aws SUMMIT TEL AVIV

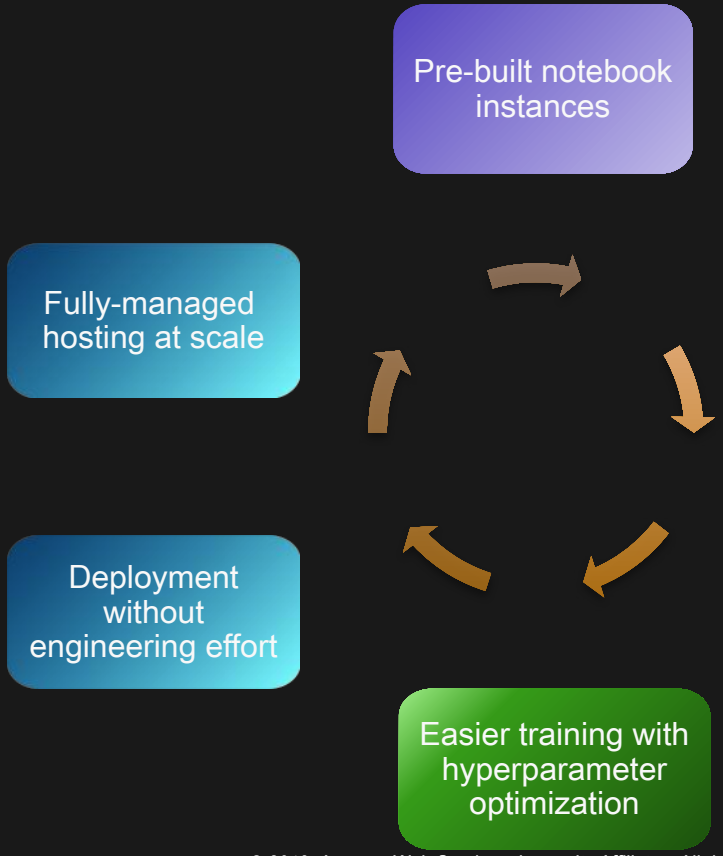
Speed up your Machine Learning workflows with built-in algorithms

Julien Simon

Principal AI/ML Evangelist, Amazon Web Services

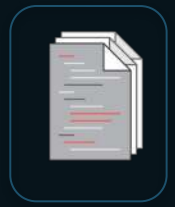
@julsimon

Amazon SageMaker



scikit learn
mxnet
GLUON
TensorFlow
PYTORCH

Amazon SageMaker: model options



Training code

- Matrix Factorization
- Regression
- Principal Component Analysis
- K-Means Clustering
- Gradient Boosted Trees
- And More!

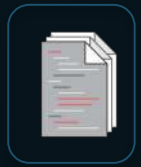
Amazon provided Algorithms

mxnet TensorFlow

Bring Your Own Script

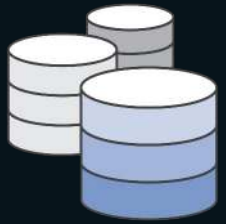
APACHE Spark

IM Estimators in Apache Spark



Bring Your Own Container

Amazon SageMaker: 10x better algorithms



Streaming datasets, for cheaper training

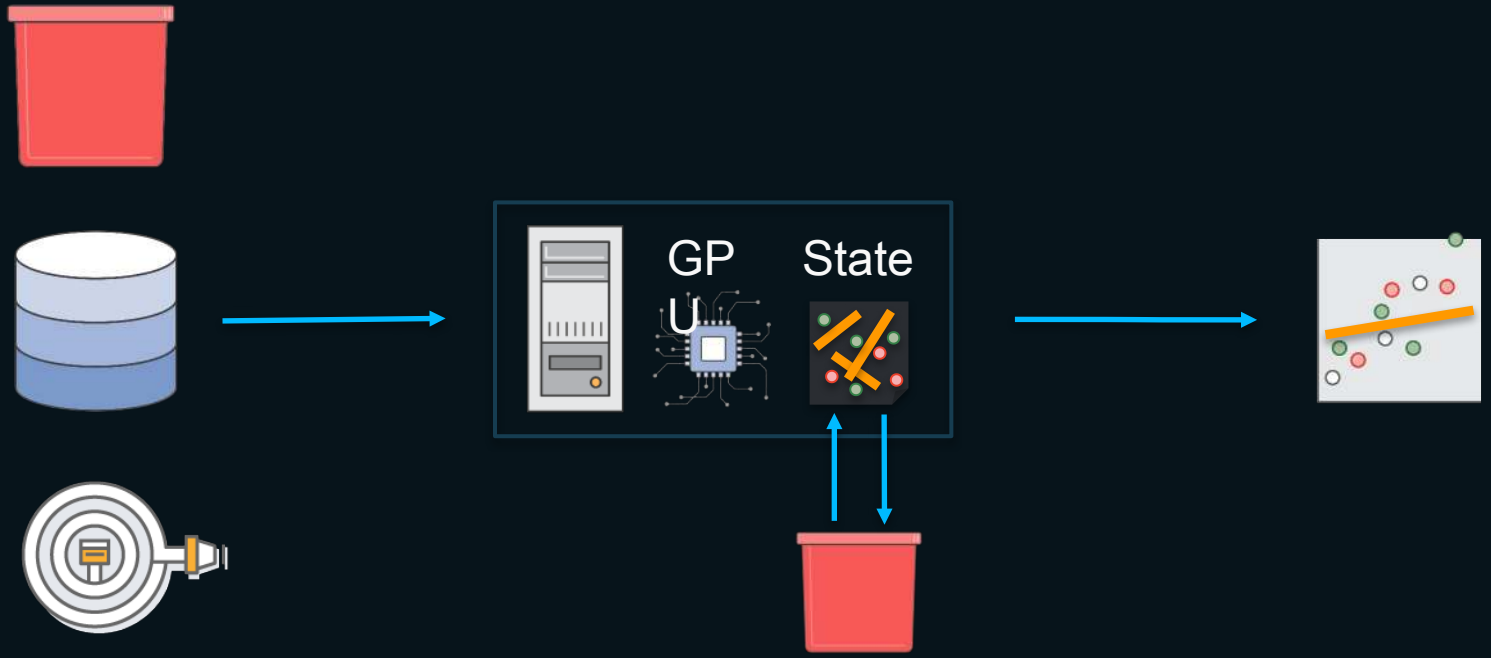
Train faster, in a single pass

Greater reliability on extremely large datasets

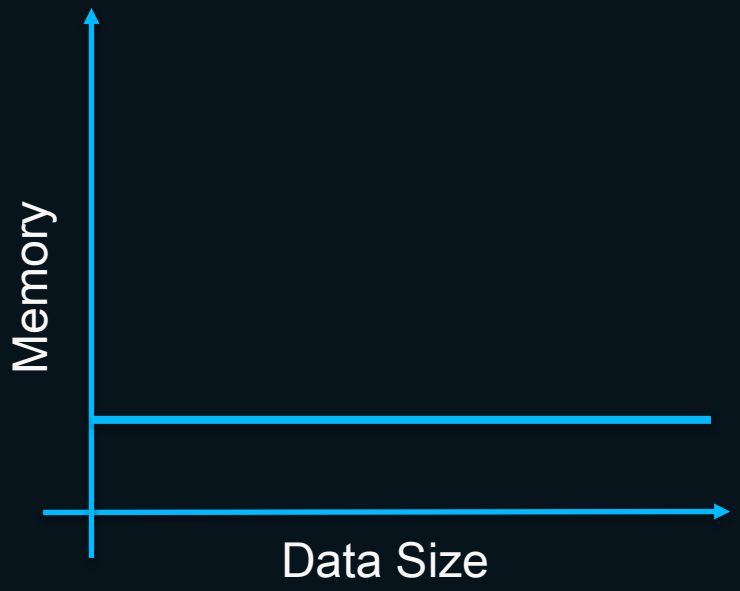
Choice of several ML algorithms

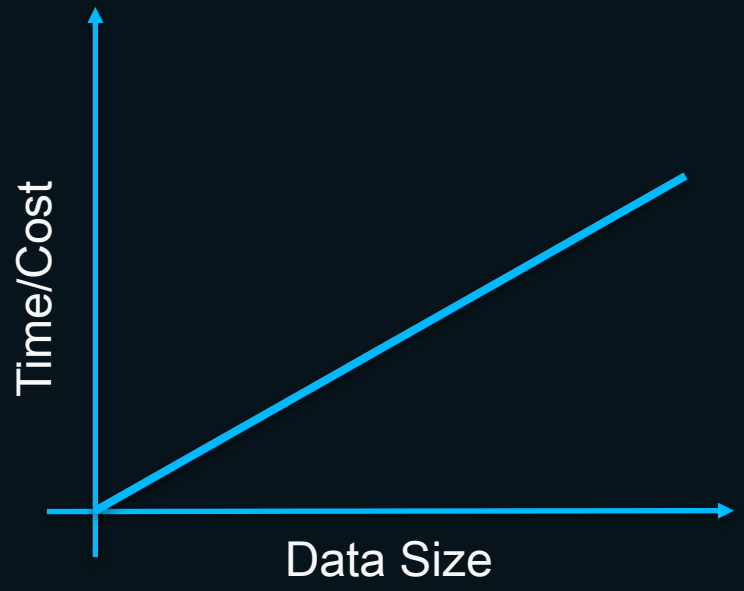
Infinitely scalable algorithms

Streaming

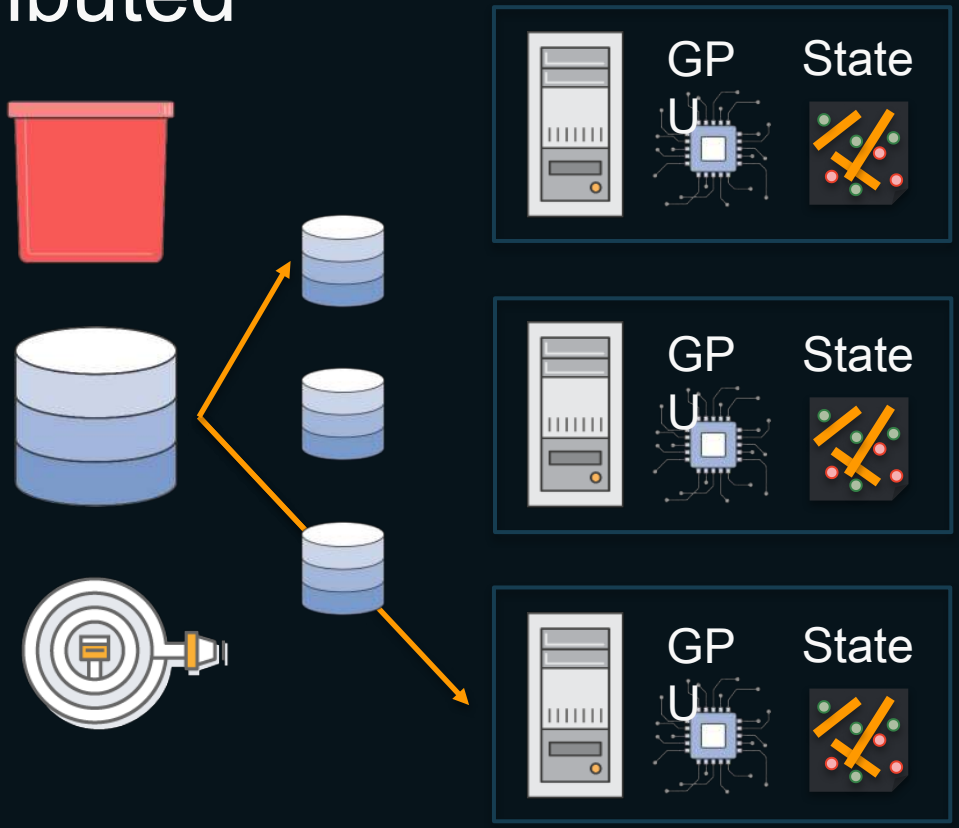


Streaming

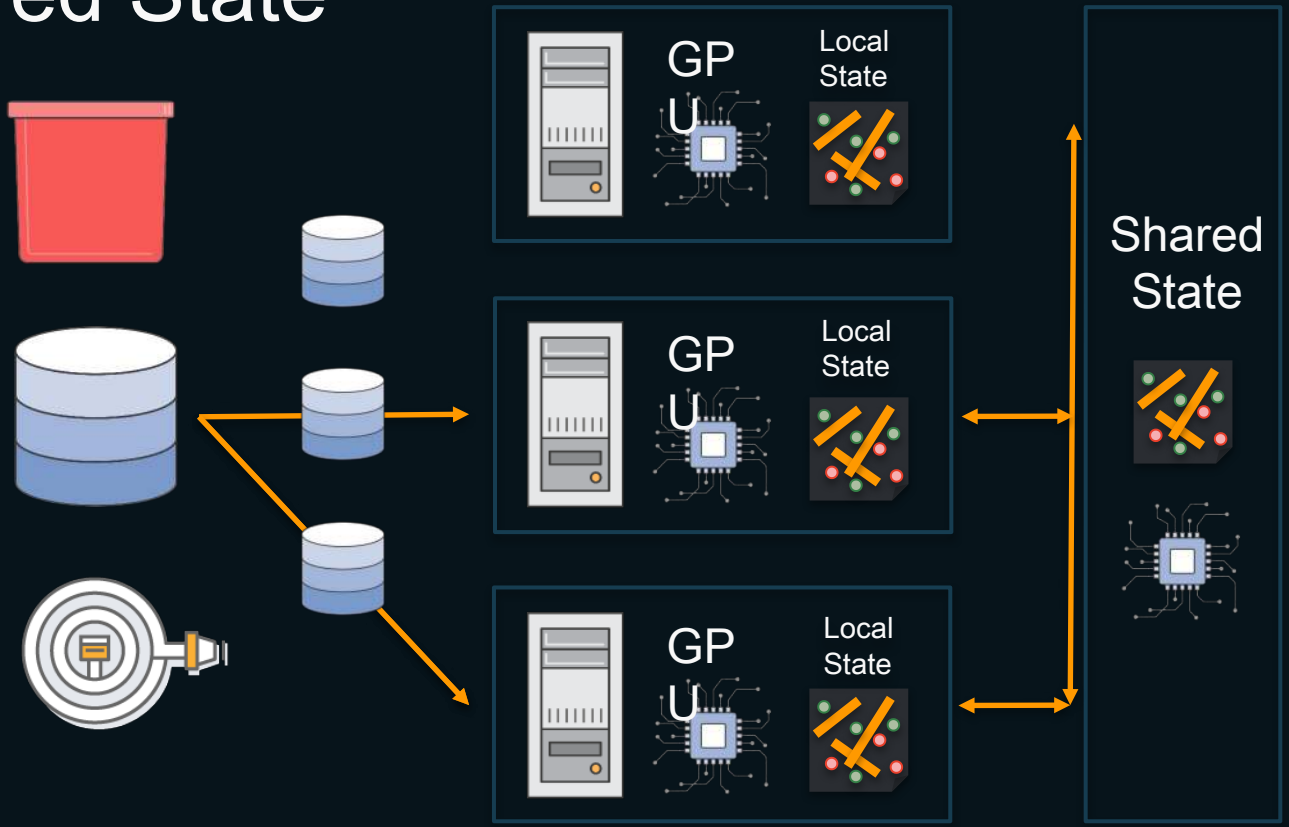




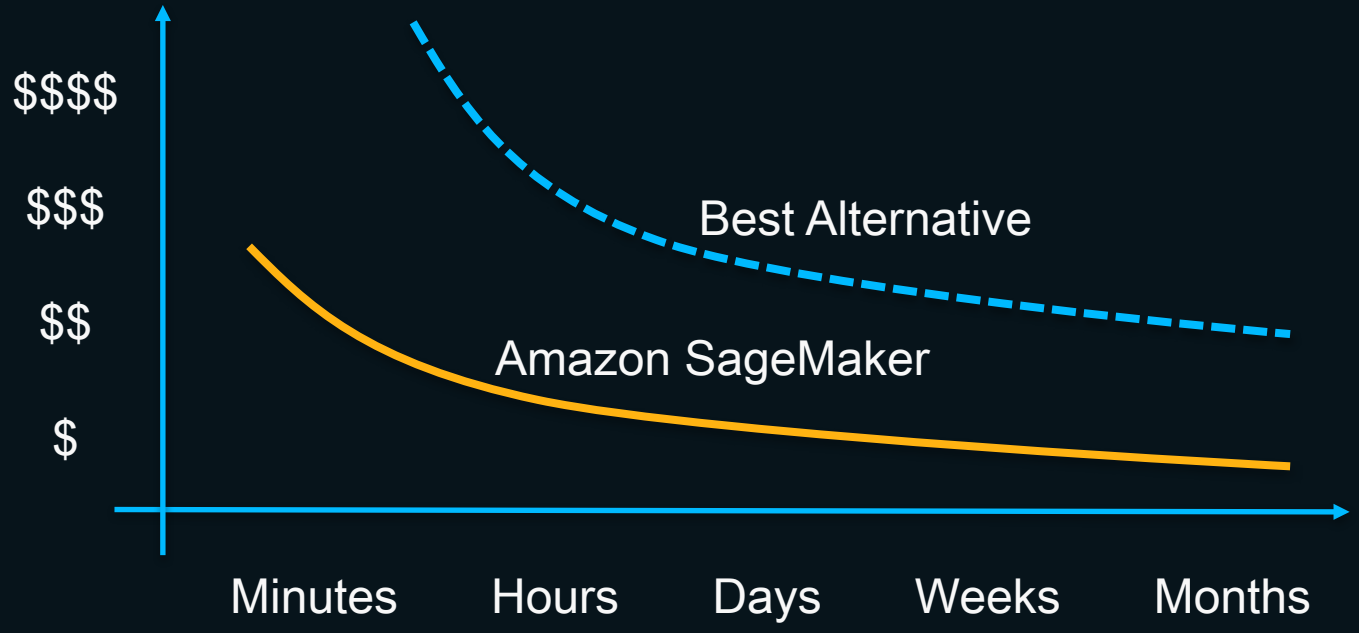
Distributed



Shared State



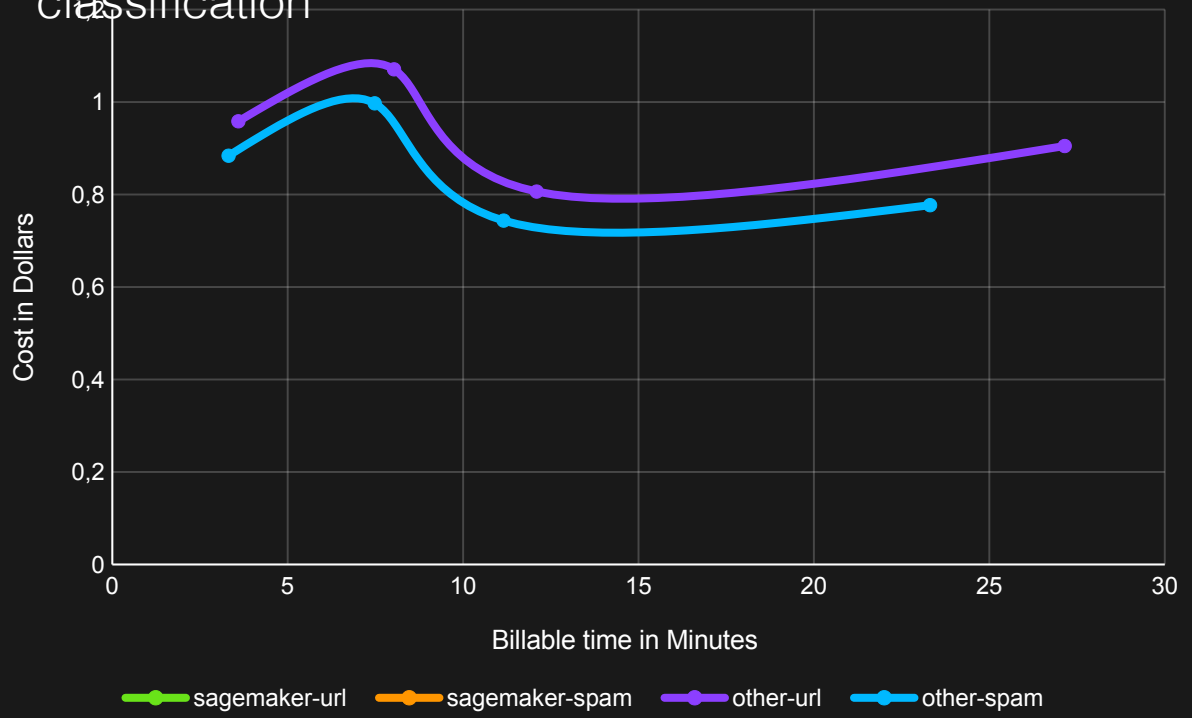
Cost vs. Time



Linear Learner

Regression (mean squared error)	
SageMaker	Other
1.02	1.06
1.09	**1.02**
0.332	**0.183**
0.086	0.129
83.3	84.5
Classification (F1 Score)	
SageMaker	Other
0.980	**0.981**
0.870	**0.930**
0.997	**0.997**
0.978	0.964
0.914	0.859
0.470	**0.472**
0.903	**0.908**
0.508	0.508

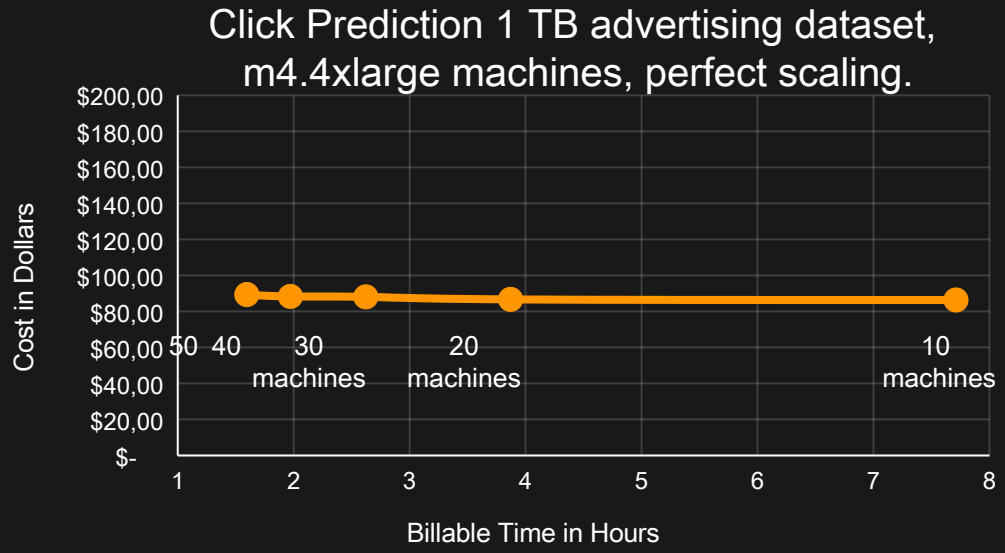
30 GB datasets for web-spam and web-url classification



Factorization Machines

$$\tilde{y} = w_0 + \langle w_1, x \rangle + \sum_{i,j>i} x_i x_j \cdot \langle v_i, v_j \rangle$$

	Log_loss	F1 Score	**Seconds**
SageMaker	0.494	0.277	820
Other (10 Iter)	0.516	0.190	650
Other (20 Iter)	0.507	0.254	1300
Other (50 Iter)	0.481	0.313	3250

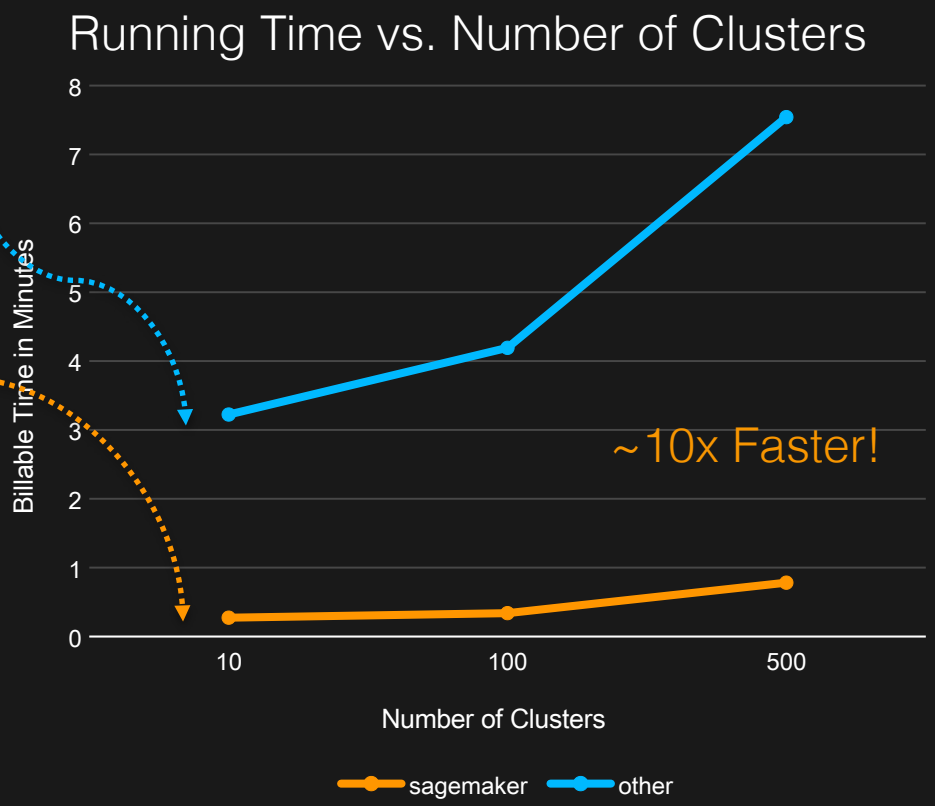


Demo: building a movie recommender with Factorization Machines

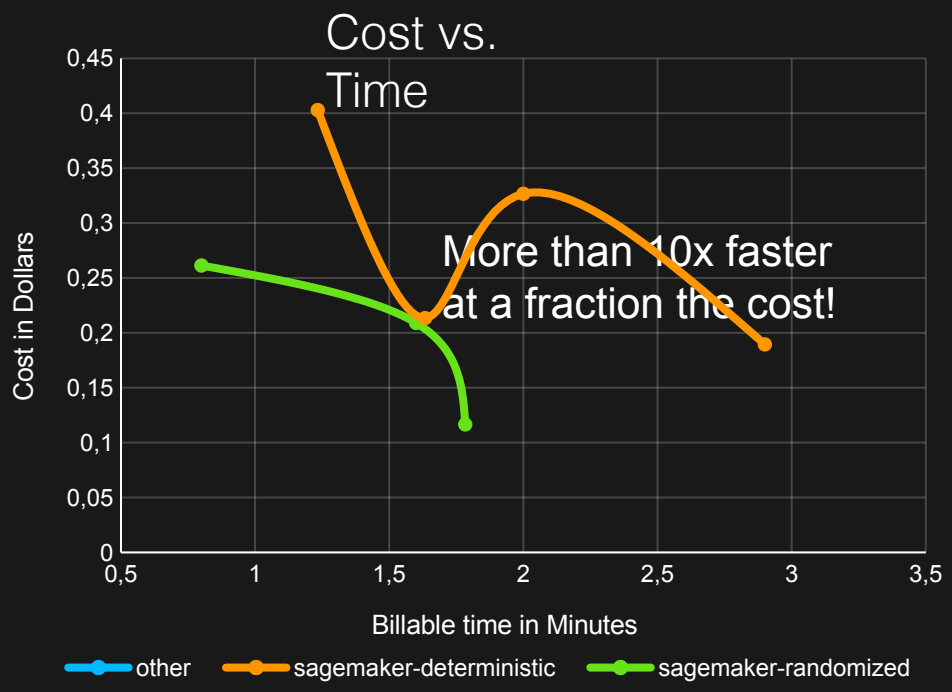
https://medium.com/@julsimon/building-a-movie-recommender-with-factorization-machines-on-amazon-sagemaker-cedbfc8c93d8

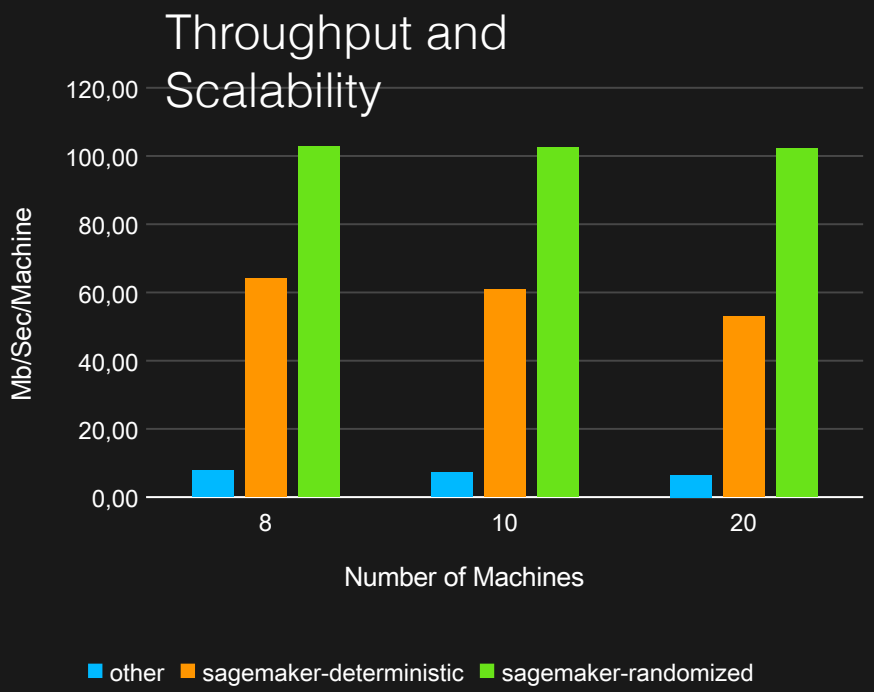
K-Means Clustering

	k	SageMaker	Other
Text 1.2GB	10	1.18E3	1.18E3
	100	1.00E3	9.77E2
	500	9.18.E2	9.03E2
Images 9GB	10	3.29E2	3.28E2
	100	2.72E2	2.71E2
	500	2.17E2	Failed
Videos 27GB	10	2.19E2	2.18E2
	100	2.03E2	2.02E2
	500	1.86E2	1.85E2
Advertising 127GB	10	1.72E7	Failed
	100	1.30E7	Failed
	500	1.03E7	Failed
Synthetic 1100GB	10	3.81E7	Failed
	100	3.51E7	Failed
	500	2.81E7	Failed

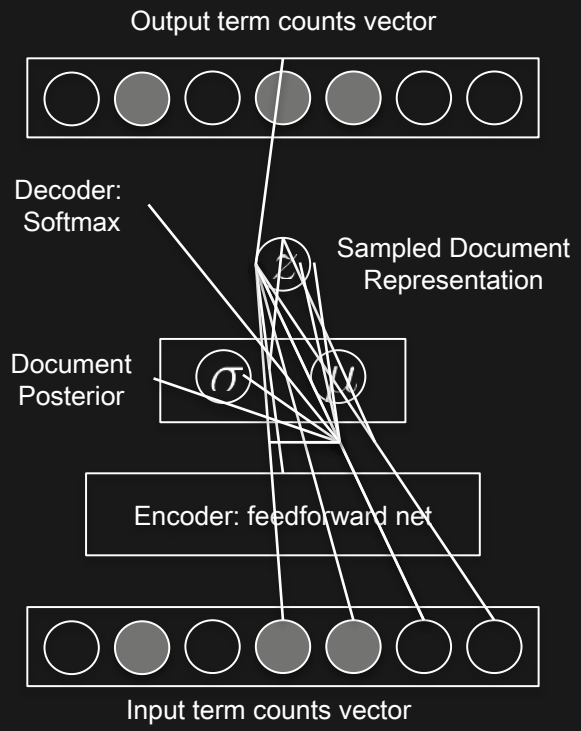


Principal Component Analysis (PCA)



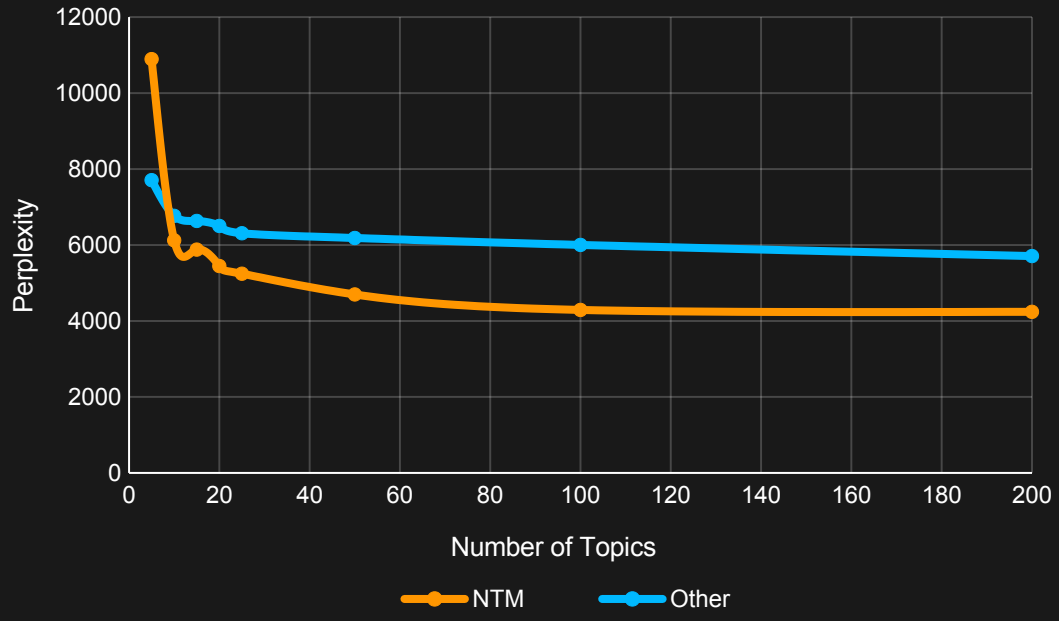


Neural Topic Modeling

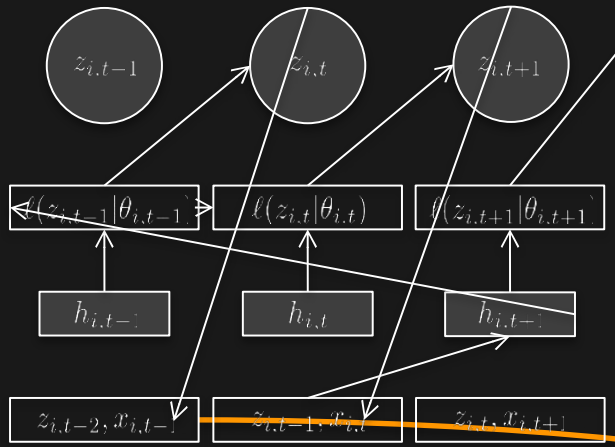


Perplexity vs. Number of Topic

(~200K documents, ~100K vocabulary)



DeepAR: Time Series Forecasting



		Mean absolute percentage error		P90 Loss	
		DeepAR	R	DeepAR	R
traffic Hourly occupancy rate of 963 Bay Area freeways		0.14	0.27	0.13	0.24
electricity Electricity use of 370 homes over time		0.07	0.11	0.08	0.09
pageviews Page view hits of websites	10k	0.32	0.32	0.44	0.31
	180k	0.32	0.34	0.29	NA

One hour on p2.xlarge, $1

DeepAR

DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks

Valentin Flunkert[*], **David Salinas**[*], **Jan Gasthaus**
Amazon Development Center
Germany
<dsalina,flunkert,gasthaus@amazon.com>

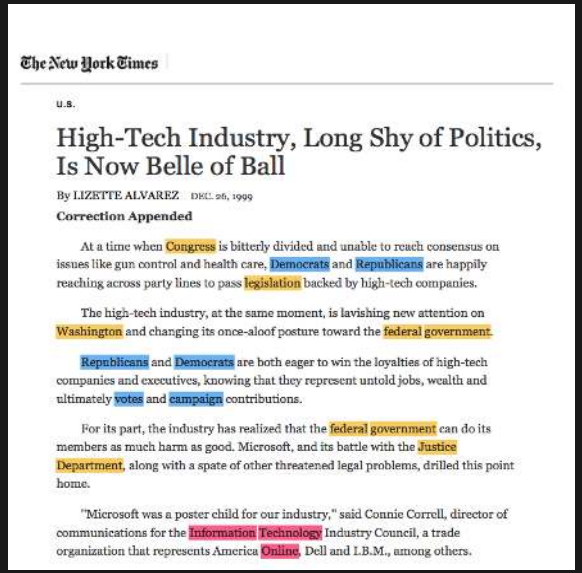
https://arxiv.org/abs/1704.04110

Demo: predicting world temperature with DeepAR

https://medium.com/@julsimon/predicting-world-temperature-with-time-series-and-deepar-on-amazon-sagemaker-e371cf94ddb5

More built-in algorithms

Spectral LDA



The New York Times

U.S.

High-Tech Industry, Long Shy of Politics, Is Now Belle of Ball

By LIZETTE ALVAREZ DEC. 26, 1999

Correction Appended

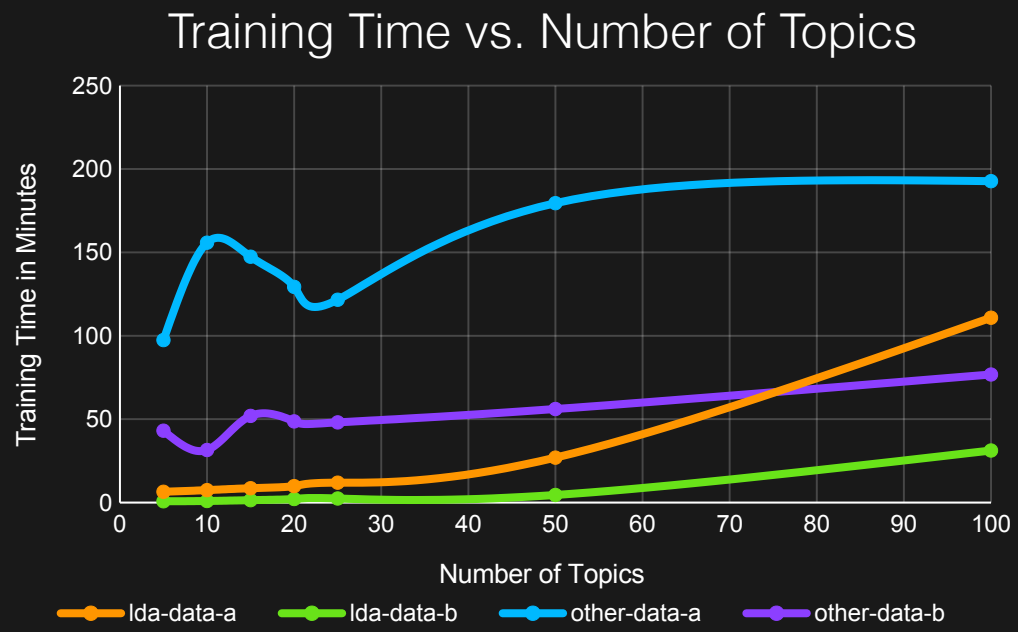
At a time when Congress is bitterly divided and unable to reach consensus on issues like gun control and health care, Democrats and Republicans are happily reaching across party lines to pass legislation backed by high-tech companies.

The high-tech industry, at the same moment, is lavishing new attention on Washington and changing its once-aloof posture toward the federal government.

Republicans and Democrats are both eager to win the loyalties of high-tech companies and executives, knowing that they represent untold jobs, wealth and ultimately votes and campaign contributions.

For its part, the industry has realized that the federal government can do its members as much harm as good. Microsoft, and its battle with the Justice Department, along with a spate of other threatened legal problems, drilled this point home.

"Microsoft was a poster child for our industry," said Connie Correll, director of communications for the Information Technology Industry Council, a trade organization that represents America Online, Dell and I.B.M., among others.

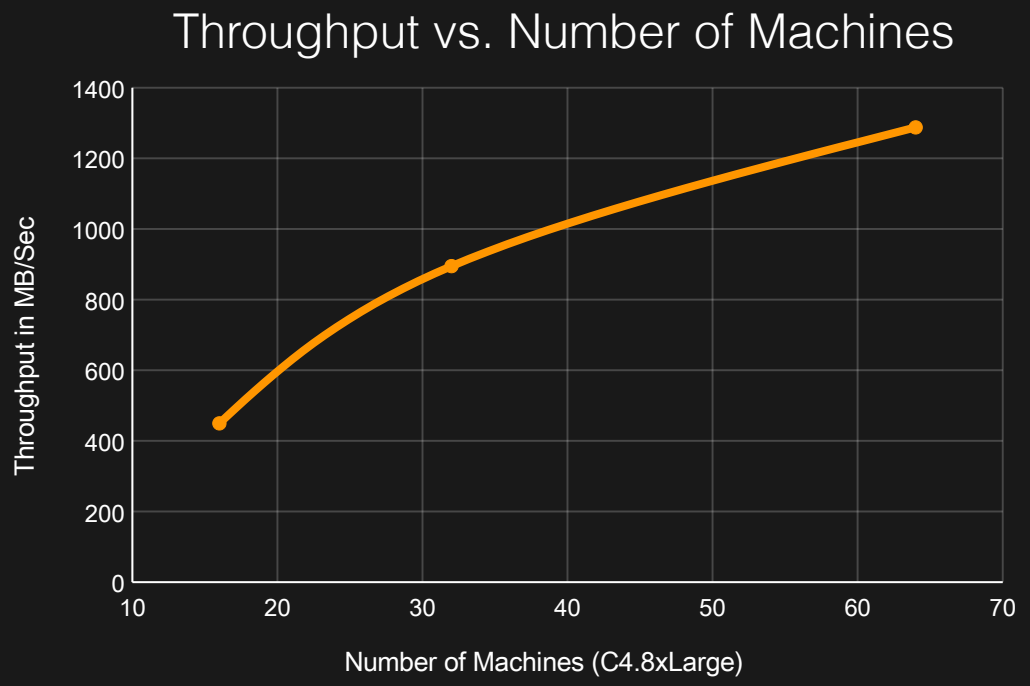


Boosted Decision Trees

XGBoost is one of the most commonly used classifiers.

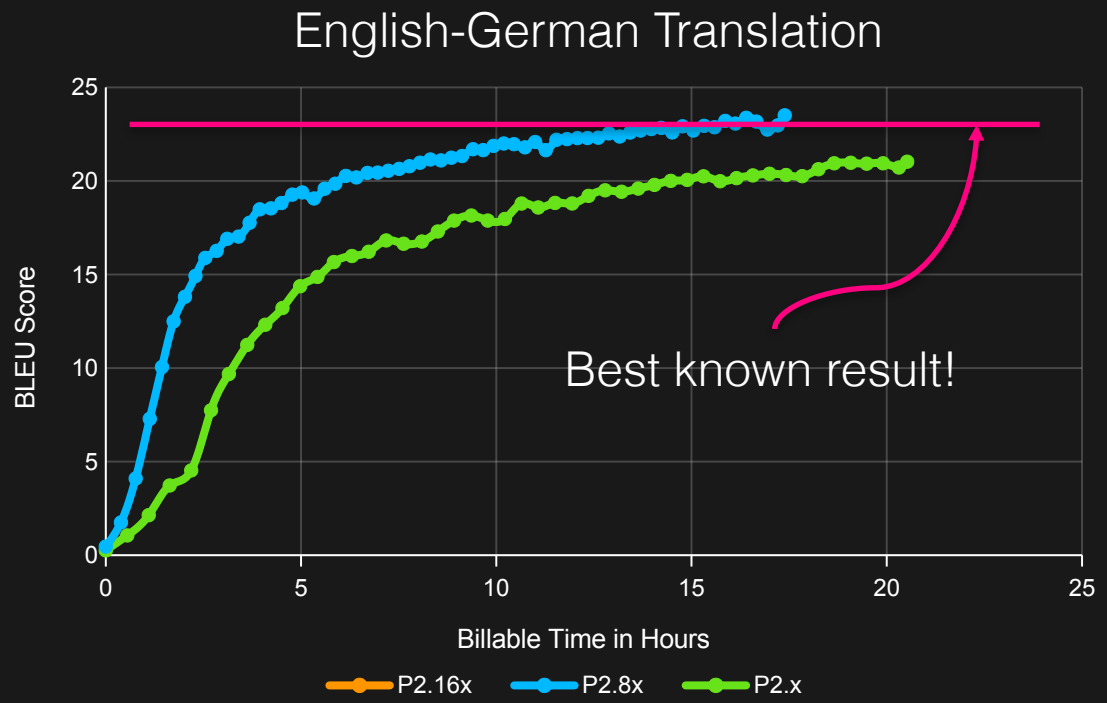
dmlc XGBoost

https://github.com/dmlc/xgboost



Sequence to Sequence

- Based on Sockeye and Apache MXNet.
- Multi-GPU.
- Can be used for Neural Machine Translation.
- Supports both RNN/CNN as encoder/decoder



Sockeye

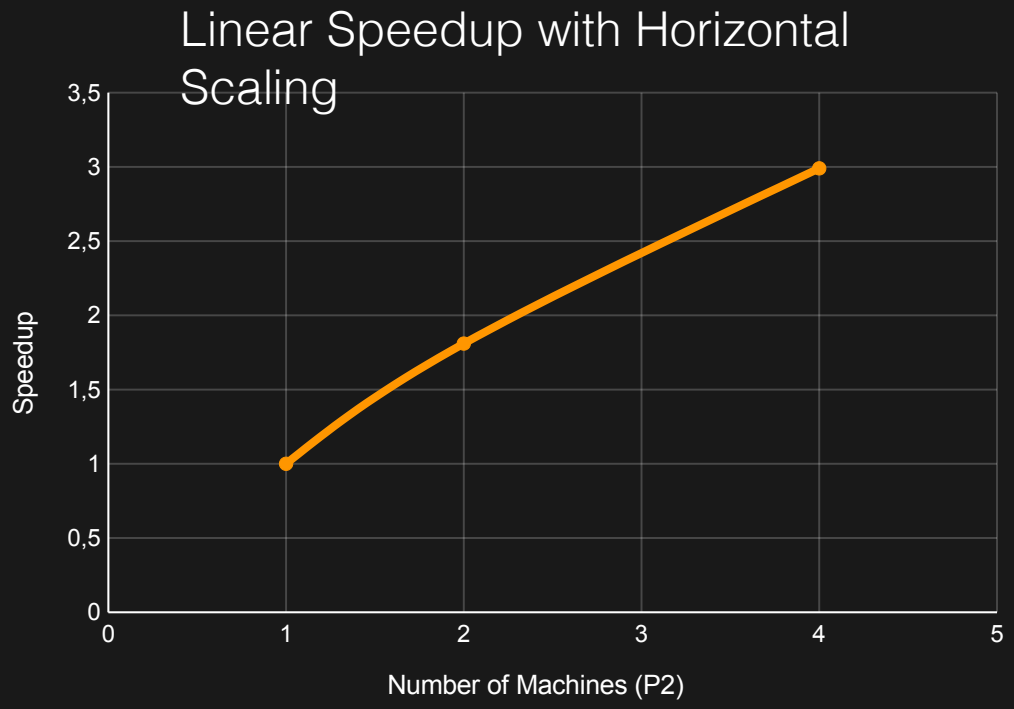
SOCKEYE: A Toolkit for Neural Machine Translation

Felix Hieber, Tobias Domhan, Michael Denkowski,
David Vilar, Artem Sokolov, Ann Clifton, Matt Post
{fhieber,domhant,mdenkows,dvilar,artemsok,acclift,mattpost}@amazon.com

https://arxiv.org/abs/1712.05690
https://github.com/awslabs/sockeye

Image Classification

- ResNet implementation with Apache MXNet.
- More networks to come.
- Transfer learning: begin with a model already trained on ImageNet!



Demo: fine-tuning an image classification model

https://medium.com/@julsimon/image-classification-on-amazon-sagemaker-9b66193c8b54

Latest addition: Blazing Text

BlazingText: Scaling and Accelerating Word2Vec using Multiple GPUs

Saurabh Gupta
Amazon Web Services
gsaur@amazon.com

Vineet Khare
Amazon Web Services
vkhare@amazon.com

https://dl.acm.org/citation.cfm?id=3146354

Resources

https://aws.amazon.com/machine-learning

https://aws.amazon.com/blogs/ai

https://aws.amazon.com/sagemaker (free tier available)

https://github.com/awslabs/amazon-sagemaker-examples

An overview of Amazon SageMaker https://www.youtube.com/watch?v=ym7NEYEx9x4

https://medium.com/@julsimon

Thank you!

Julien Simon

Principal AI/ML Evangelist, Amazon Web Services

@julsimon