H_2O ai

HANJO ODENDAAL

Scalable Analytics with H2O and AWS

SatRday 2019 - Johannesburg

www.daeconomist.com

DATA SCIENCE GOALS

If you take ONE thing away from todays talk, it should be this:
Take R100 ($7.14), spin up BIG AWS machine (r4.16xlarge) and
watch for 10 hours light up like a christmas tree while telling
your line manager your tuning you ML model. #geekgoals

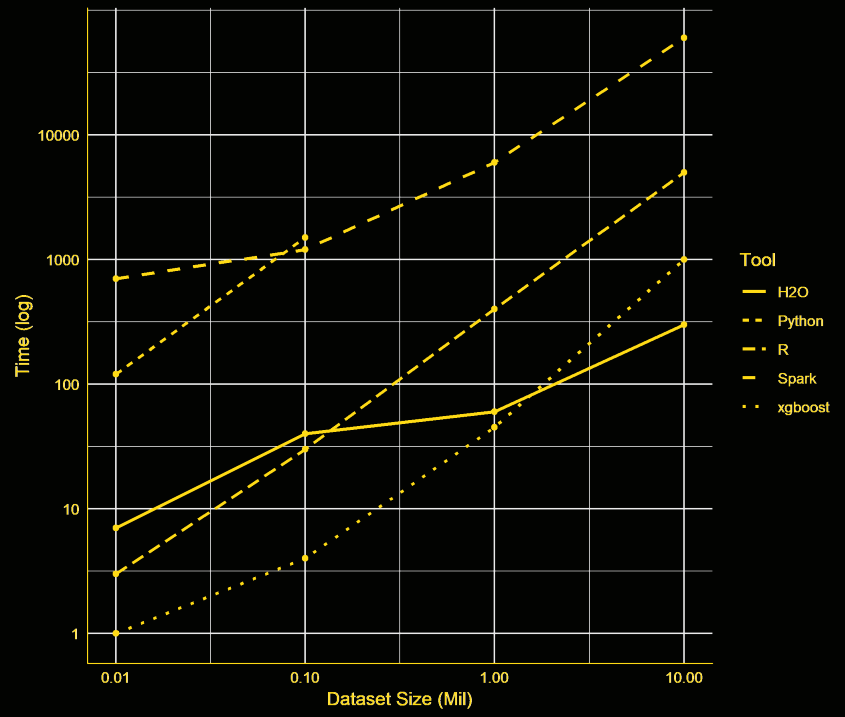

H2O.ai

NO BULLSH*T DATA SCIENCE

Szilard Pafka's[1] R/Finance 2017 presentation had a big influence on how I saw the hype around the industry

- Don't always believe the hype!
- Scaling shouldn't be an excercise in fighting your tools

Stop wasting time - if you interested in machine learning. Learn all you can about XGBoost and Random Forests[2]



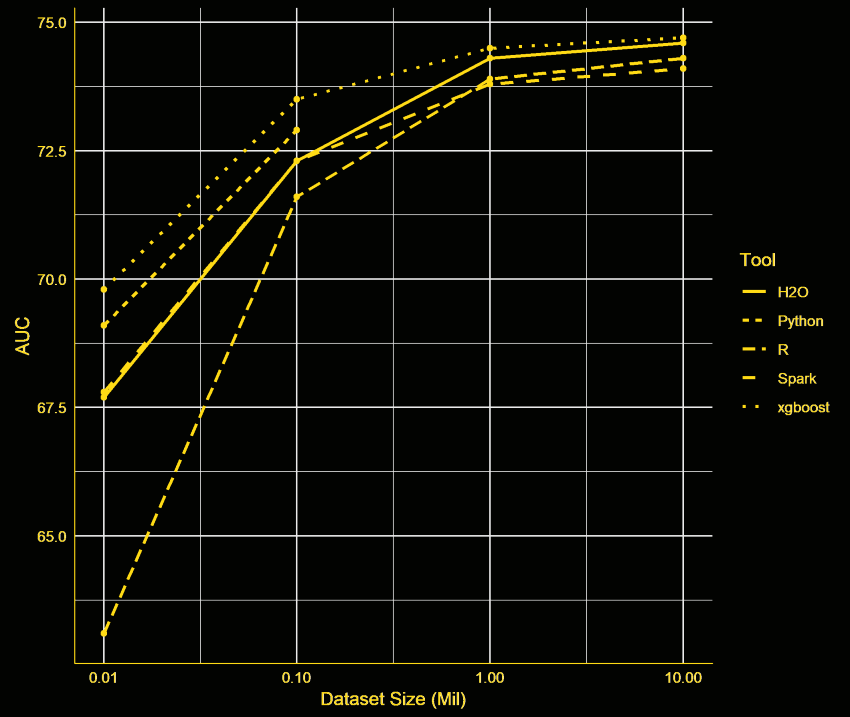
[1] Hope I am pronouncing this correctly
[2] Which means you can concentrate on the stats and maths behind each model
Talk source: https://bit.ly/2ErsBtB

NO BULLSH*T DATA SCIENCE

Szilard Pafka's[1] R/Finance 2017 presentation had a big influence on how I saw the hype around the industry

- Don't always believe the hype!
- Scaling shouldn't be an excercise in fighting your tools

Stop wasting time - if you interested in machine learning. Learn all you can about XGBoost and Random Forests[2]



[1] Hope I am pronouncing this correctly
[2] Which means you can concentrate on the stats and maths behind each model
Talk source: https://bit.ly/2ErsBtB

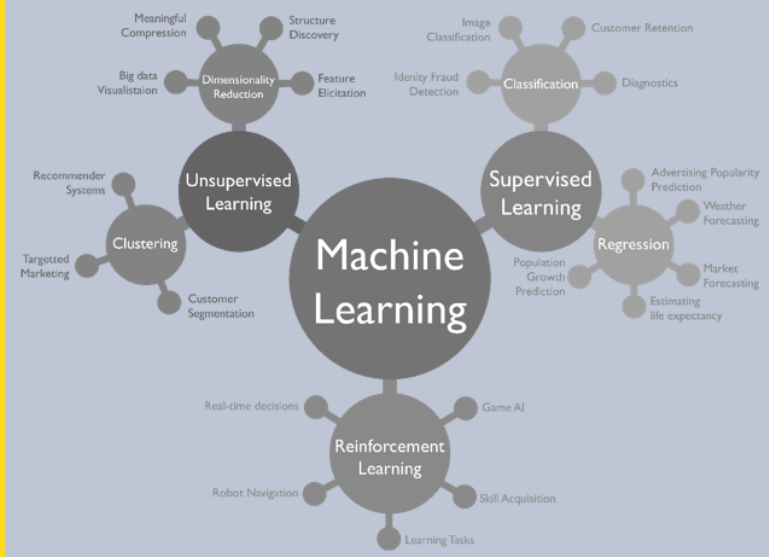
H2O

Leading open source platform for machine learning and artificial intelligence. Multiple API interface such as , , , , .

Well developed interfaces around GPUs.

AWS.EC2

Amazon Elastic Compute. For analytics, their are perfect as they come at an enourmously discounted price as well as delivering bag for buck.

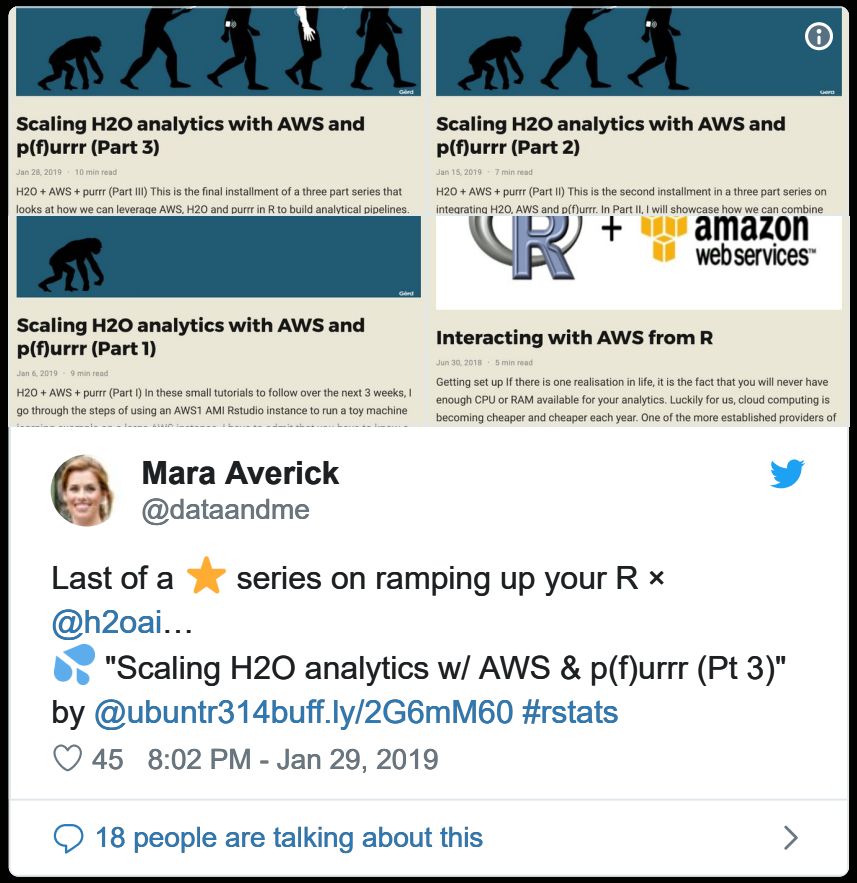


80% exploratory research still in caret

But, running H20 in live systems.

WHERE DO YOU START?

- Very nice high level functions and methods
- It really does scale nicely on large datasets
- Code interface smells good



NO EXCUSE NOT TOO PLAY

r3.xlarge

- 4 cores
- 30.5 GB RAM
- $0.0379 (R0,53/h)

r4.16xlarge

- 64 cores
- 488 GB RAM
- $0.6974 (R9,76/h)

H2O HAS A SIGNIFICANT AMOUNT OF TUNING OPTIONS:

XGBoost:

Random Forest:

HELLO WORLD H2O:

Messaging confirming the server has been intitialized sucessfully

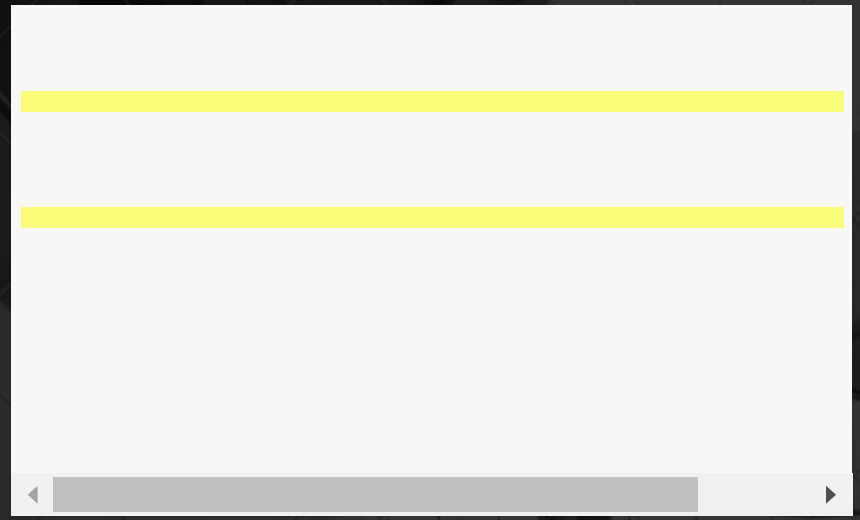
HELLO WORLD H2O:

HELLO WORLD H2O:

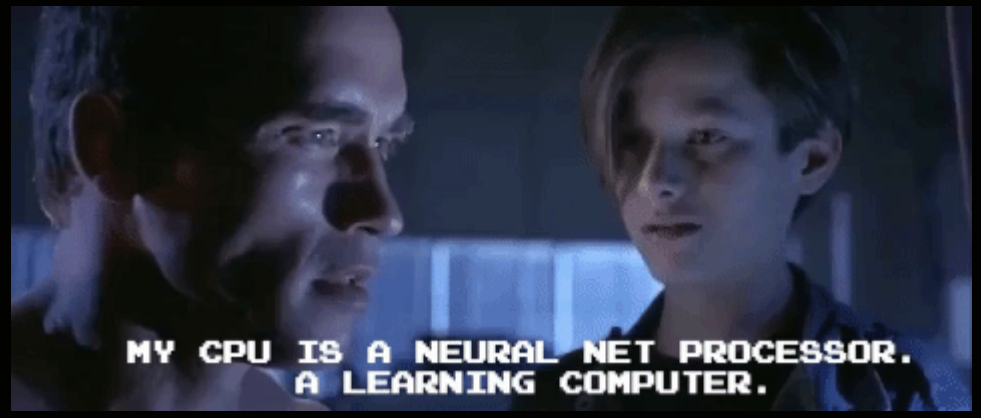
FURRR NEVER LOOKED SO GOOD

Interact with EC2 through :

to connect to server



ON TO AN EXAMPLE



THANK YOU