

TH16 The Rise of the Machines - Machine Learning for Developers

Adam Tuliper
Principal Software Engineer
Microsoft

About me

Adam Tuliper

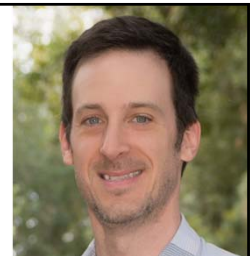
Principal Software Engineer @ Microsoft

20+ year software architect

- Cloud, gaming & mixed reality, web, bots
- Orange County Unity Meetup
- Security background
 - ASM should be taught to everyone ☺

adamt@microsoft.com | @AdamTuliper

adamtuliper.com



- Want to learn? `library("swirl")`
- `swirl()`

The need to know what could be...



Using past data to predict the future

Machine learning & predictive analytics are core capabilities that are needed throughout your business

 Churn analysis	 Social network analysis	 Recommendation engines	 Location-based tracking and services
 Vision Analytics	 Weather forecasting for business planning	 Legal discovery and document archiving	 Equipment monitoring
 Advertising analysis	 Pricing analysis	 Fraud detection	 Personalized Insurance

Machine learning - ?

Machine Learning

Subfield of computer science and statistics that deals with the construction and study of systems that can learn from data, rather than follow only explicitly programmed instructions

-Wikipedia

A computer program is said to learn from **experience** E with respect to some **task** T and some **performance measure** P , if its performance on T , as measured by P , improves with experience E .

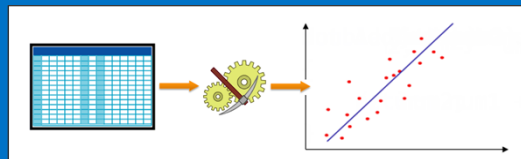
Examples

- Flash Fill
- Predictive Art
- Handwriting analysis (post office)
- Link bait suggestions
- Amazon product suggestions
- Credit card fraud detection
- Custom gestures in Kinect
- Drug discovery – advanced molecule binding



I need to predict
customer profitability...

$f()$ Age, Marital Status, Gender, Yearly Income,
Total Children, Education, Occupation,
Home Owner, Commute Distance



Some terms

- Supervised Learning
 - Here's some known data, predict from it
- Unsupervised Learning
 - Here's unknown data, categorize it and look for patterns
- Reinforced Learning
 - Robot plays nice -> Give robot cookie
 - Robot plays bad -> No cookie
- Regression
- Linear Regression

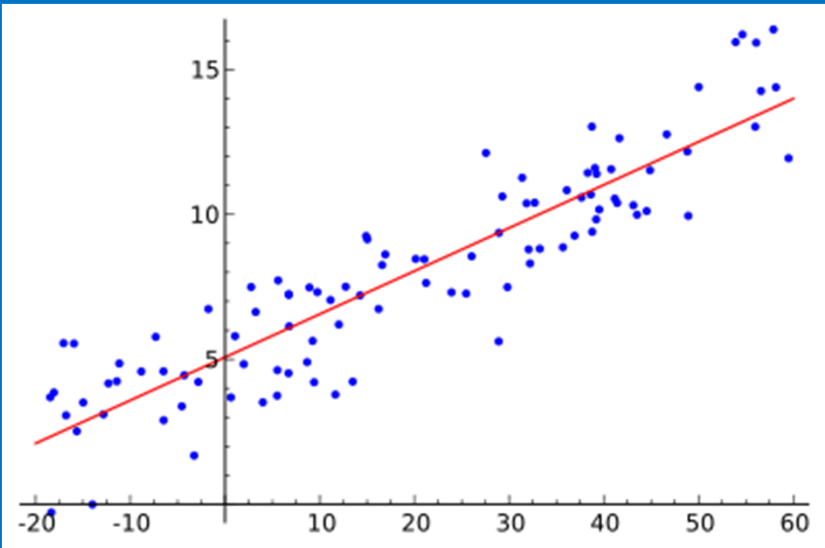
Some more terms

- Binary Classifier – Two outputs (ex true/false)
- Features – Measurable property of a phenomenon being observed
- Feature Selection – Choosing **what matters**

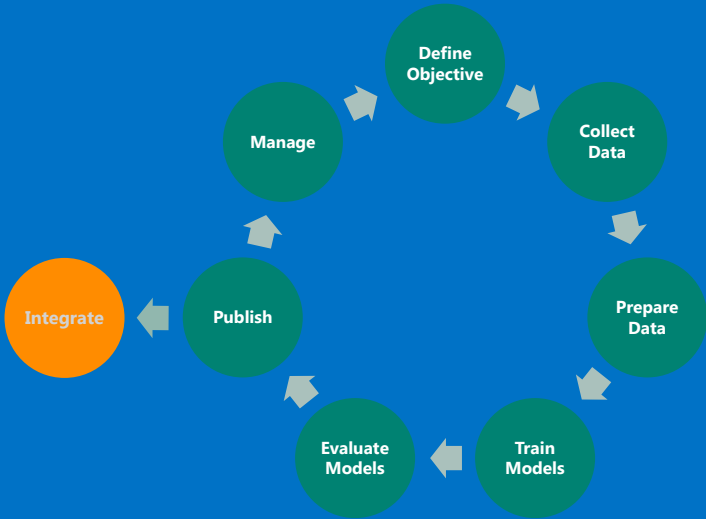
Regression

- A method for fitting a curve (not necessarily a straight line) through a set of points using some goodness-of-fit criterion. The most common type of regression is linear regression.

Linear Regression



Machine learning flow



Machine learning roles



Data scientist

A highly educated and skilled person who can solve complex data problems by employing deep expertise in scientific disciplines (mathematics, statistics or computer science)



Data professional

A skilled person who creates or maintains data systems, data solutions, or implements predictive modelling

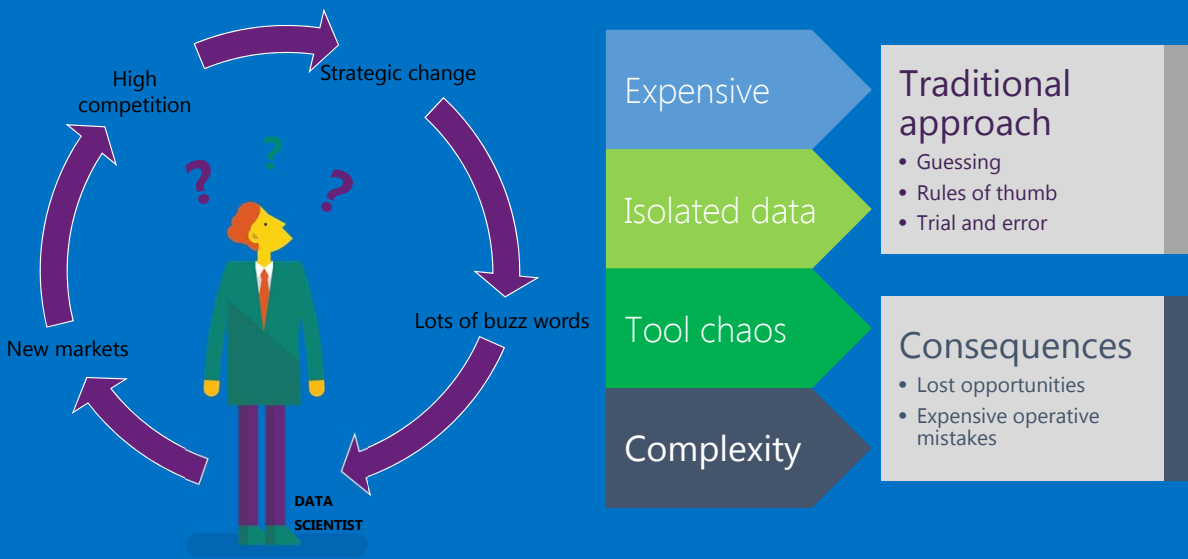
Roles: Database Administrator, Database Developer, or BI Developer



Software developer

A skilled person who designs and develops programming logic, and can apply machine learning to integrate predictive functionality into applications

Machine learning challenges



Using R for Analysis

What's R?

R

A language & environment for statistical computing

A mostly functional language

Why is it used?

Contains ML algorithms

Powerful visualization

Running R

R Studio

Microsoft procured Revolution Analytics

Microsoft R Client

Microsoft Server

Microsoft R Open

R Tools for Visual Studio

Microsoft Azure

Demo - Introduction to R



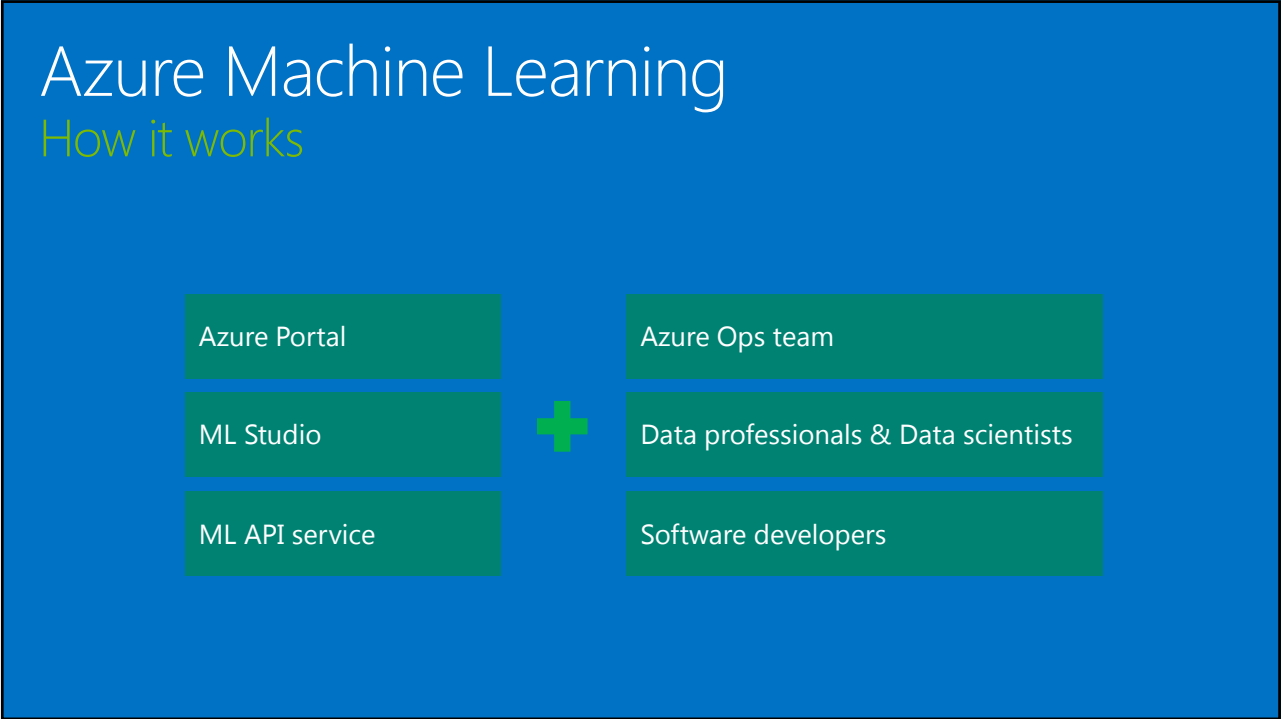
Azure Machine Learning

Azure Machine Learning

Enables powerful cloud-based predictive analytics

Professionals can easily build, deploy and share advanced analytics solutions

- Browser based, Rapid Deployment
- Connects seamlessly with other Azure data-related services, including:
 - Azure HDInsight (Big Data)
 - Azure SQL Database, and
 - Virtual Machines



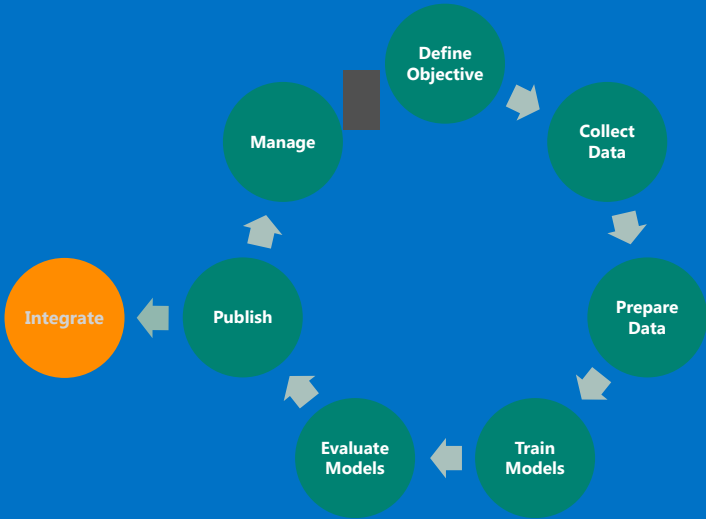
Microsoft Azure

Demo: Visual Machine Learning

A decorative graphic in the bottom left corner consists of several small, overlapping squares of different colors (blue, yellow, grey, black). Each square contains a 4-bit binary sequence (0101, 0100, 0010) arranged in a 2x2 grid. The squares are arranged in a larger, irregular grid pattern.

ML Process flow

Machine learning flow



Machine learning flow

Define
Objective

I need to predict
customer profitability...

...to deliver targeted display
advertising on the company
eCommerce web site, to:

- Present relevant product suggestions
- Increase sales and profitability

Machine learning flow

Collect
Data

Garbage in ► Garbage out ☹

- Internal sources, i.e. operational systems, data warehouse, etc.
- External sources
- Different formats, i.e. relational, multidimensional, text, map-reduce

E.g., integrate internal data to external data like weather, or market intelligence data

Machine learning flow

Prepare Data

- Transform to cleanse, reduce or reformat
- Isolate and flag abnormal data
- Appropriately **substitute** missing values
- Categorize continuous values into ranges
- **Normalize** continuous values between 0 and 1

When designing systems, give consideration to attributes that may be required as inputs for future modeling, e.g. demographic data: Birth date, gender, etc.

Machine learning flow

Train Models



Evaluate Models

- Selecting a machine learning algorithm
- Defining inputs and outputs
- Optimizing by configuring algorithm parameters

Accuracy, Reliability, Usefulness

Demo: Provisioning ML Experiment



Machine learning flow

Publish

- Transformational logic is replaced with a re-usable transformation resource
- Training logic is replaced with a trained model
- Web service inputs and outputs are added
- Module properties can be parameterized

Learn from others by discovering experiments
Contribute and showcase your experiments

Microsoft Azure

Demo: ML Web Service



Microsoft Azure

Demo - Choosing Algorithms



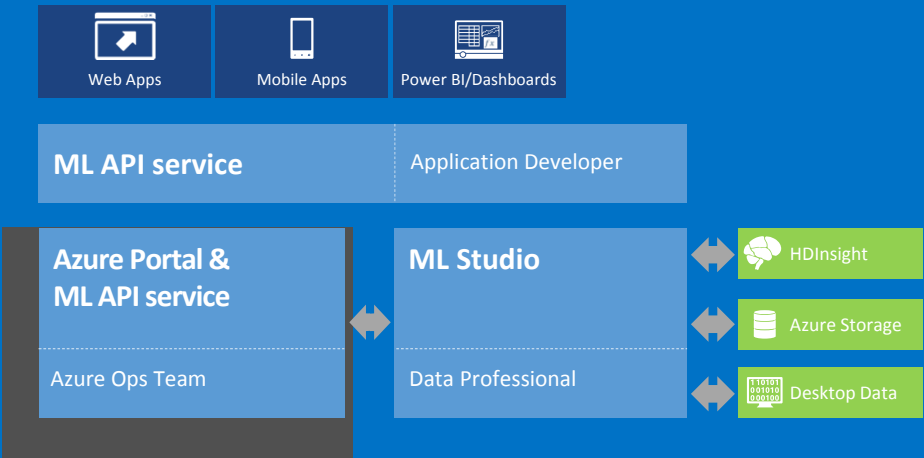
Machine learning flow

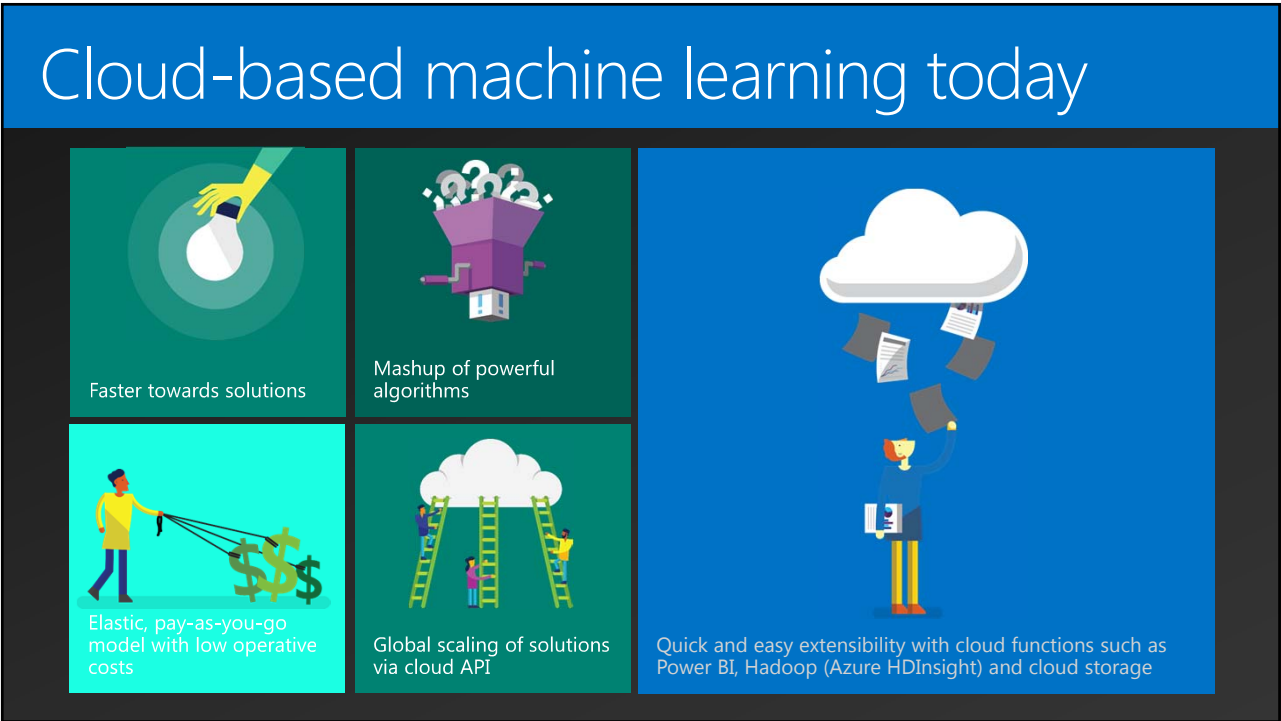
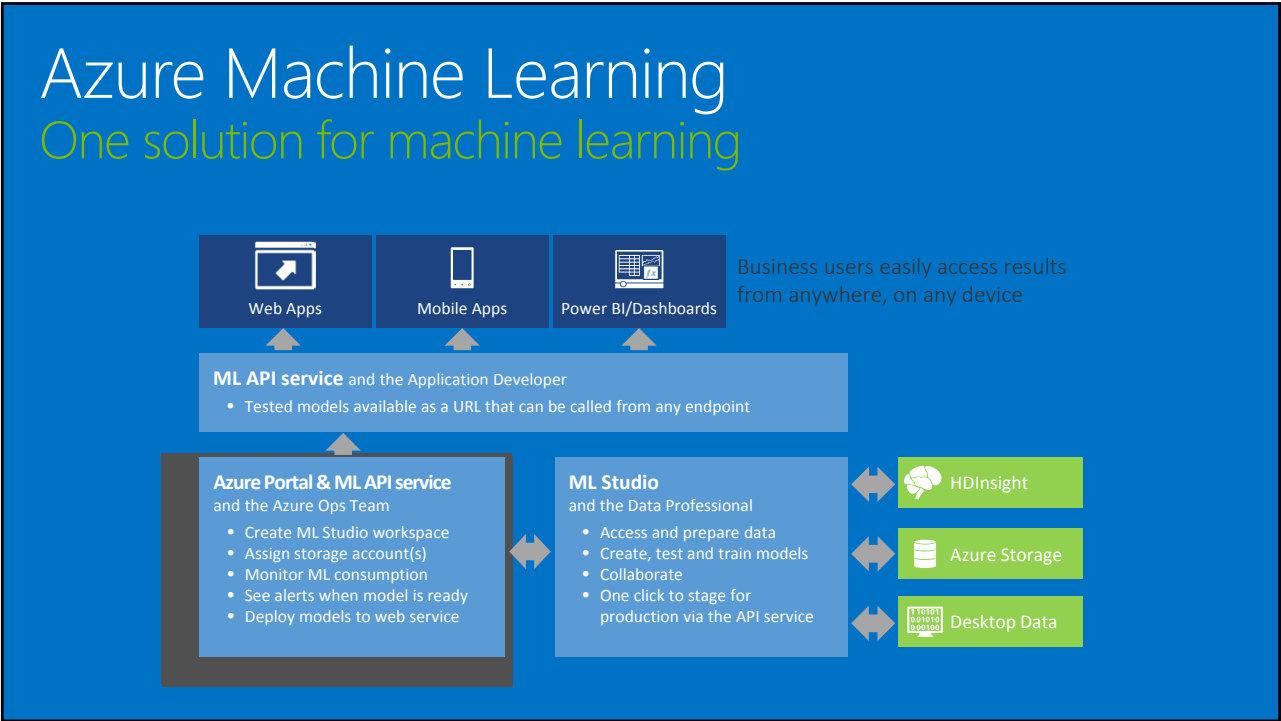


- Each web service offers two methods:
- Request/Response Service (RRS) ► Low latency, highly scalable web service
 - Batch Execution Service (BES) ► High volume, asynchronous scoring of many records

Azure Machine Learning

One solution for machine learning





Resources

<http://azure.microsoft.com/en-us/services/machine-learning>

<https://www.rstudio.com/products/rstudio/download/>

<http://azure.microsoft.com/en-us/documentation/articles/machine-learning-faq>

Q&A

