python 2	pip install
Linux	R		
Windows	Python	Python 3	anaconda
Windows	R		
Linux	R		
Windows	Python	Python 3	pip install
Windows	R		
Windows	Python	Python 3	anaconda

Familiar with Linux comm	Access (via WiFi) to Linux	Access to server with GP
Yes	Yes	No
No	No	No
Yes	Yes	No
Yes	No	No
Yes	No	No
Yes	No	No
No	No	No
Yes	Yes	Yes

Experience with machine learning	Experience with R/Python
1- heard about it	3 - can write scripts, use basic packa
3- main concepts, understand basi	3 - can write scripts, use basic packa
2- basic concepts (training, overfitt	3 - can write scripts, use basic packa
4 - trained and deployed ML mode	4 - proficient in writing R/Python cod
3 - main concepts, understand bas	2 - understand basic code, manage
4 - trained and deployed ML mode	4 - proficient in writing R/Python cod
2- basic concepts (training, overfitt	2 - can write scripts, use basic packa
4 - trained and deployed ML mode	4 - proficient in writing R/Python cod

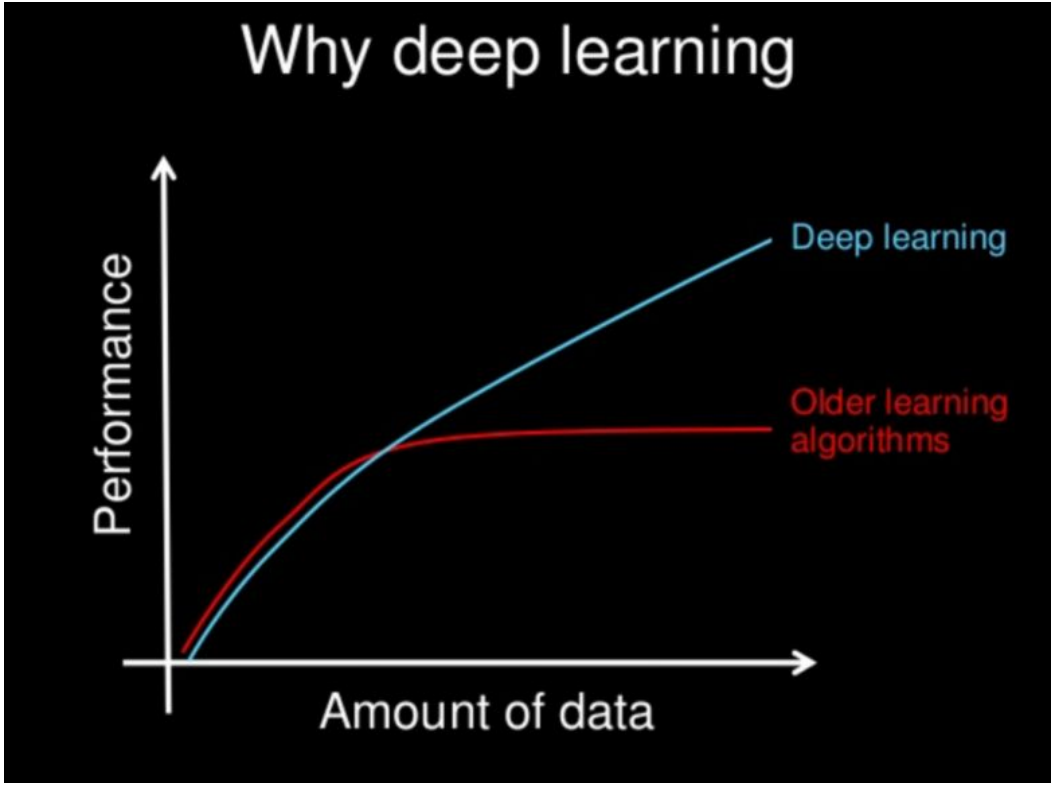
Experience with G	Experience with G	Experience with G	Experience with G	Experience with G	Experience with G
none	none	none	none	none	none
basic	none	basic	none	none	basic
none	none	none	none	none	none
basic					
none	none	basic	basic	none	none
none	none	none	none	none	none
basic	none	basic	basic	none	none
basic	advanced	basic	intermmediate	basic	

Student Intros / Goals

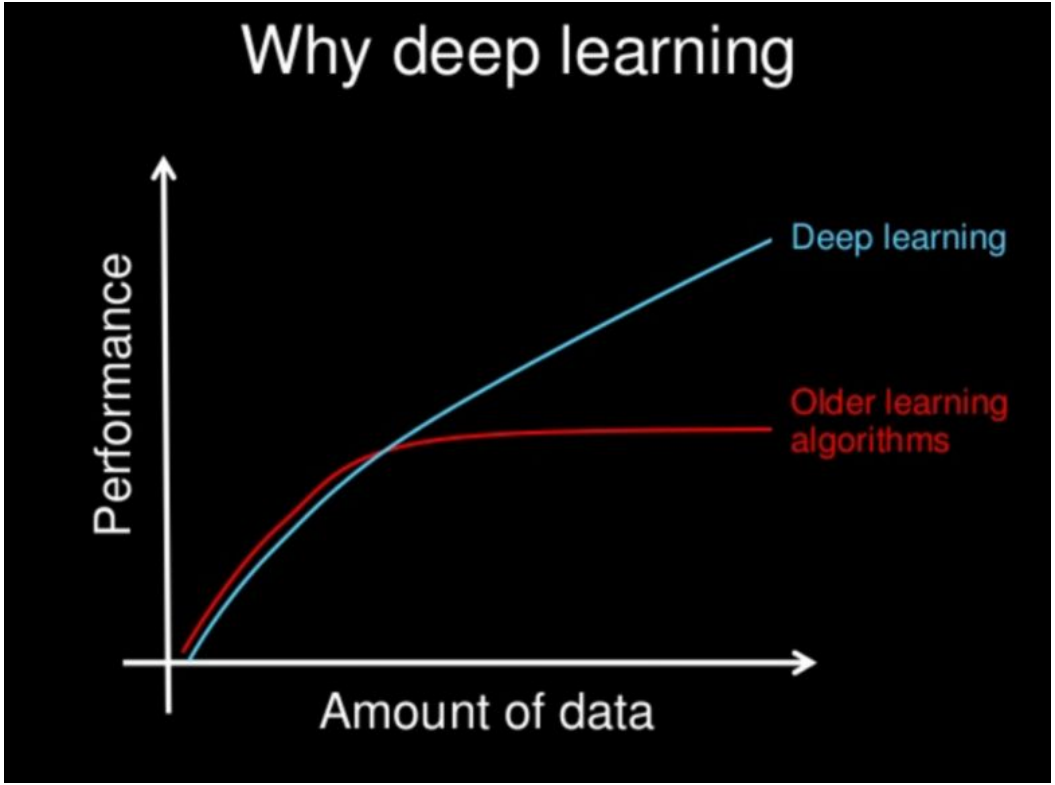
Disclaimer:

I am not representing my employer (Epoch) in this talk

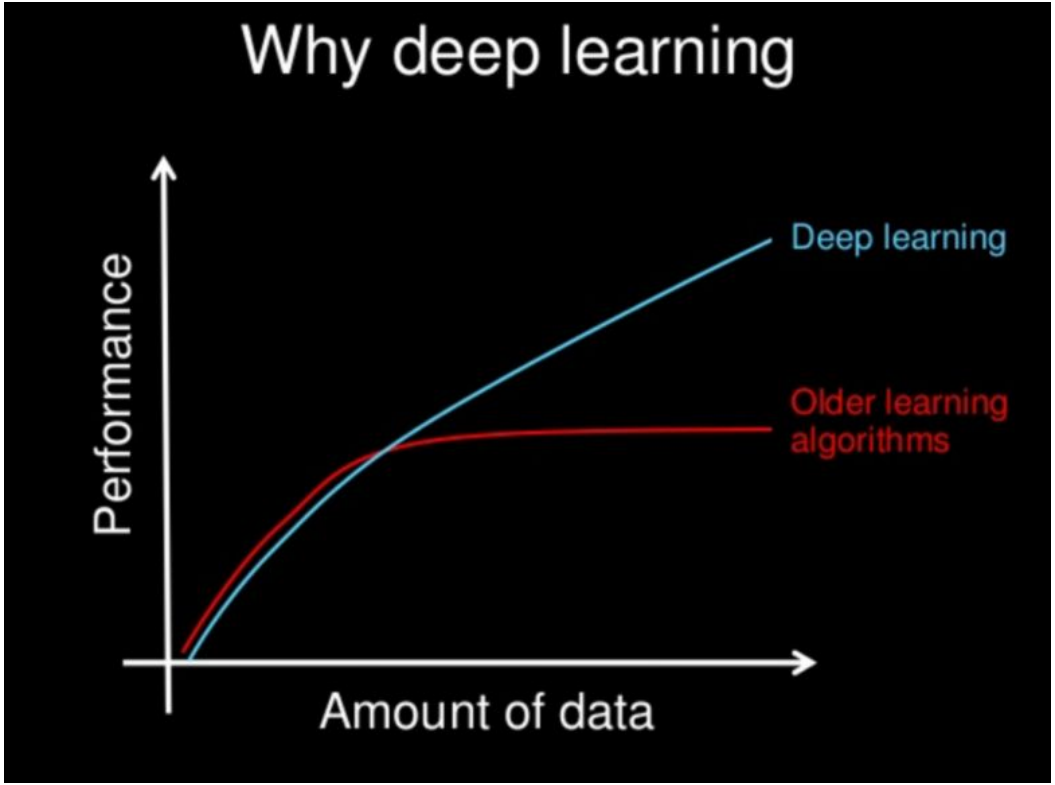
I cannot confirm nor deny if Epoch is using any of the methods, tools, results etc. mentioned in this talk



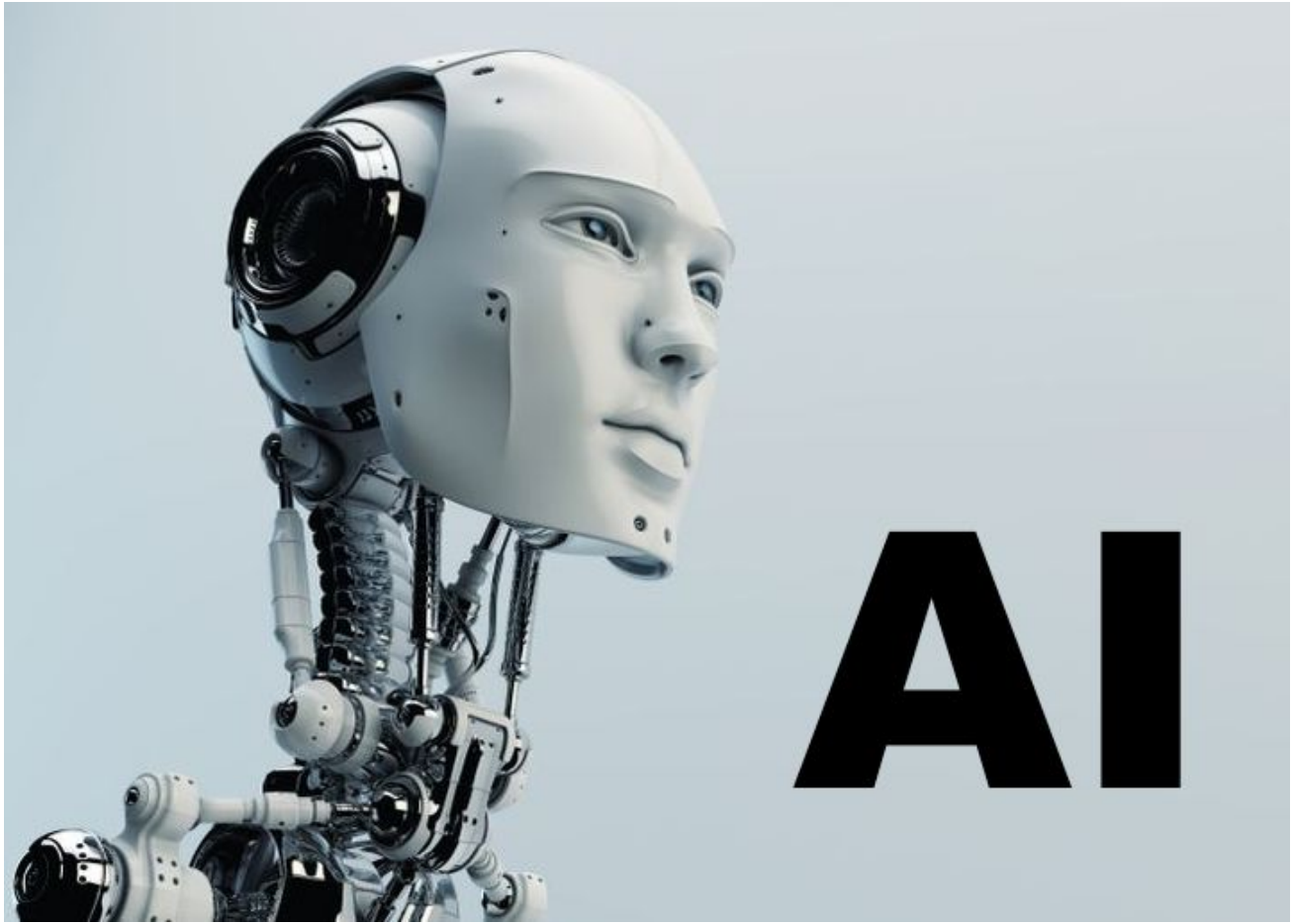
Source: Andrew Ng

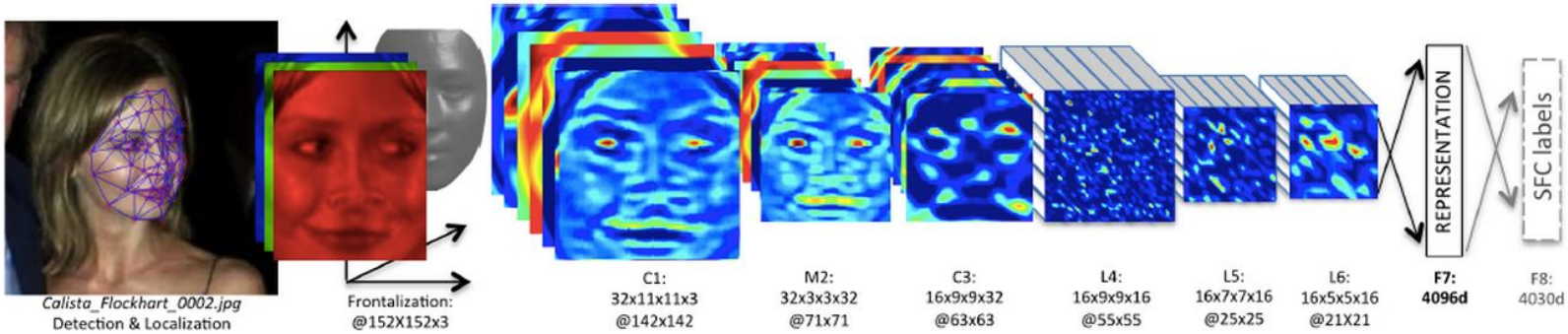


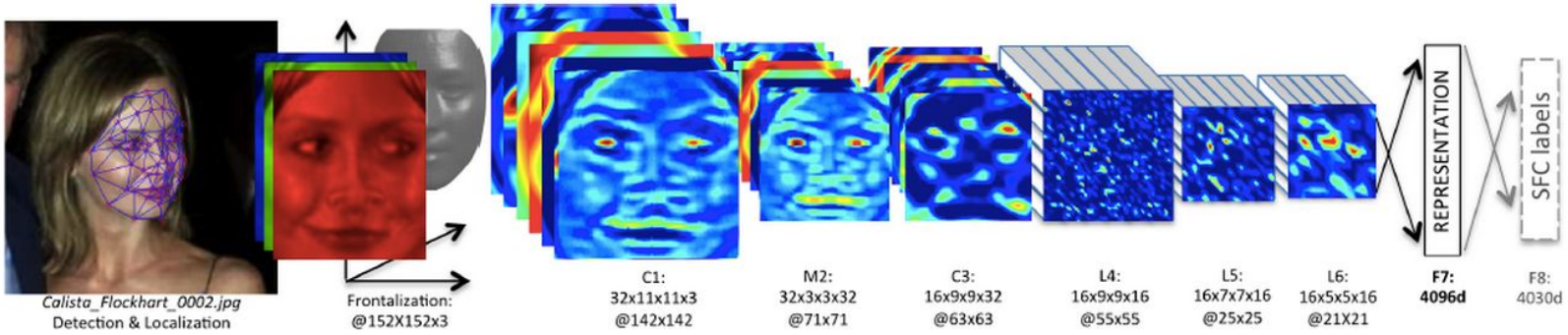
Source: Andrew Ng

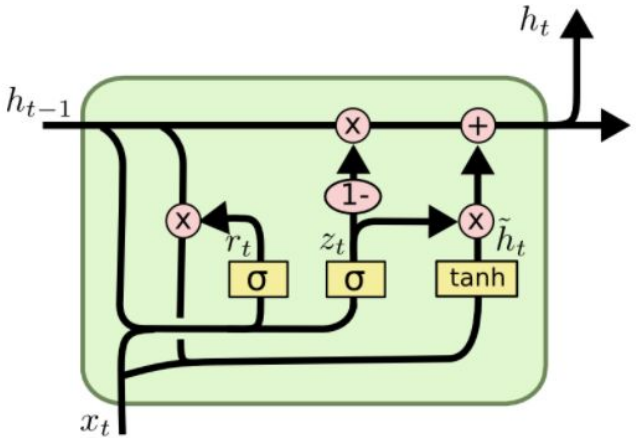


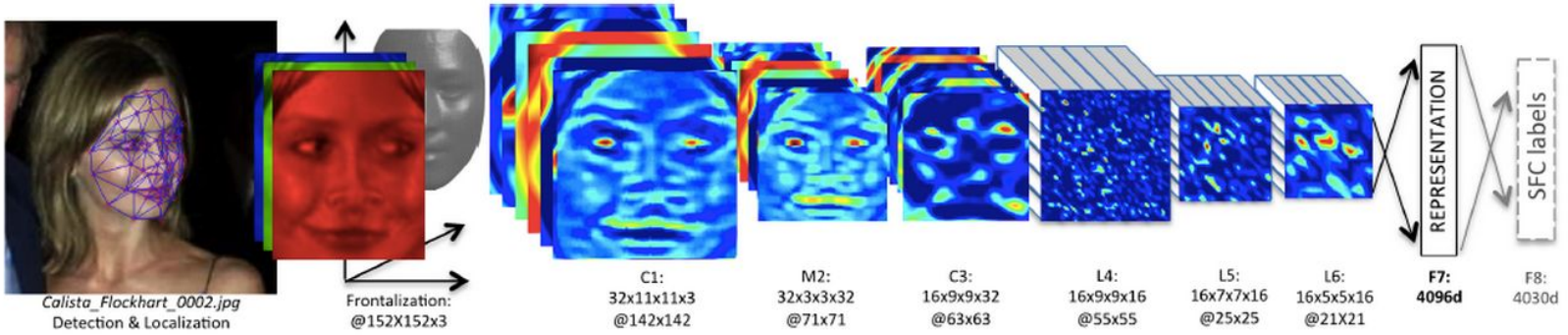
Source: Andrew Ng

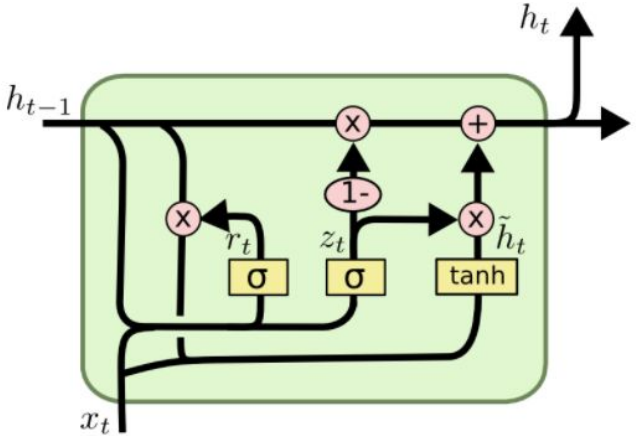

AI

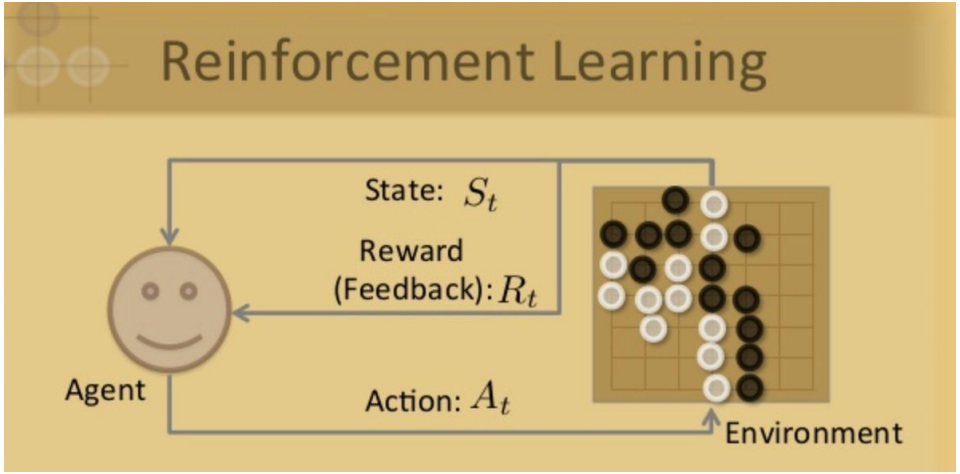

Calista_Flockhart_0002.jpg
Detection & Localization
Frontalization:
@152X152x3
C1:
32x11x11x3
@142x142
M2:
32x3x3x32
@71x71
C3:
16x9x9x32
@63x63
L4:
16x9x9x16
@55x55
L5:
16x7x7x16
@25x25
L6:
16x5x5x16
@21X21
F7:
4096d
REPRESENTATION
SFC labels
F8:
4030d


Calista_Flockhart_0002.jpg
Detection & Localization
Frontalization:
@152X152x3
C1:
32x11x11x3
@142x142
M2:
32x3x3x32
@71x71
C3:
16x9x9x32
@63x63
L4:
16x9x9x16
@55x55
L5:
16x7x7x16
@25x25
L6:
16x5x5x16
@21X21
F7:
4096d
REPRESENTATION
F8:
4030d
SFC labels


h_t
h_{t-1}
x
+
1-
x
x
r_t
z_t
$\tilde{h}_t$
σ
σ
tanh
x_t


Calista_Flockhart_0002.jpg
Detection & Localization
Frontalization:
@152X152x3
C1:
32x11x11x3
@142x142
M2:
32x3x3x32
@71x71
C3:
16x9x9x32
@63x63
L4:
16x9x9x16
@55x55
L5:
16x7x7x16
@25x25
L6:
16x5x5x16
@21X21
F7:
4096d
REPRESENTATION
F8:
4030d
SFC labels


h_t
h_{t-1}
x
+
1-
x
x
r_t
z_t
$\tilde{h}_t$
σ
σ
tanh
x_t


Reinforcement Learning
State: S_t
Reward
(Feedback): R_t
Agent
Action: A_t
Environment

Source: https://twitter.com/iamdevloper/


FRAUD
FRAUD


710
POOR
GOOD

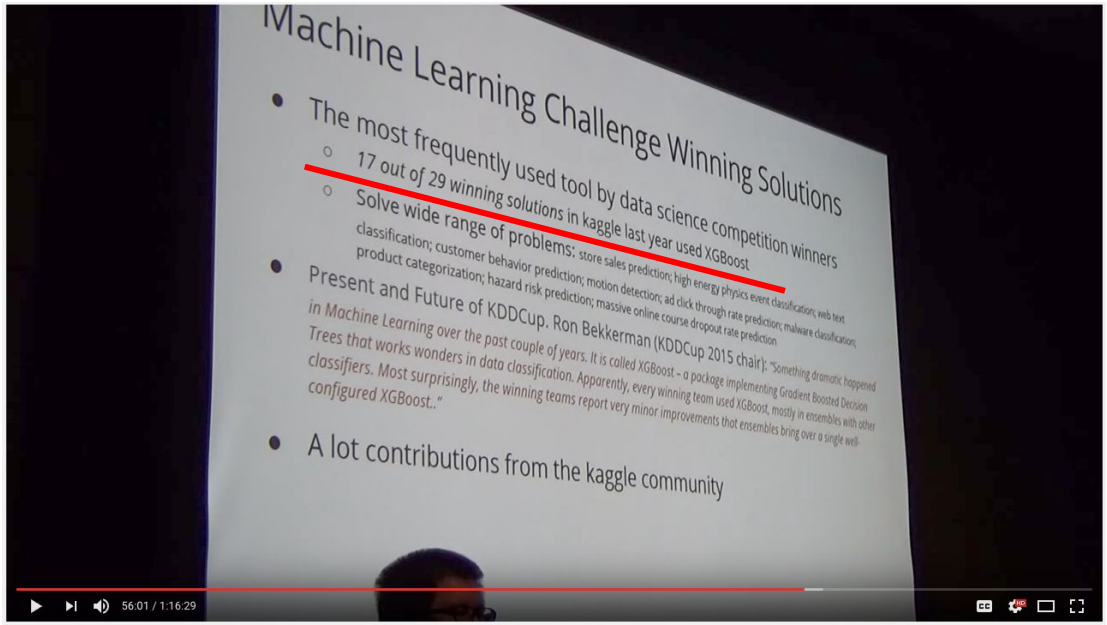

marketing
client
delivering
customers
profit
solution
concept
society
activities
creating
techniques
relationships
goal
goods
variety
communicating
satisfaction
strategy
build
brand
value
offer
business
management
development
organization
delivering
sell
term
long term
activity
sales
research
meaning
partners
process
company
adoption
objectives
sociology
activities
competitor
production
identify
profitable
knowing
buy
generate
perceived
return
support
distribution
advantage
maximize
contribute
focus
view
strong
target
wants
means

Params	AUC	Time (s)	Epochs
default: `activation = "Rectifier", hidden = c(200,200)`	73.1	270	1.8
`hidden = c(50,50,50,50), input_dropout_ratio = 0.2`	73.2	140	2.7
`hidden = c(50,50,50,50)`	73.2	110	1.9
`hidden = c(20,20)`	73.1	100	4.6
`hidden = c(20)`	73.1	120	6.7

...

`RectifierWithDropout, c(200,200,200,200), dropout=c(0.2,0.1,0.1,0)`	73.3	440	2.0
`ADADELTA rho = 0.95, epsilon = 1e-06`	71.1	240	1.7
`rho = 0.999, epsilon = 1e-08`	73.3	270	1.9
`adaptive = FALSE` default: `rate = 0.005, decay = 1, momentum = 0`	73.0	340	1.1
`rate = 0.001, momentum = 0.5 / 1e5 / 0.99`	73.2	410	0.7
`rate = 0.01, momentum = 0.5 / 1e5 / 0.99`	73.3	280	0.9
`rate = 0.01, rate_annealing = 1e-05, momentum = 0.5 / 1e5 / 0.99`	73.5	360	1
`rate = 0.01, rate_annealing = 1e-04, momentum = 0.5 / 1e5 / 0.99`	72.7	3700	8.7
`rate = 0.01, rate_annealing = 1e-05, momentum = 0.5 / 1e5 / 0.9`	73.4	350	0.9

kaggle



XGBoost A Scalable Tree Boosting System June 02, 2016

DataScience.LA

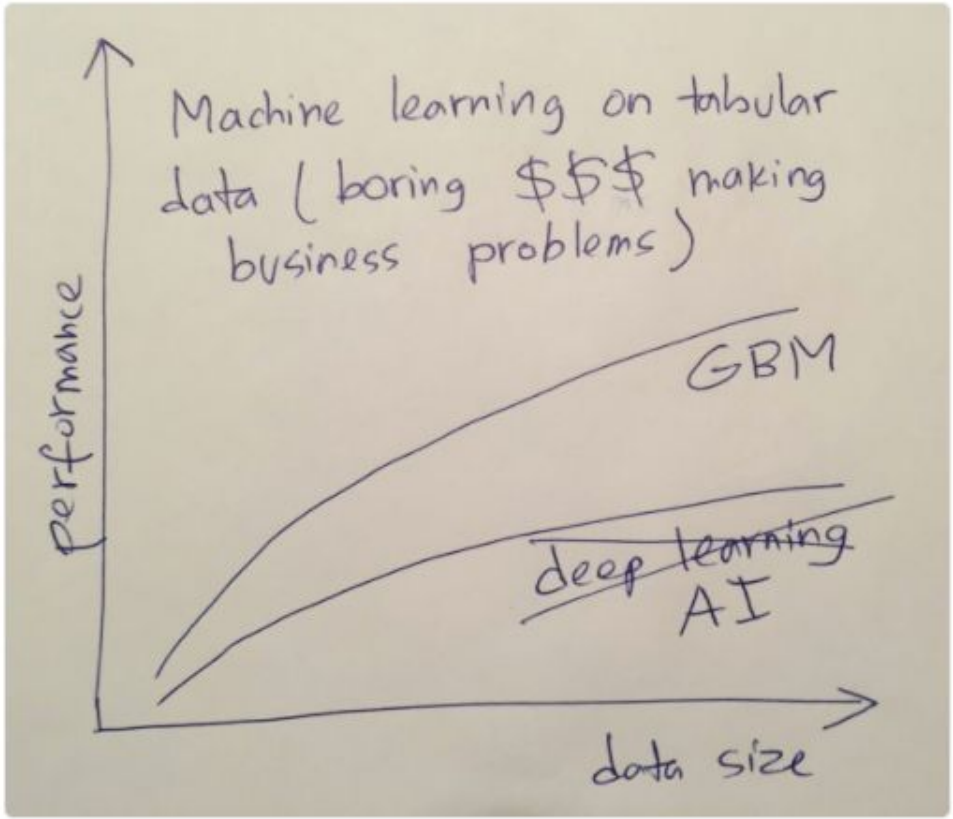
Subscribe 2.6K

5,632 views

Add to Share More 55 0

Szilard @DataScienceLA · 2 Nov 2016

Can anyone beat GBMs with deep learning (ahem, AI) on the airline dataset (or generally tabular/business data)? github.com/szilard/benchm…



2 8 16

MODEL	1ST	2ND
BST-DT	0.580	0.228
RF	0.390	0.525
BAG-DT	0.030	0.232
SVM	0.000	0.008
ANN	0.000	0.007
KNN	0.000	0.000
BST-STMP	0.000	0.000
DT	0.000	0.000
LOGREG	0.000	0.000
NB	0.000	0.000

AVG	1ST	2ND
RF	0.727	0.207
ANN	0.053	0.172
BSTDT	0.059	0.228
SVM	0.043	0.195
LR	0.089	0.132
BAGDT	0.002	0.012
KNN	0.023	0.045
BSTST	0.004	0.009
PRC	0	0
NB	0	0

An Empirical Comparison of Supervised Learning Algorithms

http://www.cs.cornell.edu/~alexn/papers/empirical.icml06.pdf

An Empirical Evaluation of Supervised Learning in High Dimensions

http://lowrank.net/nikos/pubs/empirical.pdf

MODEL	1ST	2ND
BST-DT	0.580	0.228
RF	0.390	0.525
BAG-DT	0.030	0.232
SVM	0.000	0.008
ANN	0.000	0.007
KNN	0.000	0.000
BST-STMP	0.000	0.000
DT	0.000	0.000
LOGREG	0.000	0.000
NB	0.000	0.000

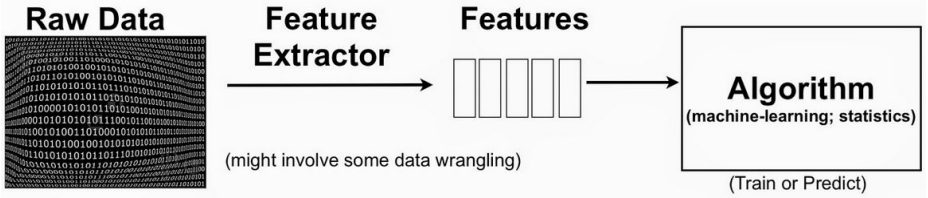
AVG	1ST	2ND
RF	0.727	0.207
ANN	0.053	0.172
BSTDT	0.059	0.228
SVM	0.043	0.195
LR	0.089	0.132
BAGDT	0.002	0.012
KNN	0.023	0.045
BSTST	0.004	0.009
PRC	0	0
NB	0	0

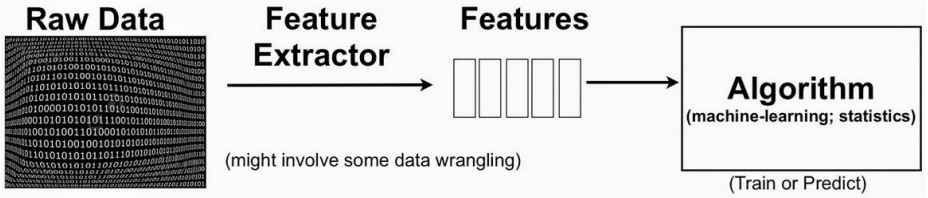
An Empirical Comparison of Supervised Learning Algorithms

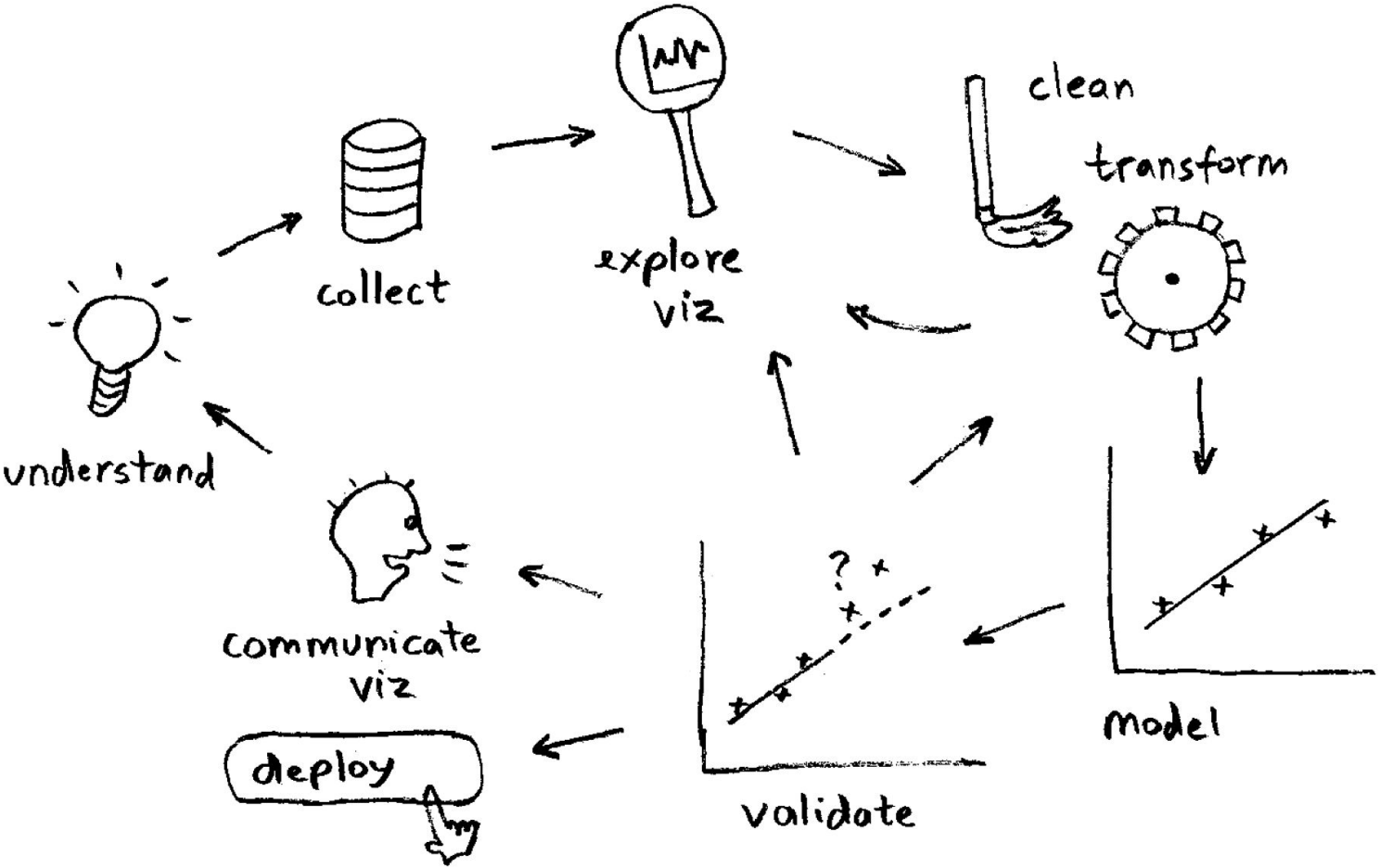
http://www.cs.cornell.edu/~alexn/papers/empirical.icml06.pdf

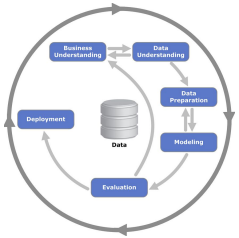
An Empirical Evaluation of Supervised Learning in High Dimensions

http://lowrank.net/nikos/pubs/empirical.pdf


Raw Data
Feature Extractor
Features
Algorithm
(machine-learning; statistics)
(might involve some data wrangling)
(Train or Predict)


Raw Data
Feature Extractor
Features
Algorithm
(machine-learning; statistics)
(might involve some data wrangling)
(Train or Predict)


clean
transform
explore
viz
collect
understand
communicate
viz
deploy
validate
model


Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment
Data

structured/tabular data: GBM (or RF)

very small data: LR

very large sparse data: LR with SGD (+L1/L2)

images/videos, speech: DL

structured/tabular data: GBM (or RF)

very small data: LR

very large sparse data: LR with SGD (+L1/L2)

images/videos, speech: DL

it depends

structured/tabular data: GBM (or RF)

very small data: LR

very large sparse data: LR with SGD (+L1/L2)

images/videos, speech: DL

it depends / try them all

structured/tabular data: GBM (or RF)

very small data: LR

very large sparse data: LR with SGD (+L1/L2)

images/videos, speech: DL

it depends / try them all / hyperparam tuning

structured/tabular data: GBM (or RF)

very small data: LR

very large sparse data: LR with SGD (+L1/L2)

images/videos, speech: DL

it depends / try them all / hyperparam tuning / ensembles

structured/tabular data: GBM (or RF)

very small data: LR

very large sparse data: LR with SGD (+L1/L2)

images/videos, speech: DL

it depends / try them all / hyperparam tuning / ensembles

feature engineering

structured/tabular data: GBM (or RF)

very small data: LR

very large sparse data: LR with SGD (+L1/L2)

images/videos, speech: DL

it depends / try them all / hyperparam tuning / ensembles

feature engineering / other goals e.g. interpretability

structured/tabular data: GBM (or RF)
very small data: LR
very large sparse data: LR with SGD (+L1/L2)
images/videos, speech: DL

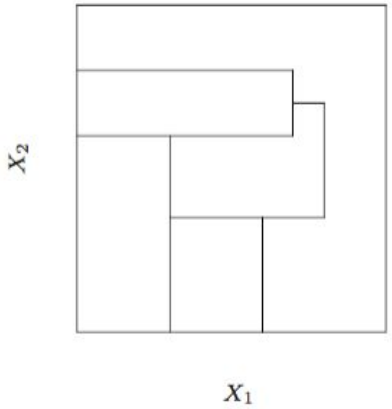
it depends / try them all / hyperparam tuning / ensembles
feature engineering / other goals e.g. interpretability

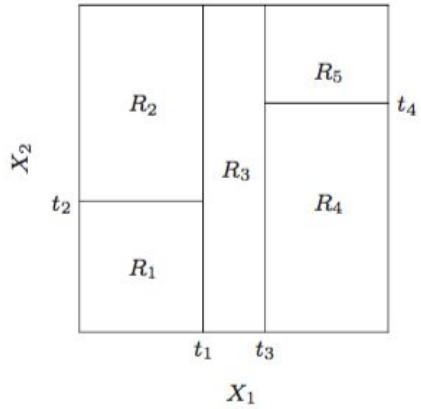
the title of this talk was misguided

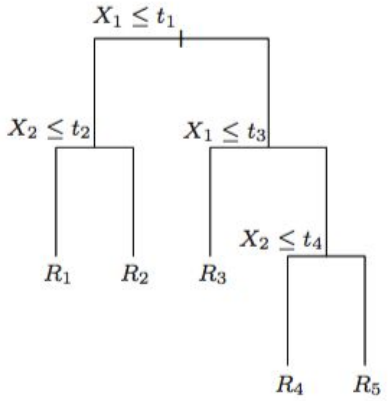
structured/tabular data: GBM (or RF)
very small data: LR
very large sparse data: LR with SGD (+L1/L2)
images/videos, speech: DL

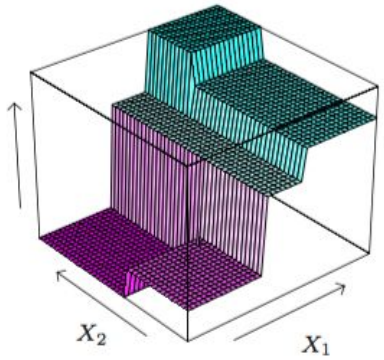
it depends / try them all / hyperparam tuning / ensembles
feature engineering / other goals e.g. interpretability

the title of this talk was misguided
but so is recently almost every use of the term AI

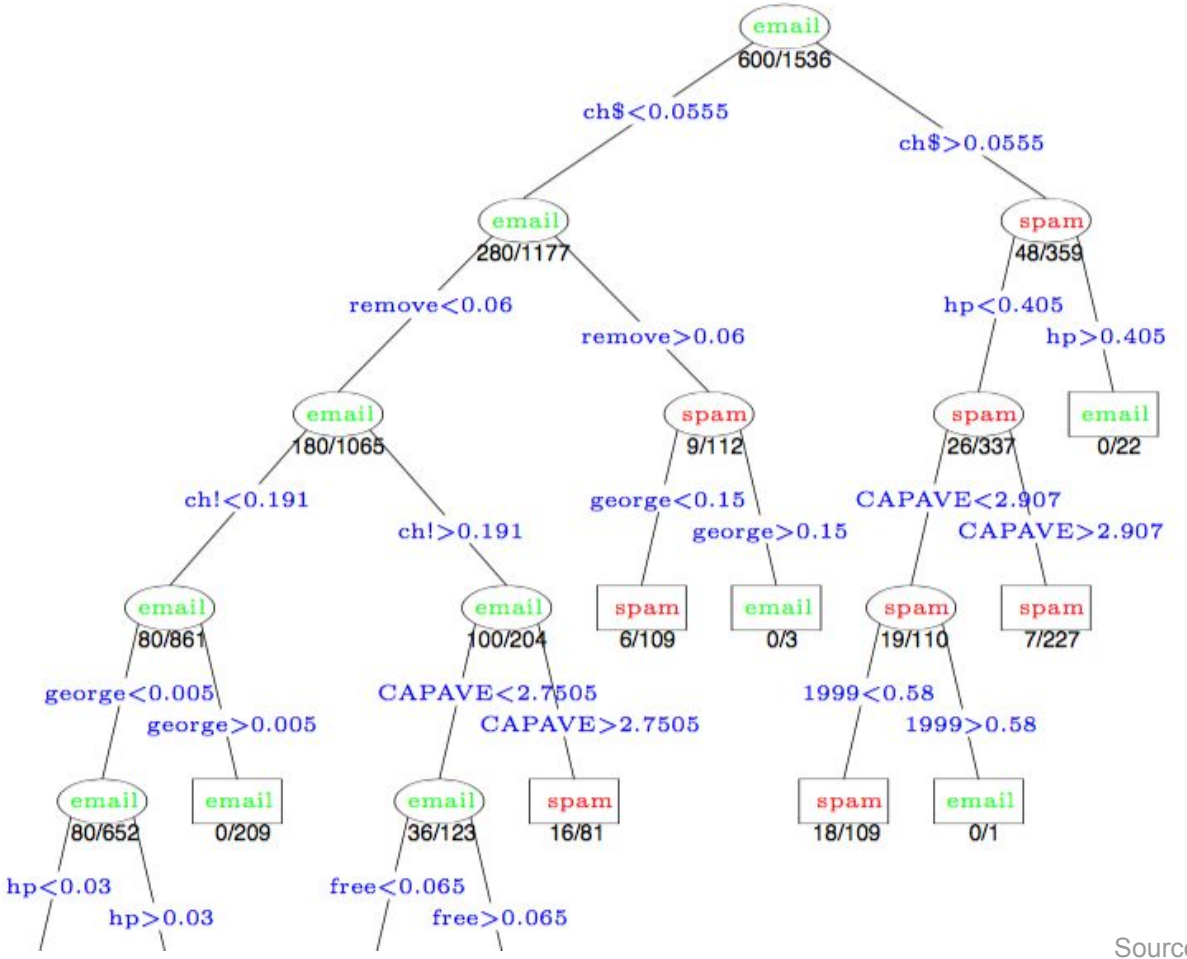








Source: Hastie etal, ESL 2ed



Source: Hastie etal, ESL 2ed

Algorithm 10.1 *AdaBoost.M1.*

1. Initialize the observation weights $w_i = 1/N,\ i = 1, 2, \ldots, N$.
2. For $m = 1$ to M:
   (a) Fit a classifier $G_m(x)$ to the training data using weights w_i.
   (b) Compute
   $$\text{err}_m = \frac{\sum_{i=1}^{N} w_i I(y_i \neq G_m(x_i))}{\sum_{i=1}^{N} w_i}.$$
   (c) Compute $\alpha_m = \log((1 - \text{err}_m)/\text{err}_m)$.
   (d) Set $w_i \leftarrow w_i \cdot \exp[\alpha_m \cdot I(y_i \neq G_m(x_i))],\ i = 1, 2, \ldots, N$.
3. Output $G(x) = \text{sign}\left[\sum_{m=1}^{M} \alpha_m G_m(x)\right]$.

Source: Hastie etal, ESL 2ed

Algorithm 10.3 *Gradient Tree Boosting Algorithm.*

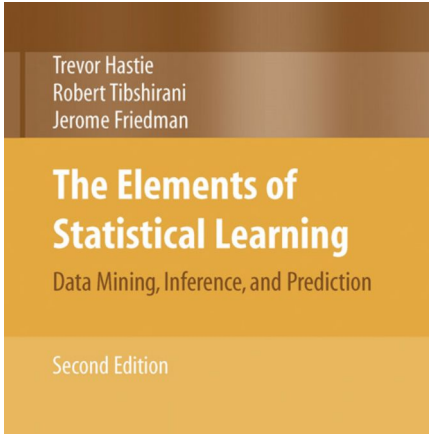
1. Initialize $f_0(x) = \arg\min_\gamma \sum_{i=1}^{N} L(y_i, \gamma)$.
2. For $m = 1$ to M:
   (a) For $i = 1, 2, \ldots, N$ compute

$$r_{im} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f=f_{m-1}}.$$

   (b) Fit a regression tree to the targets r_{im} giving terminal regions R_{jm}, $j = 1, 2, \ldots, J_m$.
   (c) For $j = 1, 2, \ldots, J_m$ compute

$$\gamma_{jm} = \arg\min_\gamma \sum_{x_i \in R_{jm}} L\left(y_i, f_{m-1}(x_i) + \gamma\right).$$

   (d) Update $f_m(x) = f_{m-1}(x) + \sum_{j=1}^{J_m} \gamma_{jm} I(x \in R_{jm})$.
3. Output $\hat{f}(x) = f_M(x)$.



Source: Hastie etal, ESL 2ed

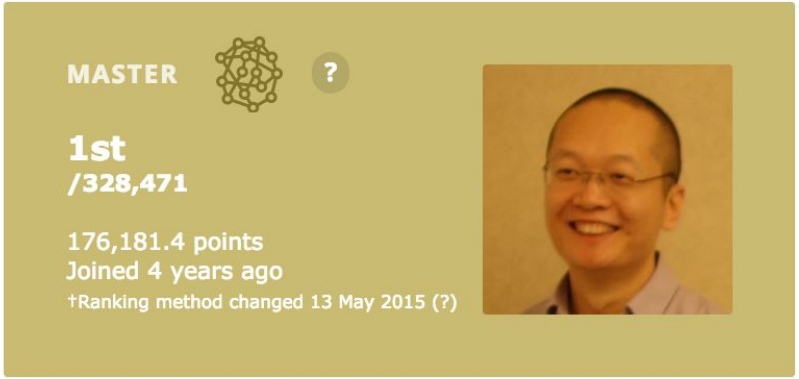
open source

- R packages
- Python scikit-learn
- Vowpal Wabbit
- H2O
- xgboost
- Spark MLlib
- a few others

I usually use other people's code [...] I can find open source code for what I want to do, and my time is much better spent doing research and feature engineering -- Owen Zhang

http://blog.kaggle.com/**2015/06/22**/profiling-top-kagglers-owen-zhang-currently-1-in-the-world/

kaggle



open source

100% FREE

Meetup

stackoverflow

open source

100% FREE

Meetup

stackoverflow

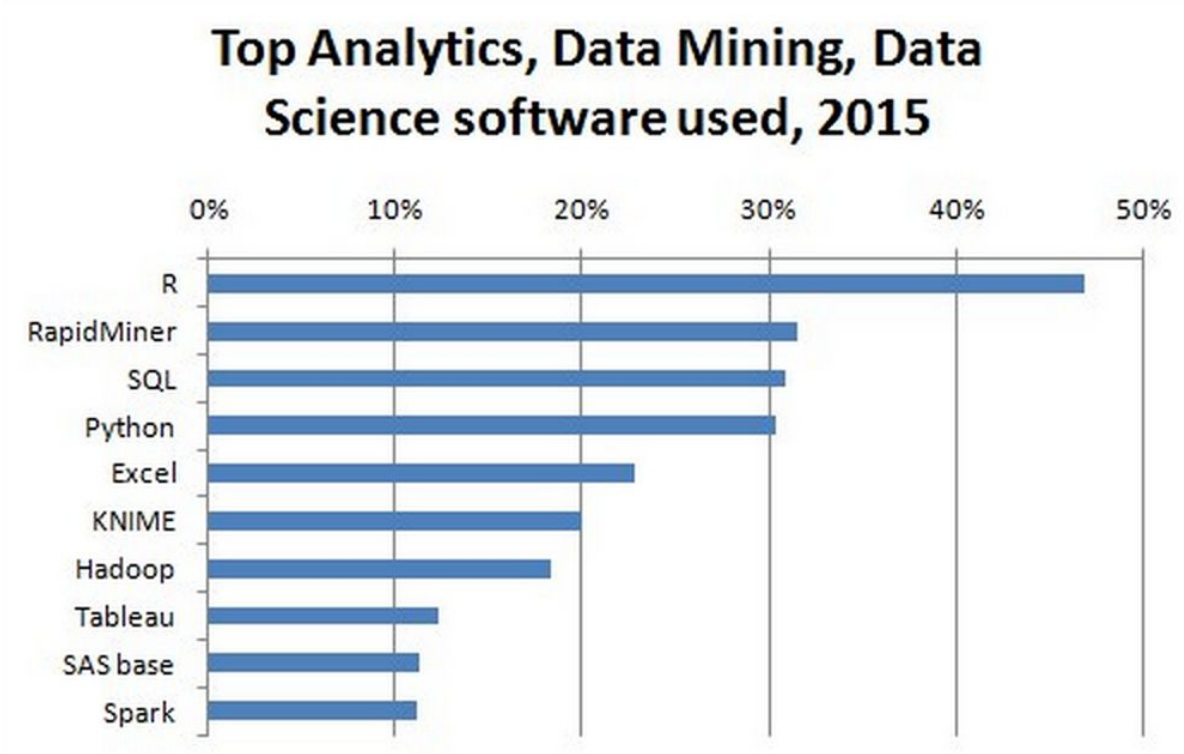
PyData

Tools Others Use (Survey)

LA Data Science/ML meetup, Apr 2014, 200 people

- Data munging: R 60%, Python 50%, SQL 40%, Hadoop (mostly Hive) 30%, Unix shell 20%, Excel 10% + Perl, Matlab, SAS, Impala, Pig, Shark...
- Visualization: R 40%,Python 30%, Tableau 10%, Javascript 10% + Matlab, Excel...
- Machine learning/modeling: R 30%, Python 30% + Vowpal Wabbit, Matlab, Mahout, SAS, SPSS...

KDnuggets


Top Analytics, Data Mining, Data Science software used, 2015
0%
10%
20%
30%
40%
50%
R
RapidMiner
SQL
Python
Excel
KNIME
Hadoop
Tableau
SAS base
Spark

KDnuggets

Poll (May 2010)

ata Mining / Analytic Tools Used Poll

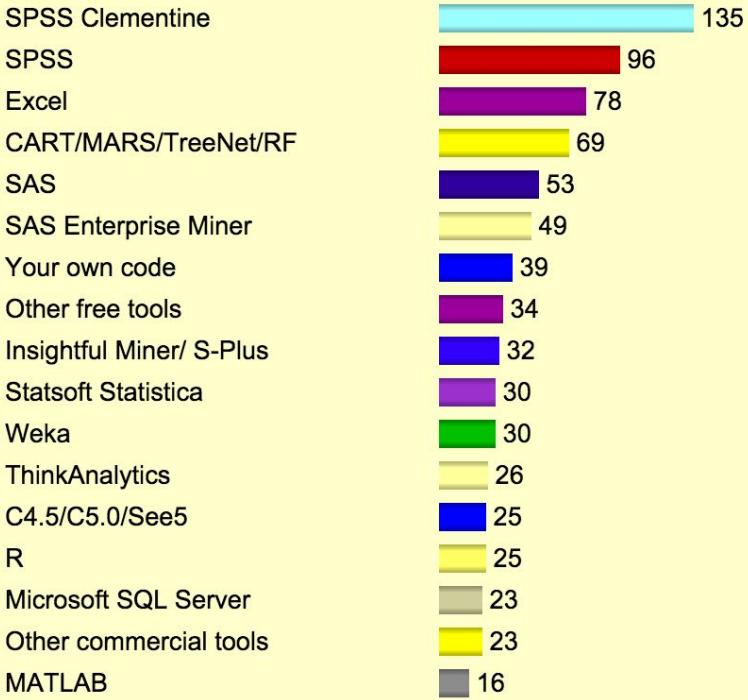
Which data mining/analytic tools you used in the past 12 months for a real project (not just evaluation) [912 voters]	
RapidMiner (345)	37.8%
R (272)	29.8%
Excel (222)	24.3%
KNIME (175)	19.2%
Your own code (168)	18.4%
Pentaho/Weka (131)	14.3%
SAS (110)	12.0%
MATLAB (84)	9.2%
IBM SPSS Statistics (72)	7.9%
Other free tools (67)	7.3%
IBM SPSS Modeler (former Clementine) (67)	7.3%
Microsoft SQL Server (63)	6.9%
Statsoft Statistica (57)	6.2%
Other commercial tools (56)	6.1%
SAS Enterprise Miner (50)	5.5%

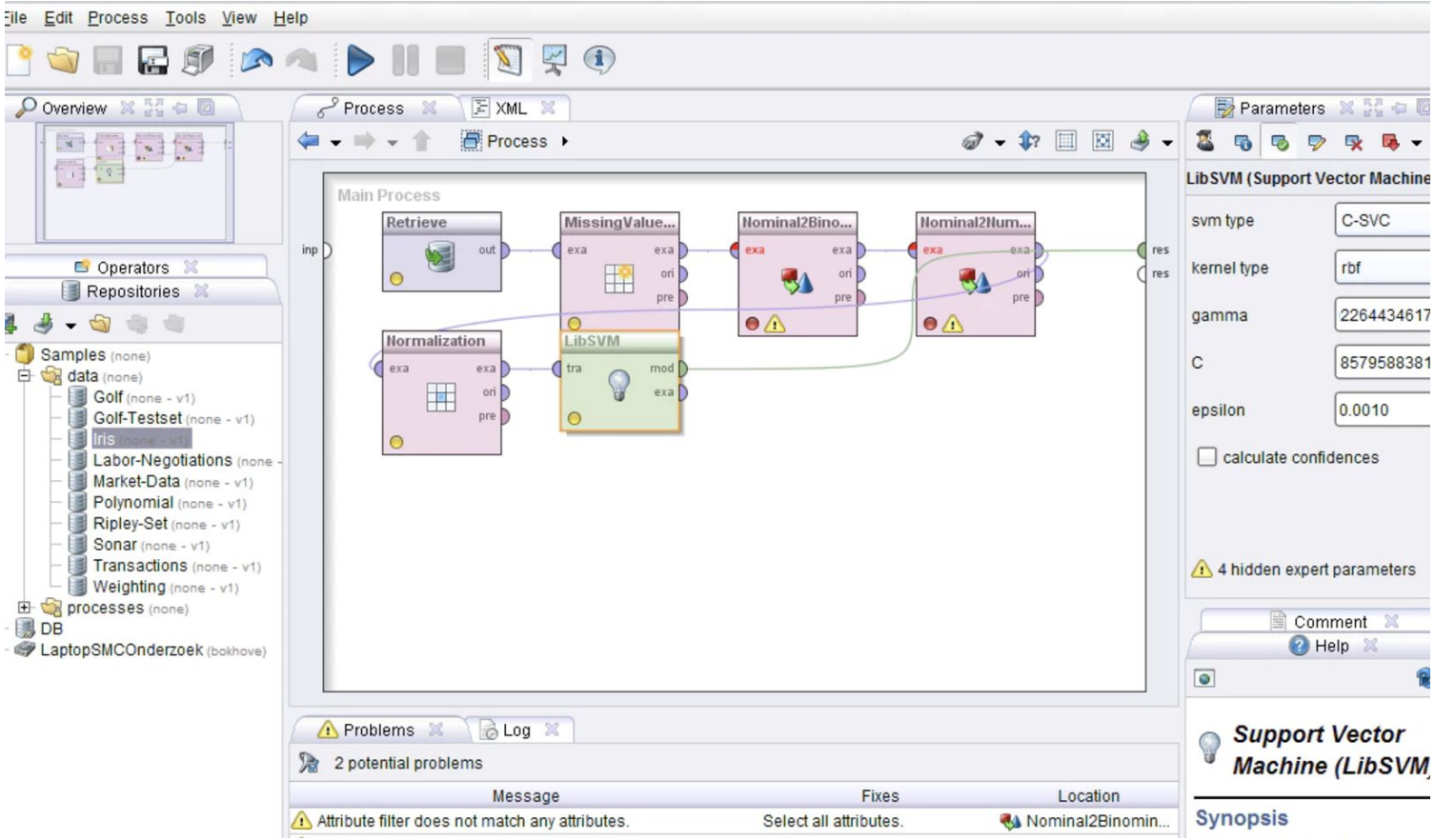
KDnuggets

: Data Mining Tools You Used in 2005 (May 2005)

Poll

Data mining/Analytic tools you used in 2005 [376 voters, 860 votes total]




File Edit Process Tools View Help
Overview
Operators
Repositories
Samples (none)
data (none)
Golf (none - v1)
Golf-Testset (none - v1)
Iris (none - v1)
Labor-Negotiations (none -
Market-Data (none - v1)
Polynomial (none - v1)
Ripley-Set (none - v1)
Sonar (none - v1)
Transactions (none - v1)
Weighting (none - v1)
processes (none)
DB
LaptopSMCOnderzoek (bokhove)
Process
XML
Process
Main Process
Retrieve
MissingValue...
Nominal2Bino...
Nominal2Num...
Normalization
LibSVM
inp
out
exa
ori
pre
tra
mod
res
Problems
Log
2 potential problems
Message
Fixes
Location
Attribute filter does not match any attributes.
Select all attributes.
Nominal2Binomin...
Parameters
LibSVM (Support Vector Machine
svm type
C-SVC
kernel type
rbf
gamma
2264434617
C
8579588381
epsilon
0.0010
calculate confidences
4 hidden expert parameters
Comment
Help
Support Vector Machine (LibSVM)
Synopsis

Alice Zhao
@adashofdata

Follow

The Great Debate: If You Can't Code, You Can't Be a Data Scientist #TeamCode wins! #Strataconf

open source

- R packages
- Python scikit-learn
- Vowpal Wabbit
- H2O
- xgboost
- Spark MLlib
- a few others

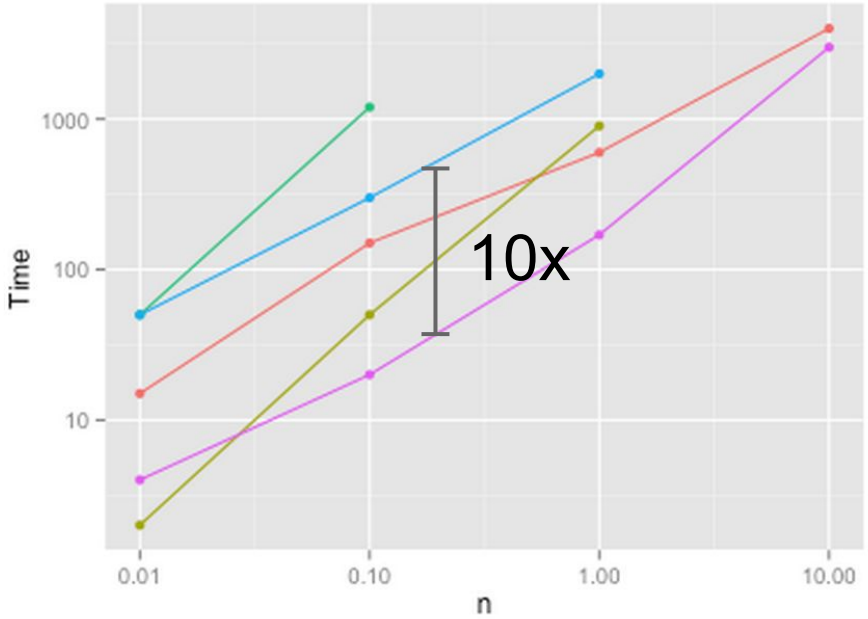
open source

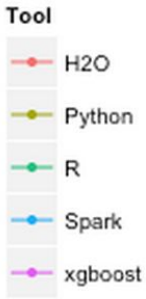
- R packages
- Python scikit-learn
- Vowpal Wabbit
- H2O
- xgboost
- Spark MLlib
- a few others

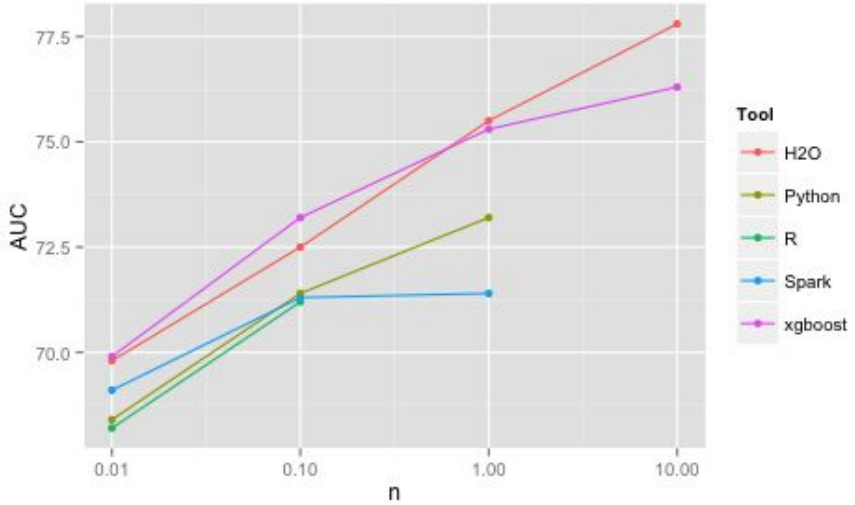
szilard / benchm-ml

★ Star 1,203

Simple/limited/incomplete benchmark


Time
1000
100
10
0.01
0.10
1.00
10.00
n
10x

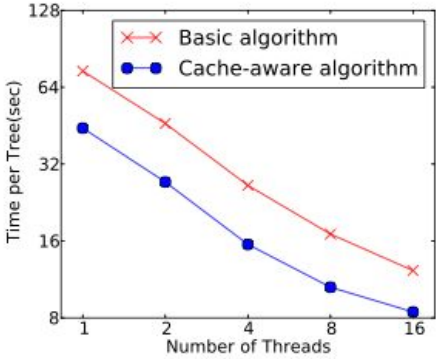

Tool
H2O
Python
R
Spark
xgboost


AUC
77.5
75.0
72.5
70.0
0.01
0.10
1.00
10.00
n
Tool
H2O
Python
R
Spark
xgboost

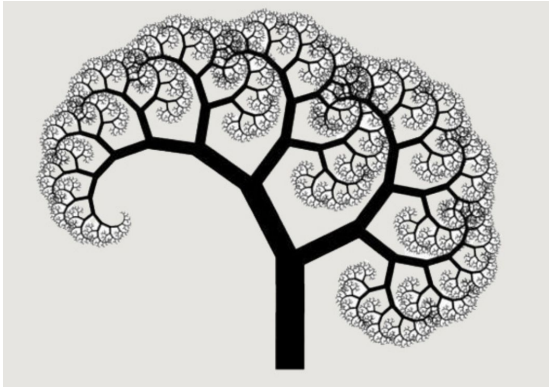
arXiv.org > cs > arXiv:1603.02754

Computer Science > Learning

XGBoost: A Scalable Tree Boosting System



H2O.ai

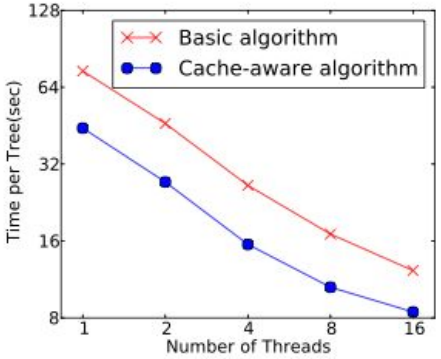


Microsoft / LightGBM

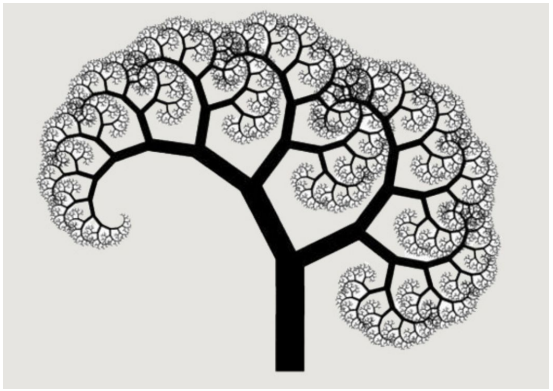
arXiv.org > cs > arXiv:1603.02754

Computer Science > Learning

XGBoost: A Scalable Tree Boosting System



H2O.ai



Microsoft / LightGBM

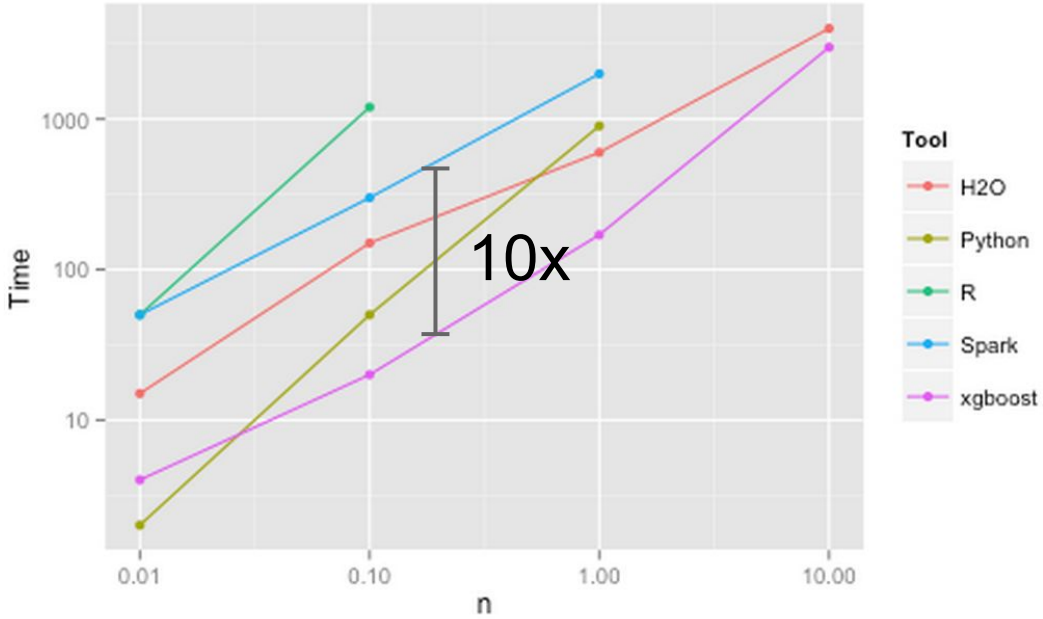
https://cran.r-project.org/web/pa

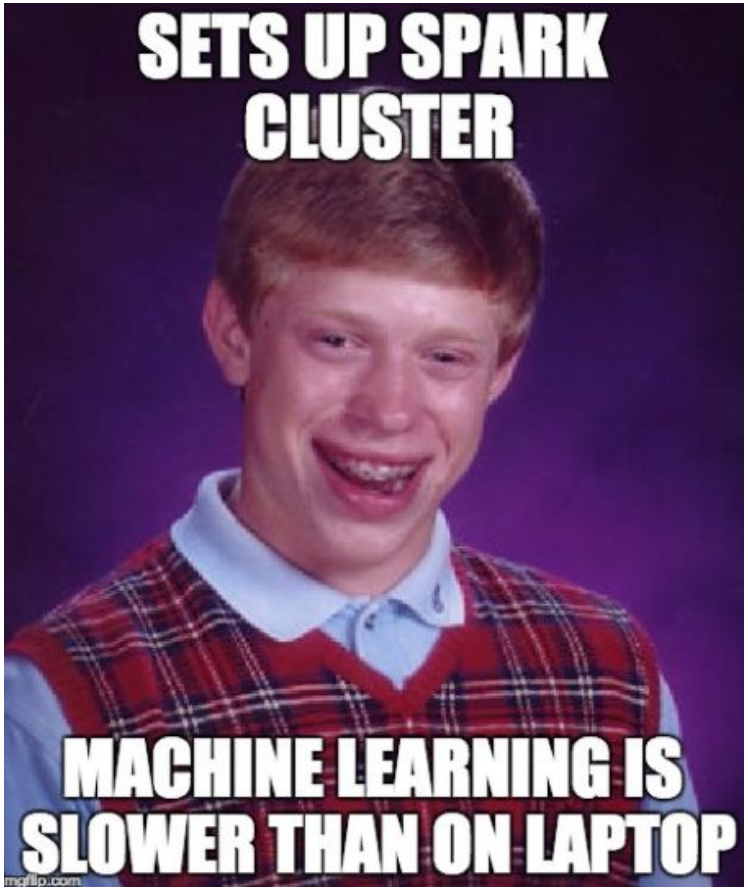
xgboost: Extreme Gradient Boosting

https://cran.r-project.org/

h2o: R Interface for H2O

PyData


1000
100
10
Time
0.01
0.10
1.00
10.00
n
10x
Tool
H2O
Python
R
Spark
xgboost


SETS UP SPARK CLUSTER
MACHINE LEARNING IS SLOWER THAN ON LAPTOP

APACHE Spark™ Lightning-fast

Download Libraries Documentati

Spark Release 2.0.0

DataFrame-based API is primary API

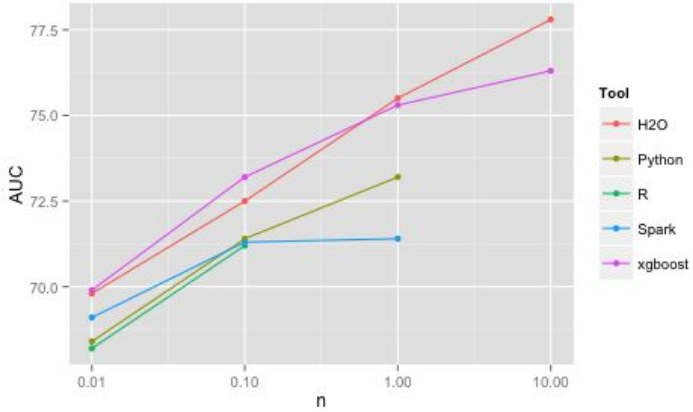
Spark random forest issues #19

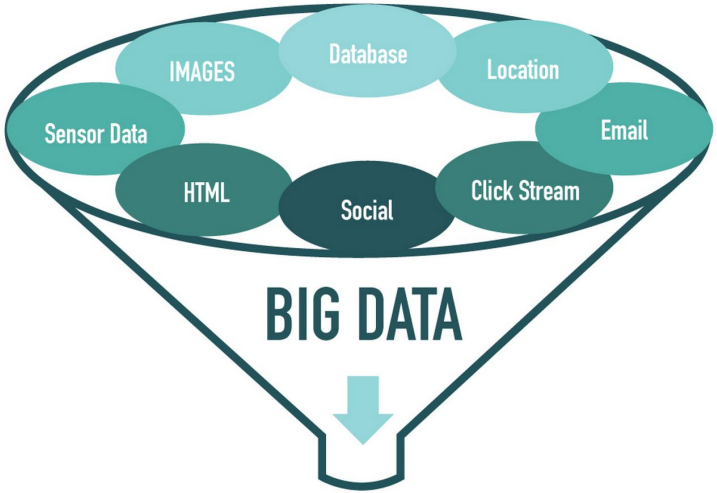
Open szilard opened this issue on Jul 23, 2015 · 15 comments

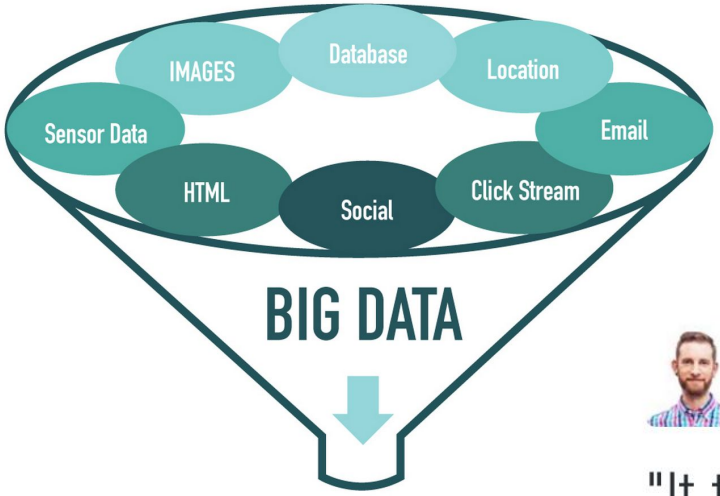
szilard commented 4 days ago

It is slower than before

```
MLlib 1.5 - 250 sec
ML    2.0 - 400 sec
```



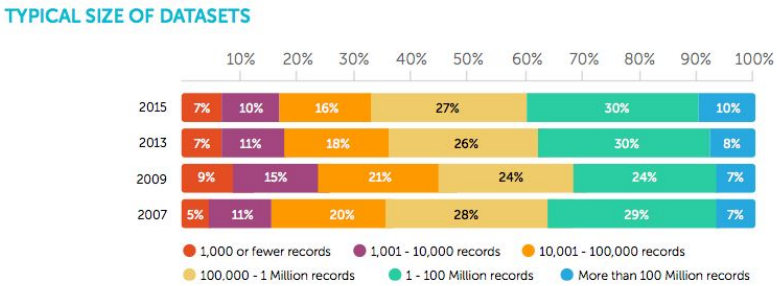

IMAGES
Database
Location
Sensor Data
Email
HTML
Social
Click Stream
BIG DATA

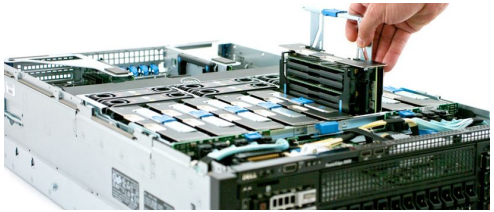


Hadley Wickham
@hadleywickham

Following

"It takes a big man to admit his data is small" — @jcheng





Kingston Technology Value RAM 128GB Kit (4x32GB) 2133MHz DDR4 ECC Reg CL15 (KVR21R15D4K4/128)

by Kingston Technology

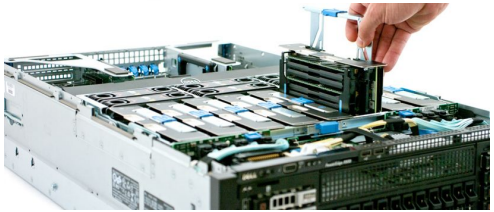
Be the first to review this item

Was: ~~$743.99~~

Price: $743.96 & **FREE Shipping**. Details

amazon web services

Model	vCPU	Mem (GiB)
r3.8xlarge	32	244
x1.32xlarge	128	1,952



Kingston Technology Value RAM 128GB Kit (4x32GB) 2133MHz DDR4 ECC Reg CL15 (KVR21R15D4K4/128)

by Kingston Technology

Be the first to review this item

Was: ~~$743.99~~

Price: $743.96 & **FREE Shipping**. Details

amazon web services

Model	vCPU	Mem (GiB)
r3.8xlarge	32	244
x1.32xlarge	128	1,952

Szilard @DataScienceLA · 18 Nov 2015

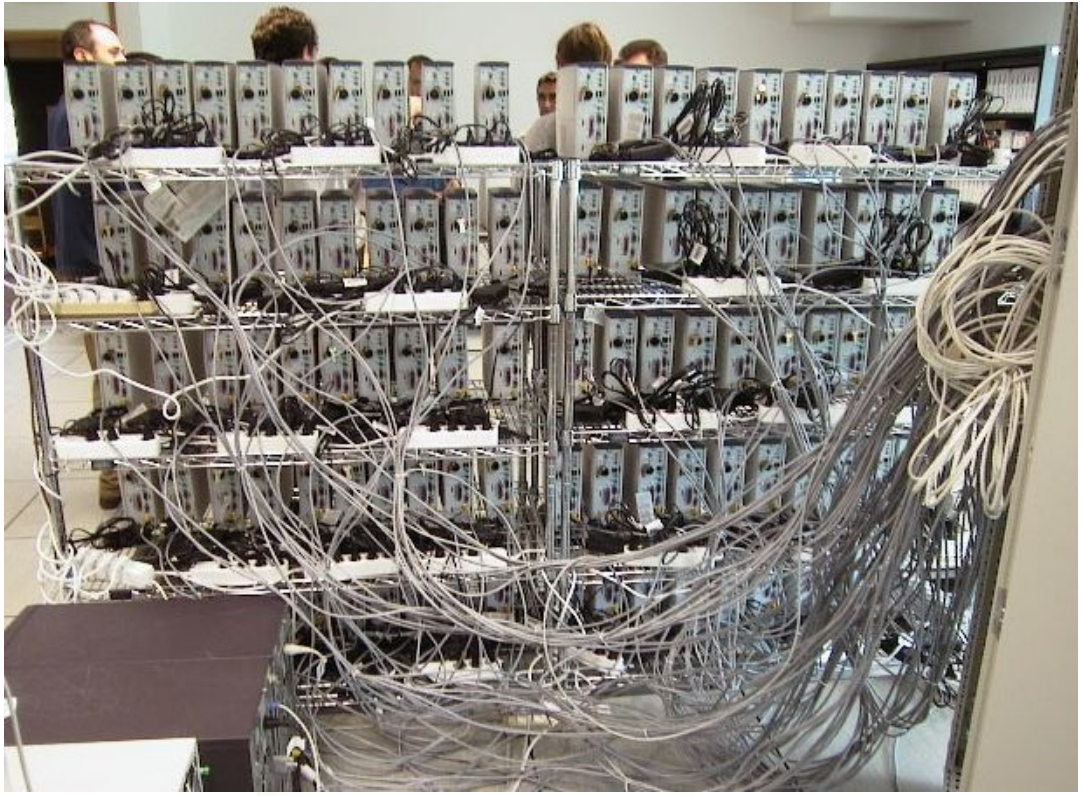
Big RAM is eating #bigdata: datasets for analytics grew 20% /yr (last decade @kdnuggets), RAM EC2 grew 50% /yr

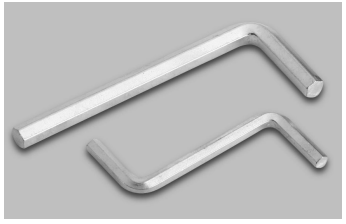
Gary Bernhardt
@garybernhardt

Consulting service: you bring your big data problems to me, I say "your data set fits in RAM", you pay me $10,000 for saving you $500,000.

RETWEETS	LIKES
997	960

3:03 PM - 19 May 2015






SYSTEM FAILURE

Szilard @DataScienceLA · Aug 3

I wish my #machinelearning worked... ("both" is not a choice 😀) #bigdata #datascience #rstats #pydata cc @h2o @databricks @cloudera @kaggle

10% on 10x bigger data

70% 10x faster

20% I don't care about either

104 votes • Final results

szilard / GBM-perf

```
git clone https://github.com/szilard/GB
cd GBM-perf/cpu
sudo docker build -t gbmperf_cpu .
sudo docker run --rm gbmperf_cpu
```

r4.8xlarge (32 cores) with software as of 2018-01-27:

Tool	Time[s] 1M	Time[s] 10M	AUC 1M	AUC 10M
h2o	21	90	0.762	0.776
xgboost	16	150	0.749	0.755
lightgbm	**6**	47	0.766	0.774

szilard / GBM-perf

```
git clone https://github.com/szilard/GB
cd GBM-perf/cpu
sudo docker build -t gbmperf_cpu .
sudo docker run --rm gbmperf_cpu
```

r4.8xlarge (32 cores) with software as of 2018-01-27:

Tool	Time[s] 1M	Time[s] 10M	AUC 1M	AUC 10M
h2o	21	90	0.762	0.776
xgboost	16	150	0.749	0.755
lightgbm	6	47	0.766	0.774

p3.2xlarge (1 GPU, Tesla V100) with software as of 2018-01-29:

Tool	Time[s] 1M	Time[s] 10M	AUC 1M	AUC 10M
h2o xgboost	18	error	0.712	error
xgboost	8	25	0.748	0.756
lightgbm	20	75	0.766	0.774

```r
X <- Matrix::sparse.model.matrix(dep_delayed_15min ~ . - 1, data = d)
## 1-hot encoding + sparse
## needs to be done *together* (train+test) for alignment (otherwise error for new cats at scoring)
## still problem in live scoring scenarios
X[1:10,1:10]
X_train <- X[idx,]
X_test <- X[-idx,]

dxgb_train <- xgb.DMatrix(data = X_train, label = ifelse(d_train$dep_delayed_15min=='Y',1,0))
## special optimized data structure


## TRAIN
system.time({
  md <- xgb.train(data = dxgb_train, objective = "binary:logistic",
          nround = 100, max_depth = 10, eta = 0.1)
})

?xgb.train


## SCORE
phat <- predict(md, newdata = X_test)
```

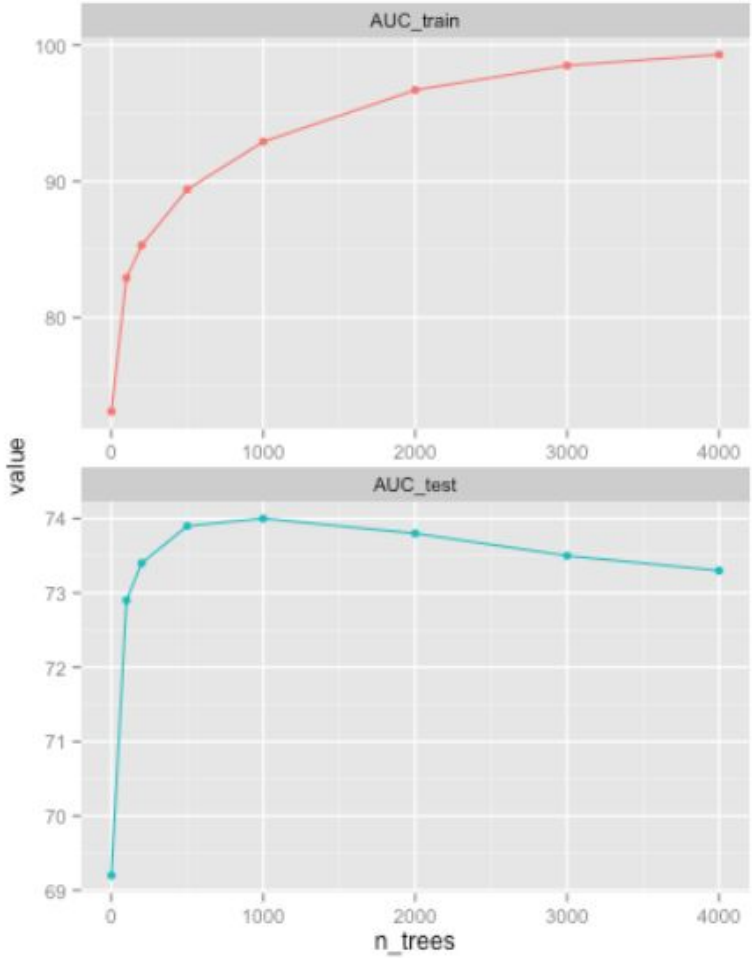

```
## exporting model for scoring

h2o.download_mojo(md_rf, path = "./h2o")
```

```
## building prediction service

# (need jetty-runner.jar ROOT.war from Steam)
java -jar jetty-runner.jar ROOT.war

curl -X POST --form mojo=@h2o_RF.zip --form jar=@h2o-genmodel.jar \
             localhost:8080/makewar > h2o_RF_MOJO.war
```

```
## run prediction service

java -jar jetty-runner.jar --port 20000 h2o_RF_MOJO.war
```

```
## score via REST API

time curl "http://localhost:20000/predict?Month=c-8&DayofMonth=c-21&DayOfWeek=c-7&DepTime=1934
# (fast scoring needs JVM to warm up with a few requests)
```

```
h2o.gbm(x, y, training_frame, model_id, checkpoint, ignore_const_cols = TRUE,
  distribution = c("AUTO", "gaussian", "bernoulli", "multinomial", "poisson",
  "gamma", "tweedie", "laplace", "quantile", "huber"), quantile_alpha = 0.5,
  tweedie_power = 1.5, huber_alpha = 0.9, ntrees = 50, max_depth = 5,
  min_rows = 10, learn_rate = 0.1, learn_rate_annealing = 1,
  sample_rate = 1, sample_rate_per_class, col_sample_rate = 1,
  col_sample_rate_change_per_level = 1, col_sample_rate_per_tree = 1,
  nbins = 20, nbins_top_level = 1024, nbins_cats = 1024,
  validation_frame = NULL, balance_classes = FALSE, class_sampling_factors,
  max_after_balance_size = 5, seed, build_tree_one_node = FALSE,
  nfolds = 0, fold_column = NULL, fold_assignment = c("AUTO", "Random",
  "Modulo", "Stratified"), keep_cross_validation_predictions = FALSE,
  keep_cross_validation_fold_assignment = FALSE,
  score_each_iteration = FALSE, score_tree_interval = 0,
  stopping_rounds = 0, stopping_metric = c("AUTO", "deviance", "logloss",
  "MSE", "AUC", "misclassification", "mean_per_class_error"),
  stopping_tolerance = 0.001, max_runtime_secs = 0, offset_column = NULL,
  weights_column = NULL, min_split_improvement = 1e-05,
  histogram_type = c("AUTO", "UniformAdaptive", "Random", "QuantilesGlobal",
  "RoundRobin"), max_abs_leafnode_pred, pred_noise_bandwidth = 0,
  categorical_encoding = c("AUTO", "Enum", "OneHotInternal", "OneHotExplicit",
  "Binary", "Eigen"))
```


AUC_train
100
90
80
AUC_test
74
73
72
71
70
69
value
0
1000
2000
3000
4000
n_trees

blog.h2o.ai/2016/06/h2o-gbm-tuning-tutorial-for-r/

Arno Candel in GBM, R, Technical, Tutorials | June 16, 2016

H2O GBM Tuning Tutorial for R

In this tutorial, we show how to build a well-tuned H2O GBM model for a supervised classification task. and use a small dataset to allow you to reproduce these results in a few minutes on a laptop. This script ca dreds of GBs large and H2O clusters with dozens of compute nodes.

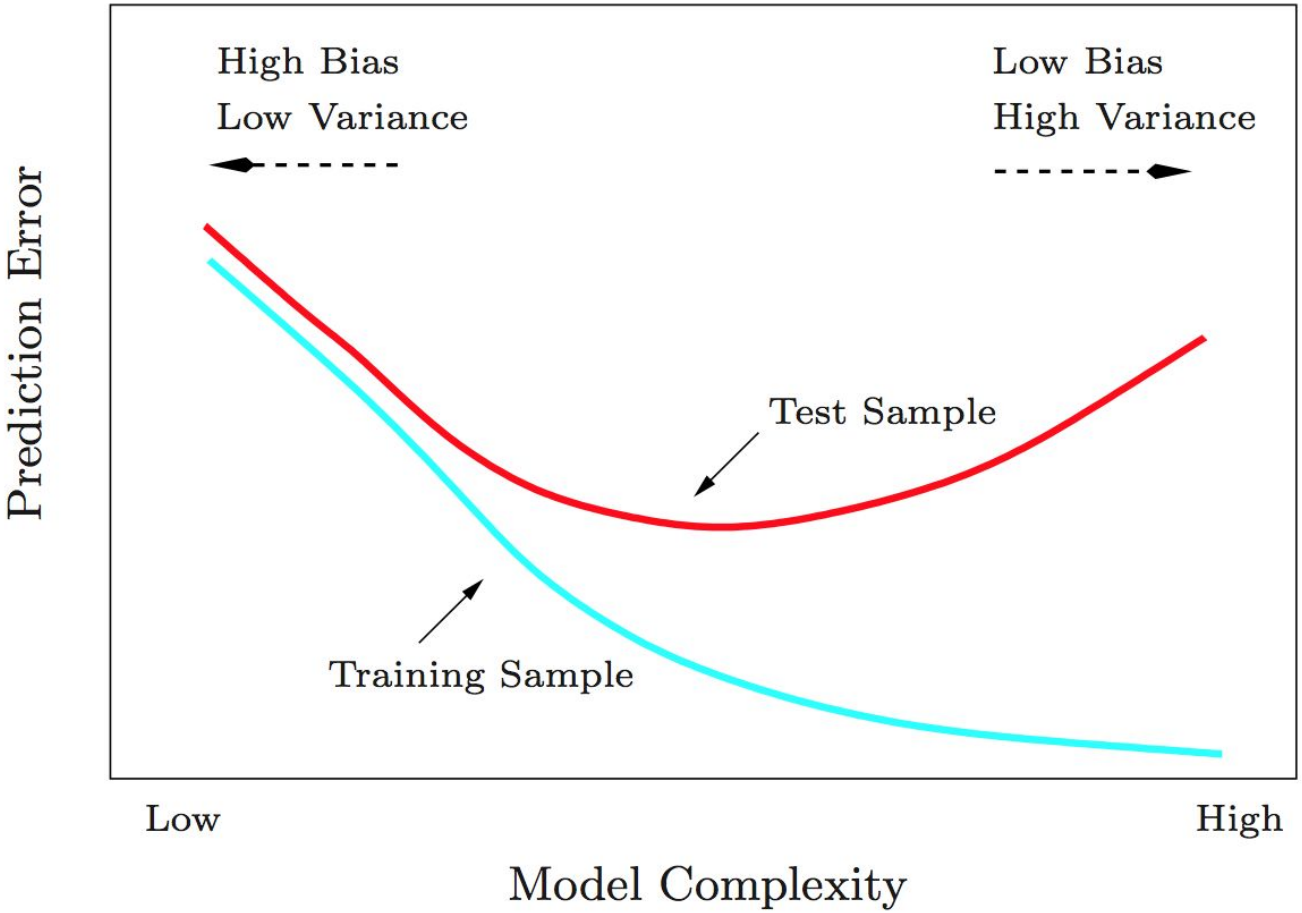
machinelearningmastery.com/configure-gradient-boosting-algorithm/

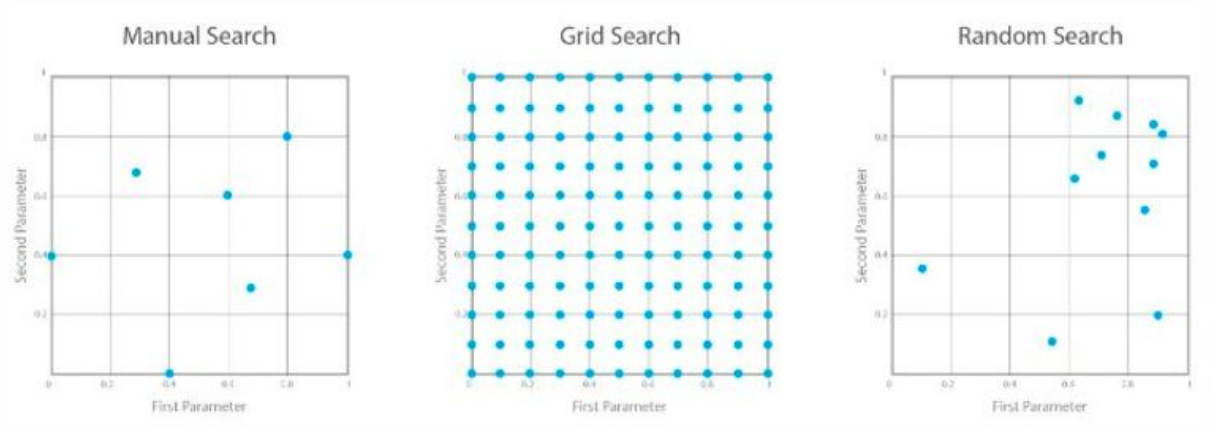
Start Here

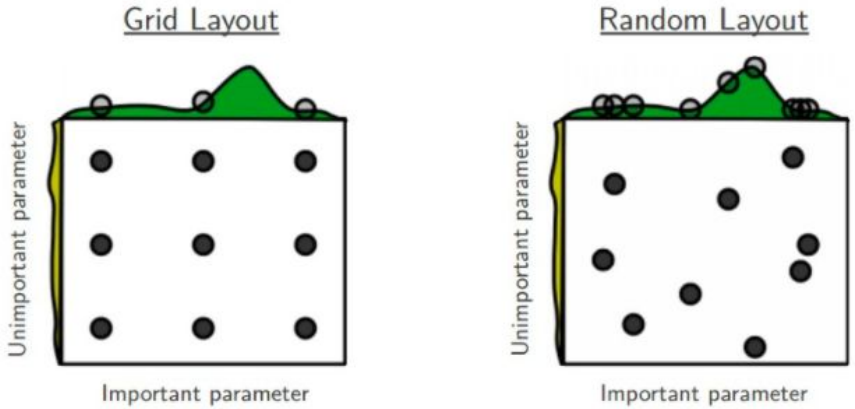
Search...

How to Configure the Gradient Boosting Algorithm

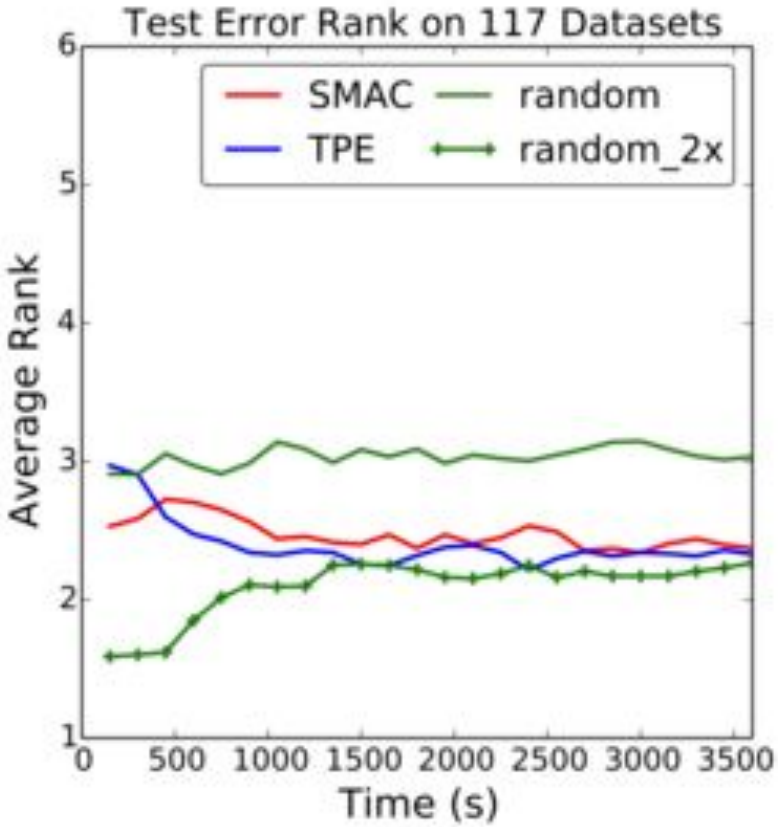
by **Jason Brownlee** on September 12, 2016 in **XGBoost**


High Bias
Low Variance
Low Bias
High Variance
Prediction Error
Test Sample
Training Sample
Low
High
Model Complexity





http://www.jmlr.org/papers/volume13/bergstra12a/bergstra12a.pdf

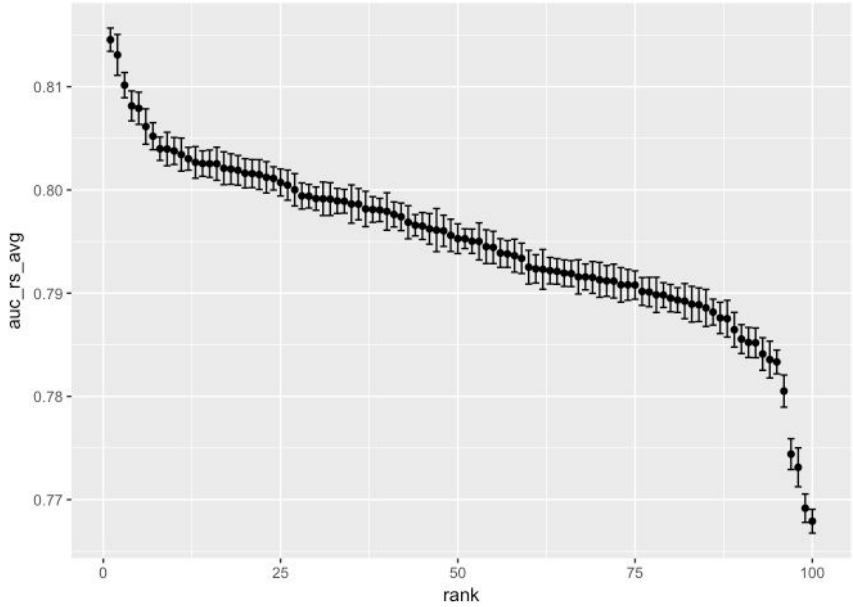


- Gaussian Processes (GP)
- Tree of Parzen Estimators (TPE)
- Sequential Model-based Algorithm Configuration (SMAC)

http://www.argmin.net/2016/06/20/hypertuning/

szilard / GBM-tune

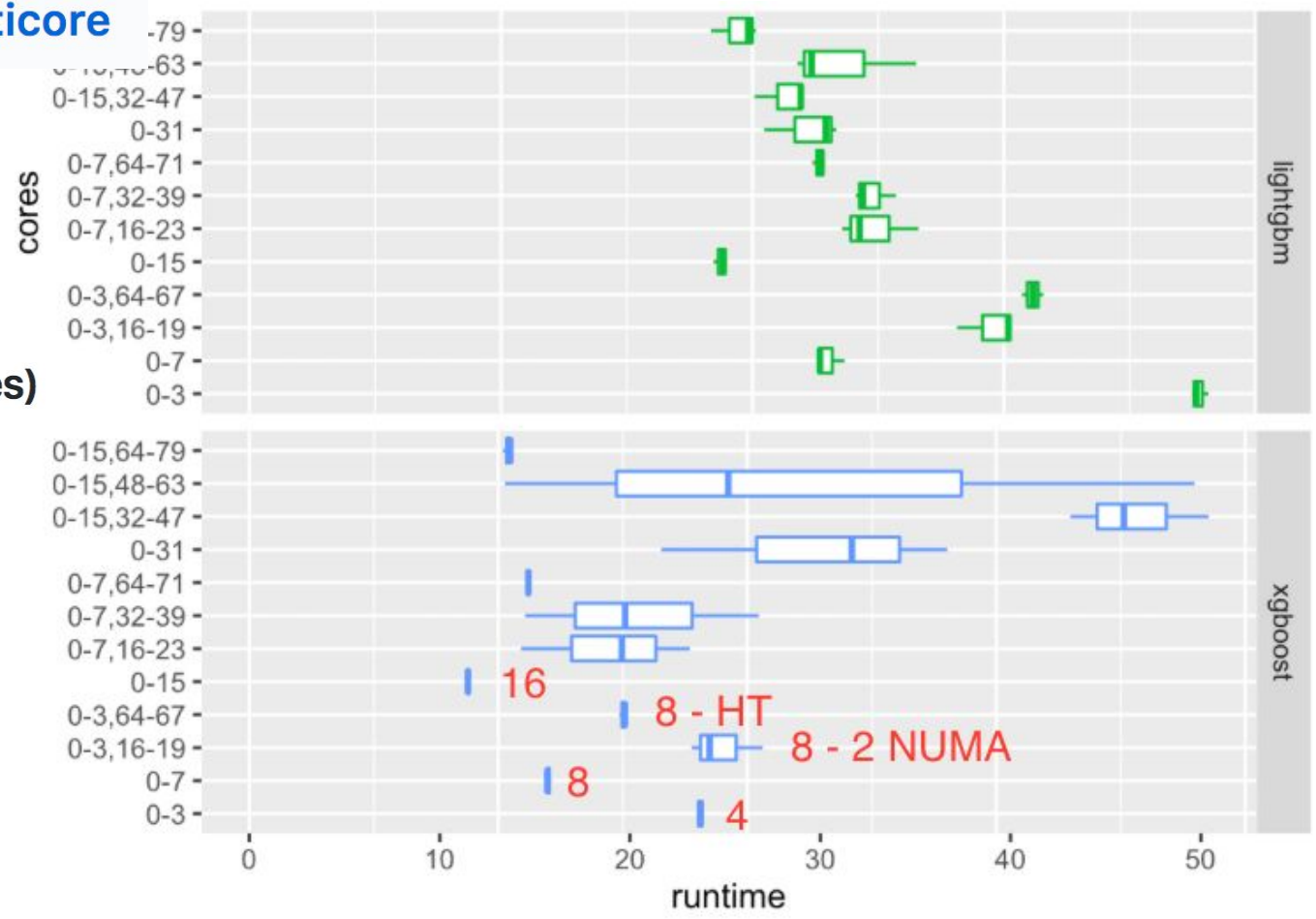
```
n_random <- 100      ## TODO: 1000?

params_grid <- expand.grid(                    # default
  num_leaves = c(100,200,500,1000,2000,5000),  # 31 (was 127)
  learning_rate = c(0.01,0.03,0.1),            # 0.1
  min_data_in_leaf = c(5,10,20,50),            # 20 (was 100)
  feature_fraction = c(0.6,0.8,1),             # 1
  bagging_fraction = c(0.4,0.6,0.8,1),         # 1
  lambda_l1 = c(0,0,0,0, 0.01, 0.1, 0.3),      # 0
  lambda_l2 = c(0,0,0,0, 0.01, 0.1, 0.3)       # 0
## TODO:
  ## min_sum_hessian_in_leaf
  ## min_gain_to_split
  ## max_bin
  ## min_data_in_bin
)
```



	num_leaves	learning_rate	min_data_in_leaf	feature_fraction
1	1000	0.03	5	0.8
2	1000	0.03	5	0.8
3	500	0.01	10	0.8
4	5000	0.01	5	0.8
5	500	0.01	10	0.8
6	5000	0.01	10	0.8
7	1000	0.10	20	0.8
8	5000	0.01	10	0.8
9	500	0.10	20	0.8
10	500	0.01	5	0.6

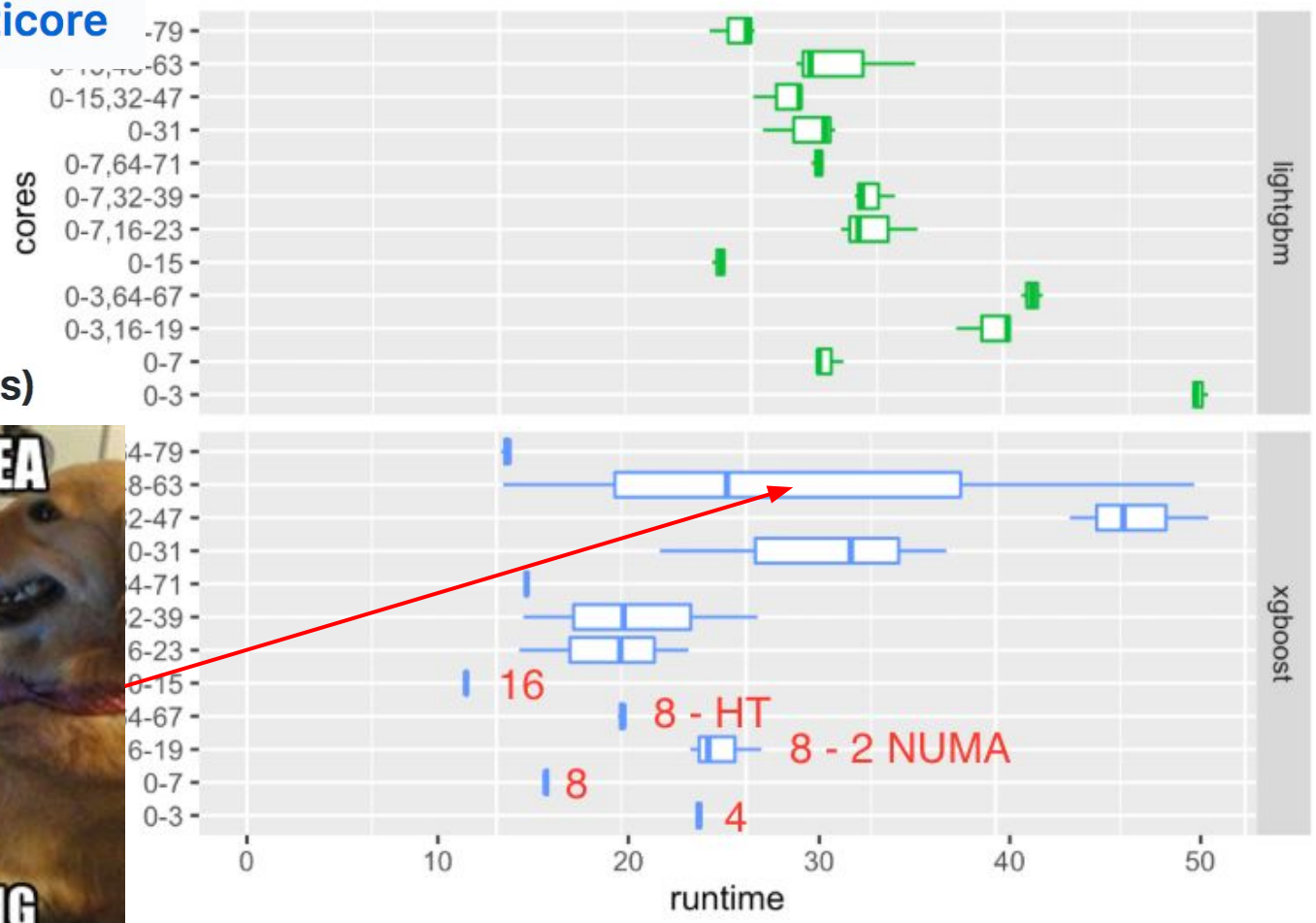
szilard / **GBM-multicore**

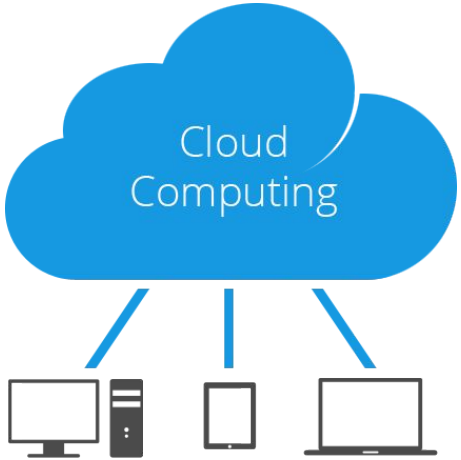
x1.32xlarge (128 cores)



szilard / GBM-multicore

x1.32xlarge (128 cores)

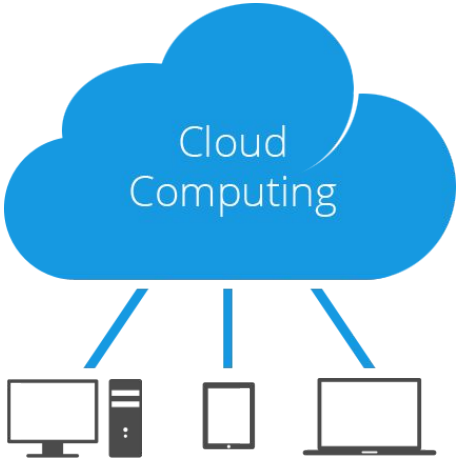




ML training:

lots of CPU cores
lots of RAM

limited time



ML training:

lots of CPU cores
lots of RAM

limited time

KDnuggets @kdnuggets · 8 Jul 2016

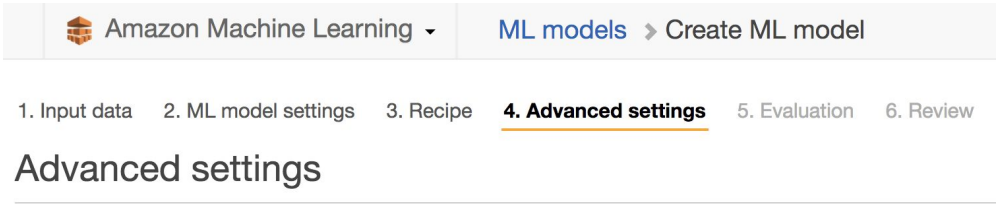
Cloud #MachineLearning Wars: Amazon vs IBM Watson vs Microsoft Azure #KDN ow.ly/w57T301V3oa

1 17 10

Szilard
@DataScienceLA

Replying to @kdnuggets

.@kdnuggets training on 10M-row dataset github.com/szilard/benchm… AmazonML 1500 sec, R (glmnet) 100 sec, Vowpal Wabbit 15 sec - all 3 same AUC



KDnuggets @kdnuggets · 8 Jul 2016

Cloud #MachineLearning Wars: Amazon vs IBM Watson vs Microsoft Azure #KDN ow.ly/w57T301V3oa

1 17 10

Szilard
@DataScienceLA

Replying to @kdnuggets

.@kdnuggets training on 10M-row dataset github.com/szilard/benchm… AmazonML 1500 sec, R (glmnet) 100 sec, Vowpal Wabbit 15 sec - all 3 same AUC

CLOUD MACHINE LEARNING ENGINE

Machine Learning on any data, of any size

TRY IT FREE VIEW CONSOLE

Amazon Machine Learning ML models › Create ML model

1. Input data 2. ML model settings 3. Recipe **4. Advanced settings** 5. Evaluation 6. Review

Advanced settings

KDnuggets @kdnuggets · 8 Jul 2016

Cloud #MachineLearning Wars: Amazon vs IBM Watson vs Microsoft Azure #KDN ow.ly/w57T301V3oa

1 17 10

Szilard
@DataScienceLA

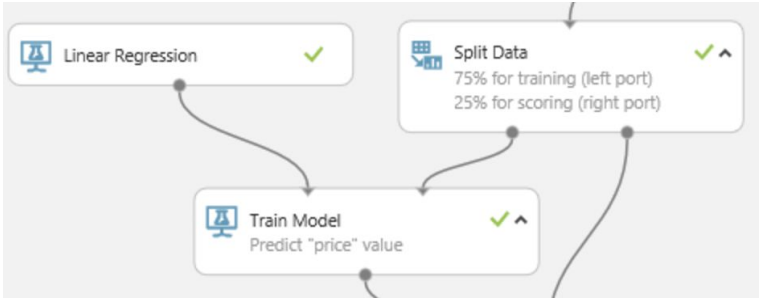
Replying to @kdnuggets

.@kdnuggets training on 10M-row dataset github.com/szilard/benchm… AmazonML 1500 sec, R (glmnet) 100 sec, Vowpal Wabbit 15 sec - all 3 same AUC

CLOUD MACHINE LEARNING ENGINE

Machine Learning on any data, of any size

TRY IT FREE VIEW CONSOLE



Amazon Machine Learning ML models › Create ML model

1. Input data 2. ML model settings 3. Recipe **4. Advanced settings** 5. Evaluation 6. Review

Advanced settings

KDnuggets @kdnuggets · 8 Jul 2016

Cloud #MachineLearning Wars: Amazon vs IBM Watson vs Microsoft Azure #KDN ow.ly/w57T301V3oa

1 17 10

Szilard
@DataScienceLA

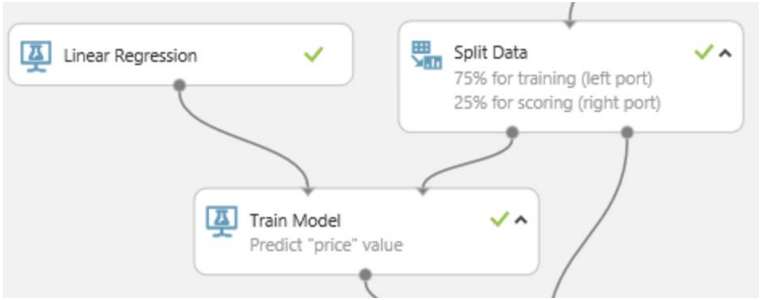
Replying to @kdnuggets

.@kdnuggets training on 10M-row dataset github.com/szilard/benchm… AmazonML 1500 sec, R (glmnet) 100 sec, Vowpal Wabbit 15 sec - all 3 same AUC

CLOUD MACHINE LEARNING ENGINE

Machine Learning on any data, of any size

TRY IT FREE VIEW CONSOLE



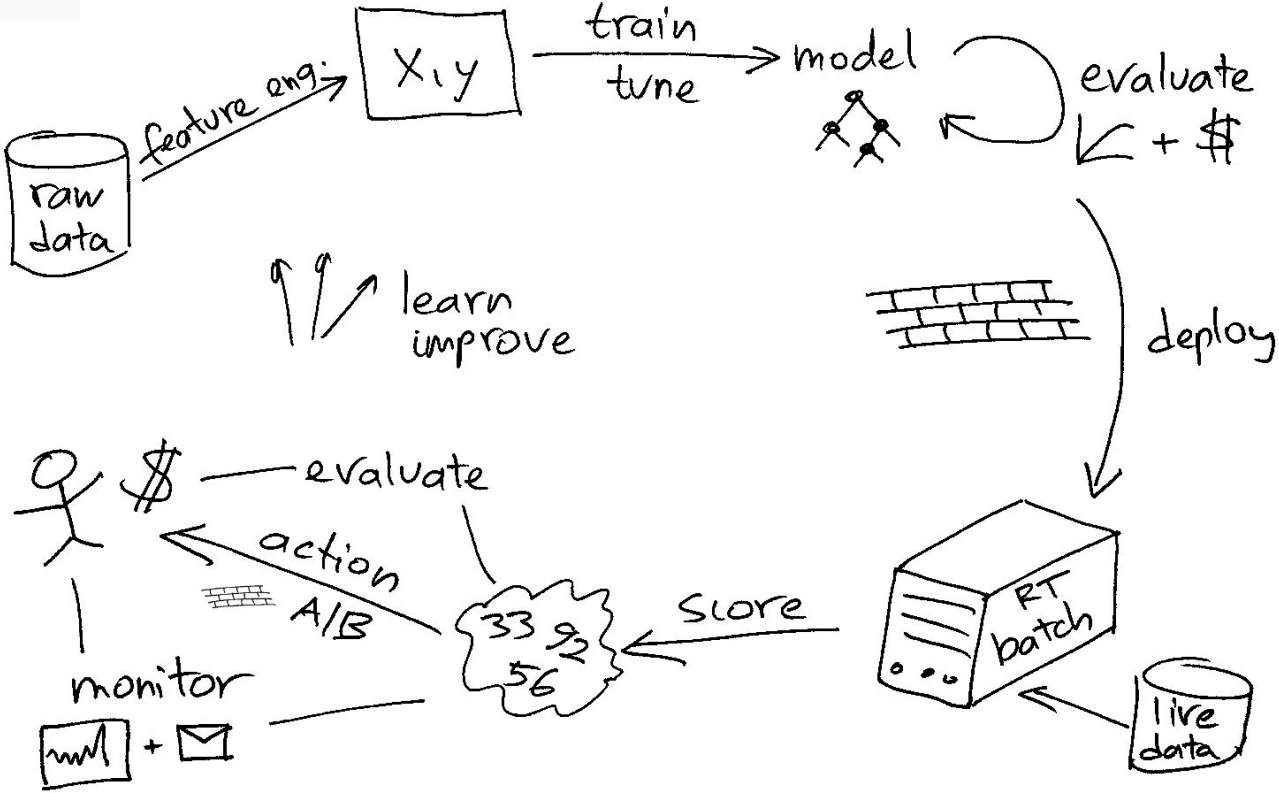
Amazon Machine Learning ML models › Create ML model

1. Input data 2. ML model settings 3. Recipe **4. Advanced settings** 5. Evaluation 6. Review

Advanced settings

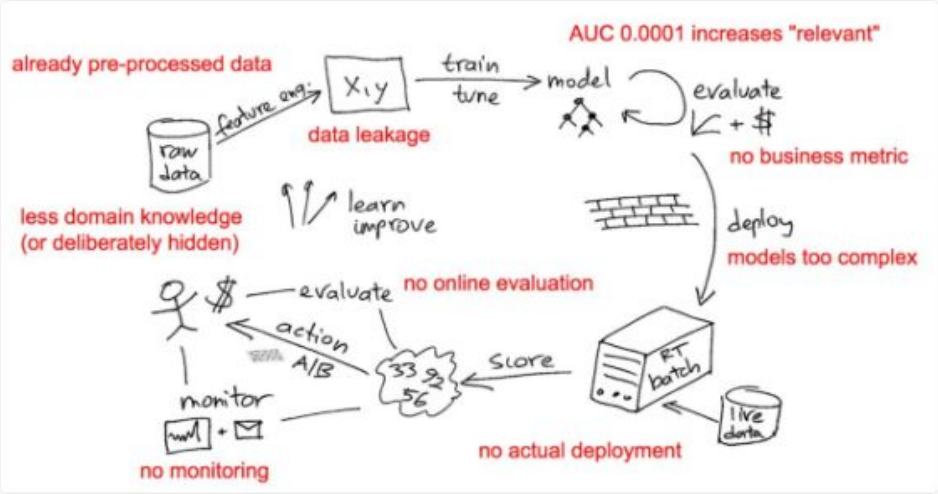
“people that know what they’re doing just use open source [...] the same open source tools that the MLaaS services offer” - Bradford Cross

szilard / **ml-prod**



Szilard
@DataScienceLA

If you do #kaggle to learn #machinelearning, you are missing on 80% of things you need for ML in real life/production



3:41 PM - 24 Aug 2017

252 Retweets **323** Likes

13 252 323

	gbm (R pkg)	xgboost	lightgbm	h2o
easy R install	cran	cran	linux OK	java+cran
maintained	no	yes	yes	yes
preprocessing	not needed	1-hot	1-hot/categ int	not needed
new cats scoring	yes	no	no	yes
early stopping	no	yes	yes	yes
speed (CPU)	1 core	ok	fastest	slow (small data)
GPU supported	no	yes	yes	via xgboost
speed GPU	NA	fastest	ok/slow	indirectly/slower
REST scoring	no	no	no	yes
other algos	no	RF	RF	RF/GLM/NN
best for		Kaggle	Kaggle	prod/real-time

Szilard
@DataScienceLA

Friday fun: what's your favorite gradient boosting machine (GBM) library?

58% xgboost

16% lightgbm

24% h2o

2% spark mllib

127 votes • Final results

3:21 PM - 11 May 2018

Szilard
@DataScienceLA

Friday fun: what's your favorite gradient boosting machine (GBM) library?

58% xgboost

16% lightgbm

24% h2o

2% spark mllib

no-one is using this crap

127 votes • Final results

3:21 PM - 11 May 2018

Szilard @DataScienceLA · May 16

If you are using gradient boosting machines(GBM), are you running it (training) on GPUs or CPUs?

7% GPU

93% CPU

55 votes • Final results

Szilard @DataScienceLA · Apr 23

What algo do you use the most for supervised learning?

33% lin/logistic regression

56% random forest/GBM

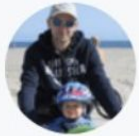
4% shallow neural nets

7% deep learning

57 votes • Final results

Szilard @DataScienceLA · Apr 25

Final fucking results. Sure, the results are biased, but still 😂🤣😍 At least my followers know better than using #deeplearning for every fucking problem. #machinelearning twitter.com/DataScienceLA/…



Szilard @DataScienceLA · Apr 23

What algo do you use the most for supervised learning?

33% lin/logistic regression

56% random forest/GBM

4% shallow neural nets

7% deep learning

57 votes • Final results

2 5 28

More:

szilard / benchm-ml ★ Star 1,203

szilard / teach-data-science-UCLA-master-appl-stats

szilard / teach-ML-CEU-master-bizanalytics

szilard / talks

szilard / GBM-workshop

szilard / ML-scoring

szilard / GBM-tune

GitHubGist Search...

szilard / GBM-perf

szilard / h2o_scoring.R

szilard / GBM-multicore

szilard / ML_with_H2O.R

spafka@gmail.com

@DataScienceLA

linkedin.com/in/szilard

github.com/szilard