

Machine Learning

DataCrux Insights @Copyright 2018

A Few Quotes

- “A breakthrough in machine learning would be worth ten Microsofts” (Bill Gates, Chairman, Microsoft)
- “Machine learning is the next Internet” (Tony Tether, Director, DARPA)
- “Machine learning is the hot new thing” (John Hennessy, President, Stanford)
- “Web rankings today are mostly a matter of machine learning” (Prabhakar Raghavan, Dir. Research, Yahoo)
- “Machine learning is going to result in a real revolution” (Greg Papadopoulos, CTO, Sun)
- “Machine learning is today’s discontinuity” (Jerry Yang, CEO, Yahoo)

DataCrux Insights @Copyright 2018

Contents:

1. Introduction/Definition
2. Where and Why ML is used
3. Types of Learning

DataCrux Insights @Copyright 2018

So What Is Machine Learning?

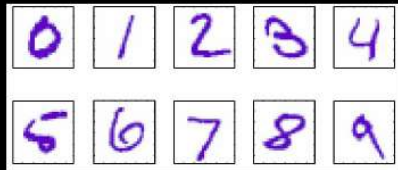
- Automating automation
- Getting computers to program themselves
- Writing software is the bottleneck
- Let the data do the work instead!

SAS definition 'Machine learning is a method of data analysis that automates analytical model building'

A computer program is said to 'learn' from experience **E** with respect to some class of tasks **T** and performance measure **P**, if its performance at tasks in **T**, as measured by **P**, improves with experience **E**

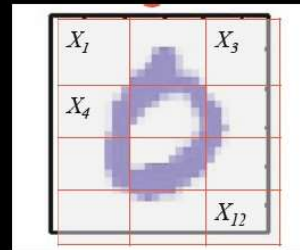
DataCrux Insights @Copyright 2018

Example:
Classification: Digit Recognition



Input (X_i): Image Features

Output (Y): Class Labels $\{y^0, y^1, \dots, y^9\}$



Features(X_i):
 Proportion of pixels in
 Each of the 12 cells
 X_i where $i=1, 2, \dots, 12$

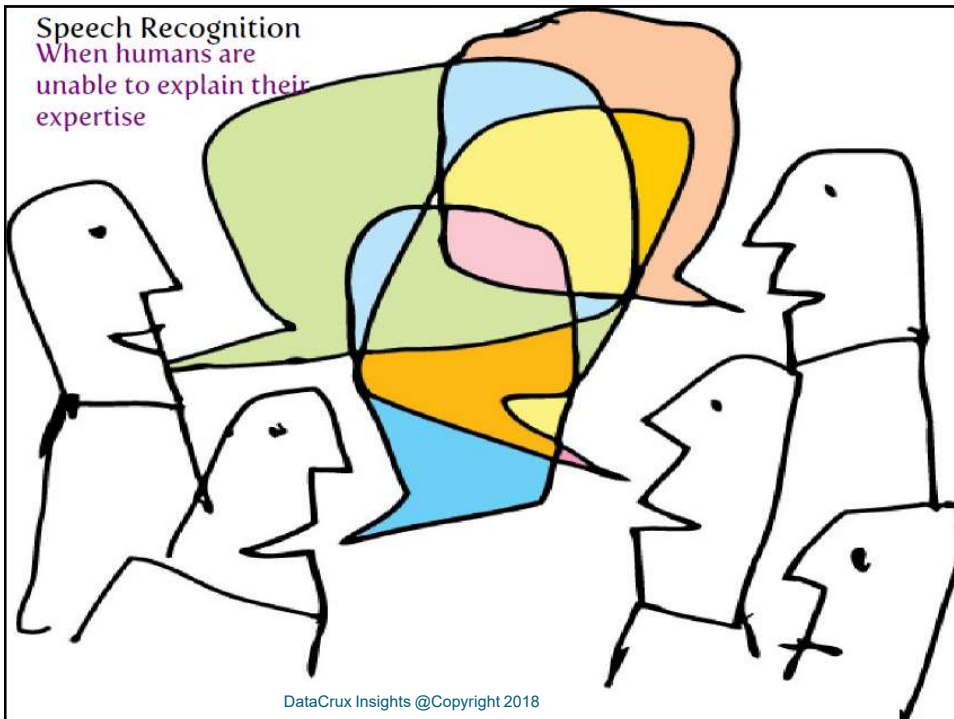
$x_i^0 = 0-10\%$
 $x_i^1 = 10-20\%$
 \dots

$Val(X_i) = 10$

DataCrux Insights @Copyright 2018

No of parameters = $10^{12} - 1$

Speech Recognition
 When humans are
 unable to explain their
 expertise



DataCrux Insights @Copyright 2018

Where **ML** is used...

DataCrux Insights @Copyright 2018



DataCrux Insights @Copyright 2018

Google applied machine learning

Secure https://www.google.co.in/search?q=applied+machine+learning&oq=applied+machine+learning&aqs=ch

About 57,30,000 results (0.49 seconds)

Scholarly articles for **applied machine learning**
 Gaussian processes for **machine learning** - Rasmussen - Cited by 9339
 Weka: Practical **machine learning** tools and techniques ... - Witten - Cited by 480
 ... : Representing model uncertainty in **deep learning** - Gal - Cited by 116

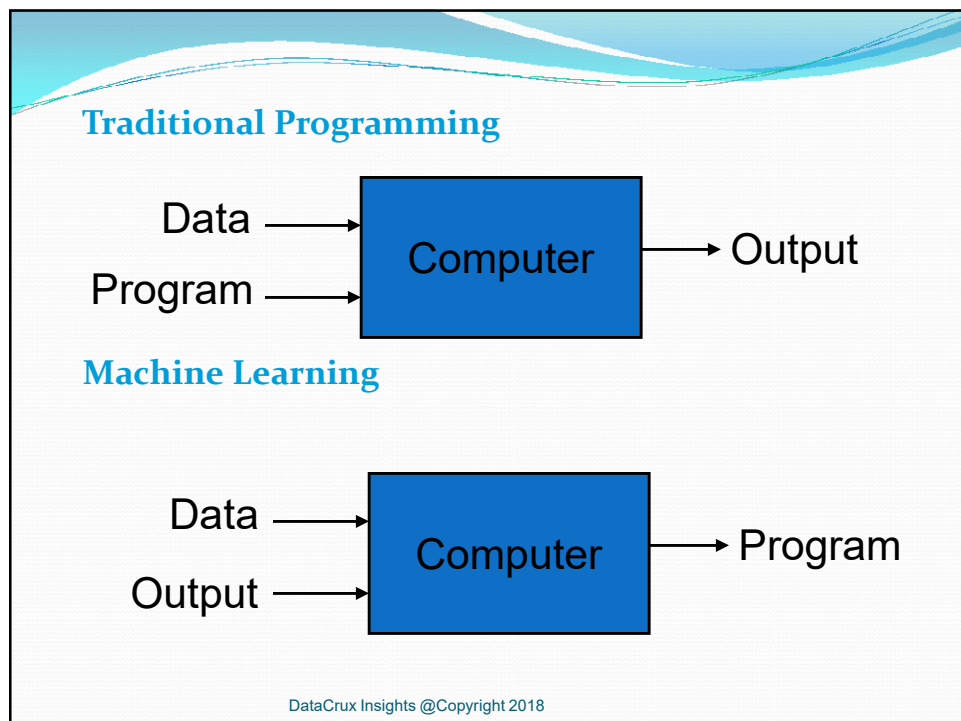
Machine learning is a rapidly growing field at the intersection of computer science and statistics that is concerned with finding patterns in data.

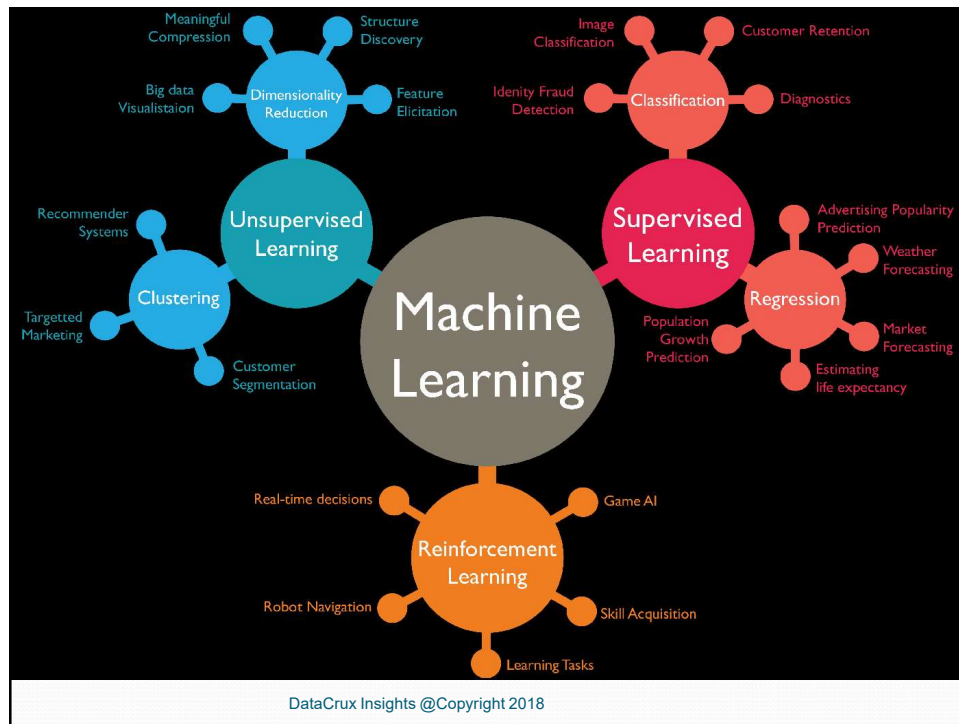
Applied Machine Learning - DataScience@Berkeley
<https://datascience.berkeley.edu/academics/curriculum/applied-machine-learning/>

Applied Machine Learning | edX
<https://www.edx.org/course/applied-machine-learning-microsoft-dat203-3x-1>
 Explore theory and practice, and work with tools like R, Python, and Azure Machine Learning to solve advanced data science problems.

Applied Machine Learning in Python | Coursera
<https://www.coursera.org/learn/python-machine-learning>
 Who is this class for: This course is part of "Applied Data Science with Python" and is intended for learners who have basic python or programming background, and want to apply statistics, machine learning, information visualization, social network analysis, and text analysis techniques to gain new insight into data.

DataCrux Insights @Copyright 2018

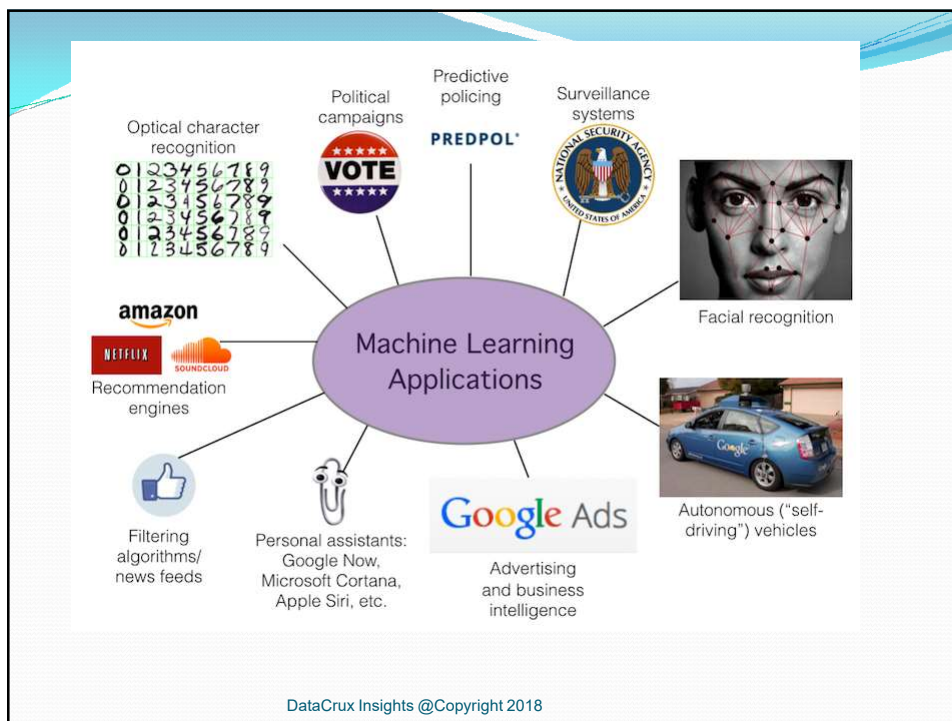




Sample Applications

- Web search
- Computational biology
- Finance
- E-commerce
- Space exploration
- Robotics
- Information extraction
- Social networks
- Debugging

DataCrux Insights @Copyright 2018



ML in a Nutshell

- Tens of thousands of machine learning algorithms
- Hundreds new every year
- Every machine learning algorithm has three components:
 - **Representation**
 - **Evaluation**
 - **Optimization**

Representation

- Decision trees
- Sets of rules / Logic programs
- Instances
- Graphical models (Bayes/Markov nets)
- Neural networks
- Support vector machines
- Model ensembles
- Etc.

DataCrux Insights @Copyright 2018

Evaluation

- Accuracy
- Precision and recall
- Squared error
- Likelihood
- Posterior probability
- Cost / Utility
- Margin
- Entropy
- K-L divergence
- Etc.

DataCrux Insights @Copyright 2018

Optimization

- Combinatorial optimization
 - E.g.: Greedy search
- Convex optimization
 - E.g.: Gradient descent
- Constrained optimization
 - E.g.: Linear programming

DataCrux Insights @Copyright 2018

