

The Analyst's perspective on learning from data.

Jason Bedford

At Product School, Santa Monica

June 14 2017

Part 1: Two perspectives on learning from data.

Machine(s) Learning (from data)

Goal: Use statistics to predict outcomes given some relevant data.

Relies on Math (usually Linear Algebra)

Machine(s) Learning (from data)

Goal: Use statistics to predict outcomes given some relevant data.

Relies on: Math (Linear Algebra)

The Analyst (Humans Learning from data)

Goal: use statistics to quantitatively describe data

Relies on: Aggregations and Visualization

Statistics to: Predict V. Describe

Are they really just the same thing?

Statistics to: Predict V. Describe

Are they really just the same thing?

Yes, there is a lot of overlap.

But ...

Statistics to: Predict V. Describe

Are they really just the same thing?

Yes But

If you could only do one thing well what would that be?

Statistics to: Predict V. Describe

If you could only do one thing well what would that be?

Predicting: Optimizing Metrics

Describing: Conveying Understanding

Machine(s) Learning (from data)

Goal: Use statistics to predict outcomes given some relevant data.

Relies on: Math (Linear Algebra)

Prioritizes: Optimizing Metrics

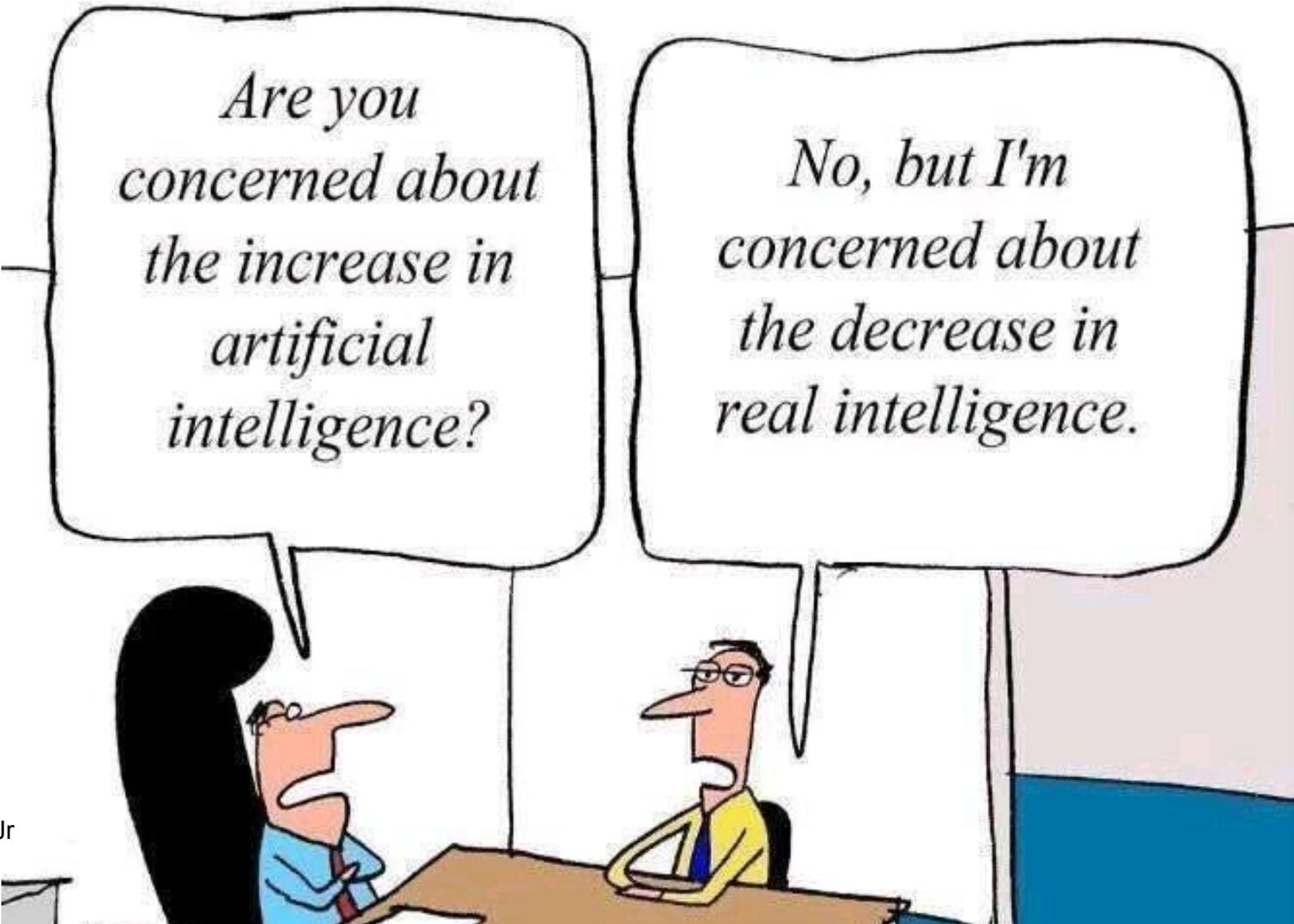
The Analyst (Humans Learning from data)

Goal: use statistics to quantitatively describe data

Relies on: Aggregations and Visualization

Prioritizes: Conveying Understanding

What is the future?



*Are you
concerned about
the increase in
artificial
intelligence?*

*No, but I'm
concerned about
the decrease in
real intelligence.*

Part 2: How an Analyst works through a project.

The data were going to be using today is from:

It's is a file hosting and syncing service.
(literally a dropbox clone)

Giving away one gigabyte of storage.

Plus, an additional 500MB for each
person referred (upto 4 people)



The data: Current storage usage for each user.

What kind of metadata would you like to have?

Looking at: Current storage usage

The data: Signup date for each user.

What kind of metadata would you like to have?

Looking at: Signup date

The data: acquisition channel for each user

Advertisements: if you click on a banner ad and signed up you would be given 2Gb of free storage instead of just 1Gb.

Referring Friends: if you referred a friend who signed up you would be given an additional 500Mb of storage for free.

The data: acquisition channel for each user

What kind of metadata would you like to have?

