Instrumentation, Observability, and Monitoring of Machine Learning Models
About Me

- Google Engineer (2007-11)
- Cloudera’s Director of Data Science (2011-15)
- Slack’s Director of Data Engineering (2015-2017)
- Slack Engineer (now)
Hey team, hoping to have that proposal ready for the Alaska clients by 3pm today, how are we doing? I can chip in wherever needed!

I'm just about finished putting together the estimate portion of it, I could use some feedback. Here's the google doc I'm working on... docs.google.com/bin/proposal

Q3 OOH — Cost Estimate
Google Drive Document

The numbers look pretty good, I tweaked a few things, but we're good to go!

I'm just wrapping up the sketches, I'll post them here once I'm done!
The Genesis of This Talk

Visibility and Monitoring for Machine Learning Models (launchdarkly.com)

64 points by heitortsargent on Feb 20, 2018 | hide | past | web | favorite | 6 comments

mav3r1ck on Feb 20, 2018 [-]

"The tricky thing, though, is in order to get good at machine learning, you need to be able to do deploys as fast as humanly possible and repeatedly as humanly possible. Deploying a machine learning model isn’t like deploying a regular code patch or something like that, even if you have a continuous deployment system." -Josh

Sounds a lot more like a DevOps problem then a Machine Learning problem to me+. But really, in general, this is something any one doing any sort of software deployment should be doing to begin with. If I encountered a continuous deployment system that doesn't already do this, then I usually take the time to get it as close as possible. Still haven't gotten anywhere close to Netflix's level, maybe some day.

+ This is the buzziest comment I have probably ever made.

infinitone on Feb 21, 2018 [-]

Apart from the occasional pun or allegorical comment... frankly that presentation ended very abruptly with very little actual substance. Or maybe i was just expecting more?
Machine Learning In the Wild
Data Scientist (n.): Person who is better at statistics than any software engineer and better at software engineering than any statistician.

9:55 AM - 3 May 2012

1,679 Retweets  1,374 Likes
Some History

- Flickr: 10 deployments per day, 2009
- Etsy: 50+ deployments per day, 2014-2015
- Amazon Web Services: Deployments to production every 11.6 seconds
Logs via the ELK Stack
Metrics with Prometheus
Prometheus Architecture
A Word About Cardinality

Figure 7: Step-wise breakdown of executing a Scuba query.
Make Good Decisions By Avoiding Bad Decisions
The ML Test Score

Figure 1. **ML Systems Require Extensive Testing and Monitoring.** The key consideration is that unlike a manually coded system (left), ML-based system behavior is not easily specified in advance. This behavior depends on dynamic qualities of the data, and on various model configuration choices.
The map is not the territory, the word is not the thing it describes. Whenever the map is confused with the territory, a ‘semantic disturbance’ is set up in the organism. The disturbance continues until the limitation of the map is recognized.
Monitor Model Decay
Build Lots of Models
Deploy Your Models Like They Are Code*

(deck)
Stand On The Shoulders of Giants

- Ensembles
- Experiments
- Dark Tests
- Canary
- Sanity Checks
Tag All The Things
Circle of Competence

WHAT YOU KNOW

DANGER ZONE

WHAT YOU THINK YOU KNOW
Garbage In...
Linking Online and Offline Metrics
Handling Cross-Language Feature Engineering

Machine Learning

Input → Feature extraction → Classification → Output

Car
Not Car

Deep Learning

Input → Feature extraction + Classification → Output

Car
Not Car
Know Your Dependencies
Monitoring For Critical Slices
Second-Order Thinking
On Razors
http://slack.com/careers