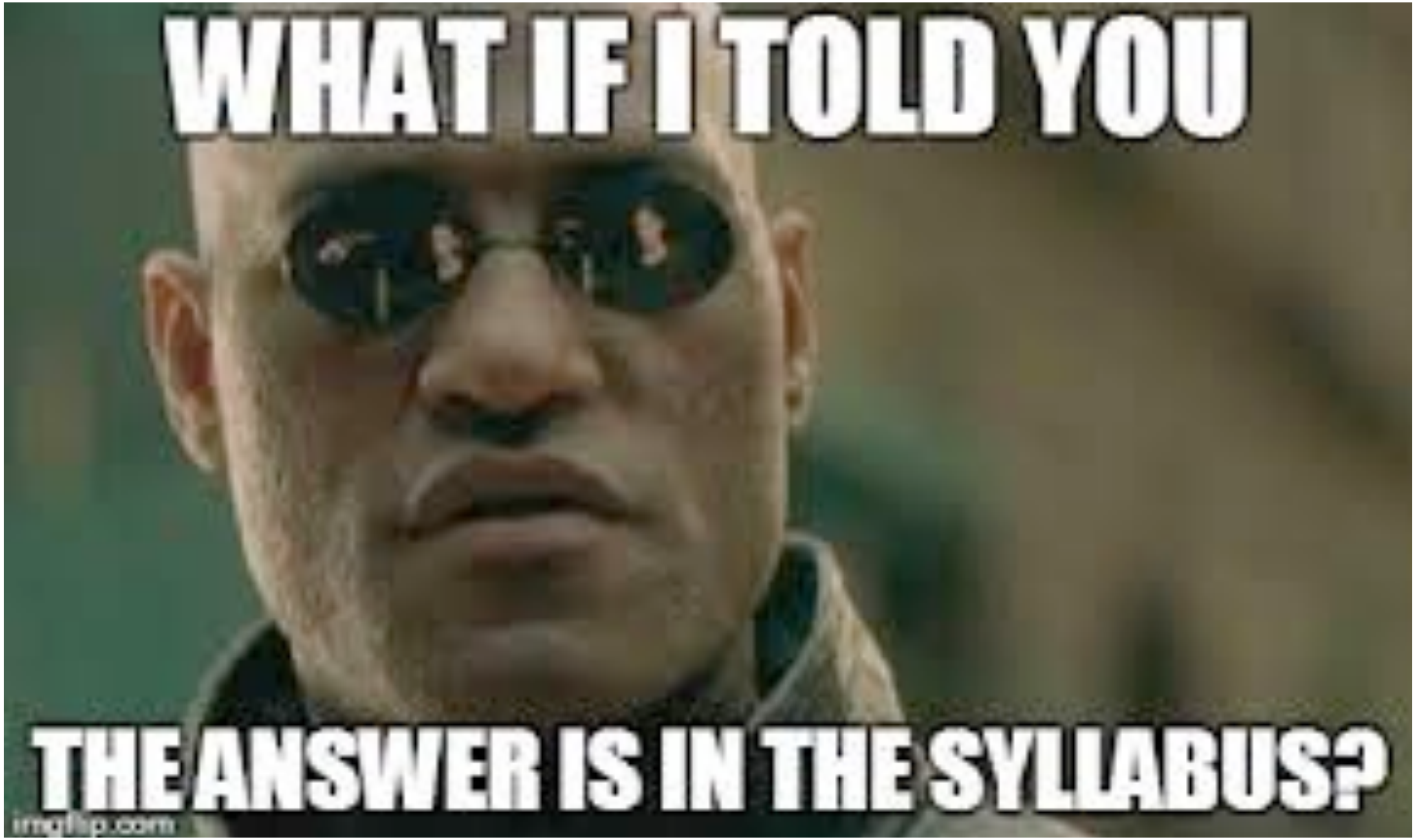
INTRODUCTION

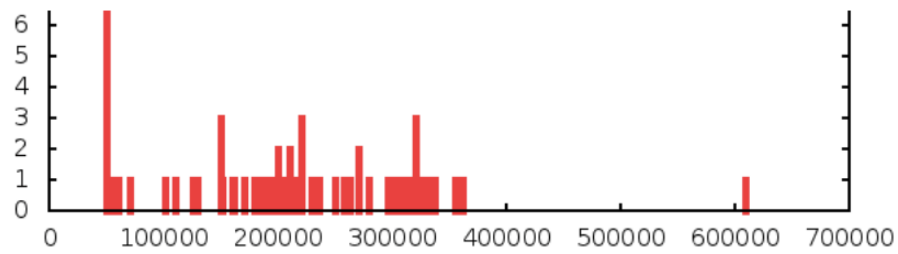
Pattern Recognition

Syllabus



Registration

- Graduate students
- 12 slots sec 2
- If filled, register as V/W only
- For undergrads, sec 21



- Signup sheet for sit-ins going around the room

Tools

- Python
- Python
- Python
- Jupyter
- Numpy
- Scipy
- Pandas
- Tensorflow, Keras

Plagiarism Policy

- You shall not show other people your code or solution
- Copying will result in a score of zero for both parties on the assignment
- Many of these algorithms have code available on the internet, do not copy paste the codes

Plagiarism vs. Cheating

What is the difference?

Courseville

- 2110597.21 (2017/1)
- https://www.mycourseville.com/?q=courseville/course/register/2110597.21_2017_1&spin=on

Password: cattern

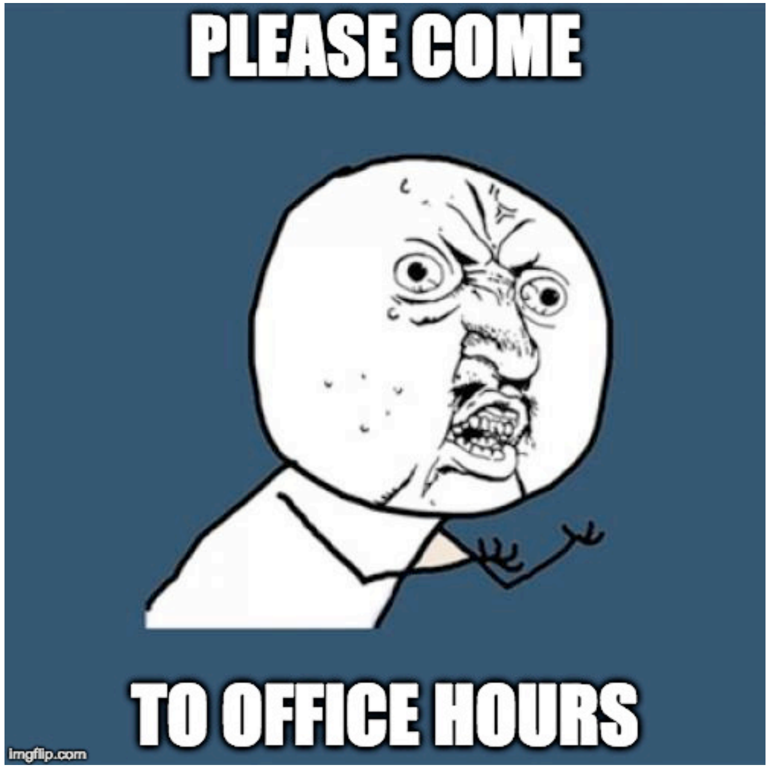
Piazza

- http://piazza.com/chula.ac.th/fall2017/2110597
- Requires chula.ac.th email

- 5 points of participation score comes from piazza

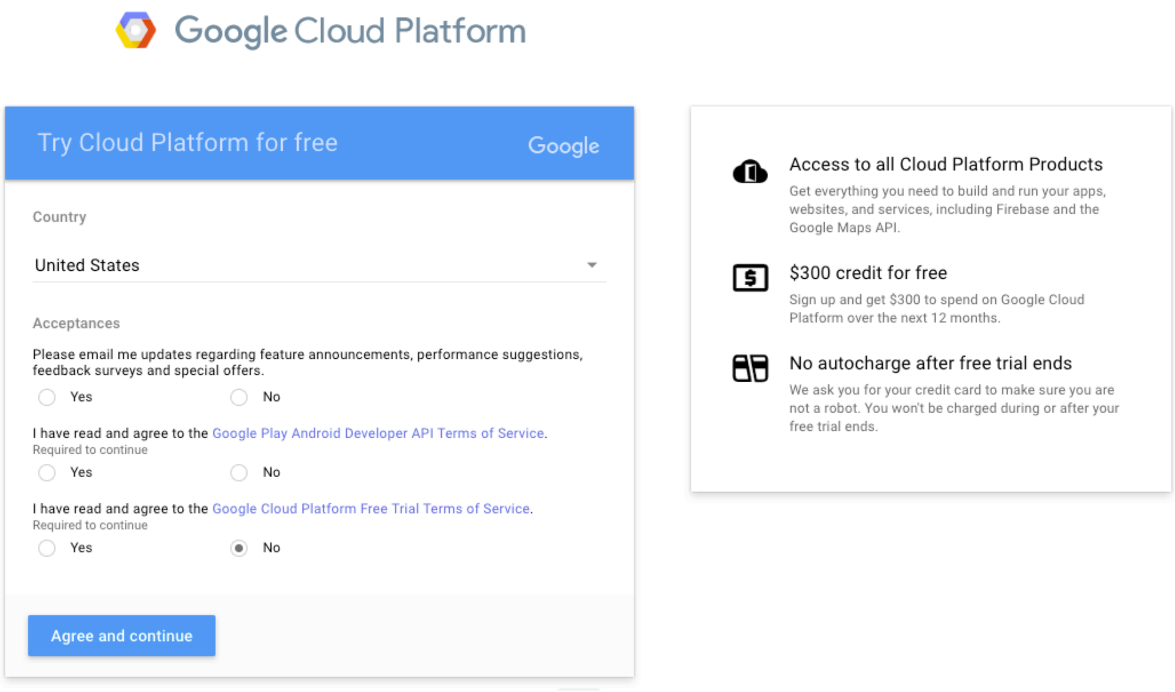
Office hours

- Thursdays 16.30-18.30 starting from Aug 31st
- Location TBA



Cloud

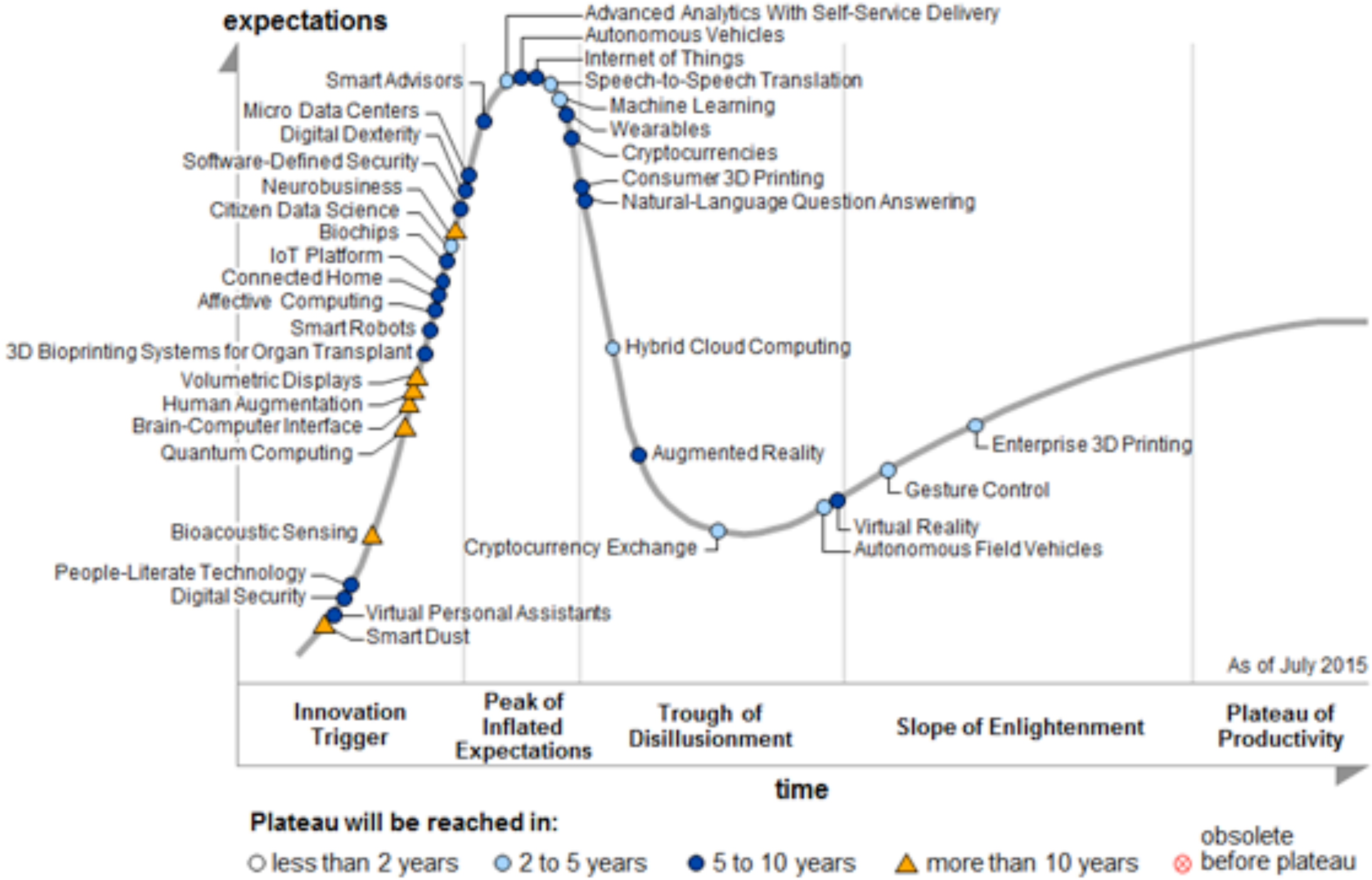
- Gcloud
- Credit card



Course project

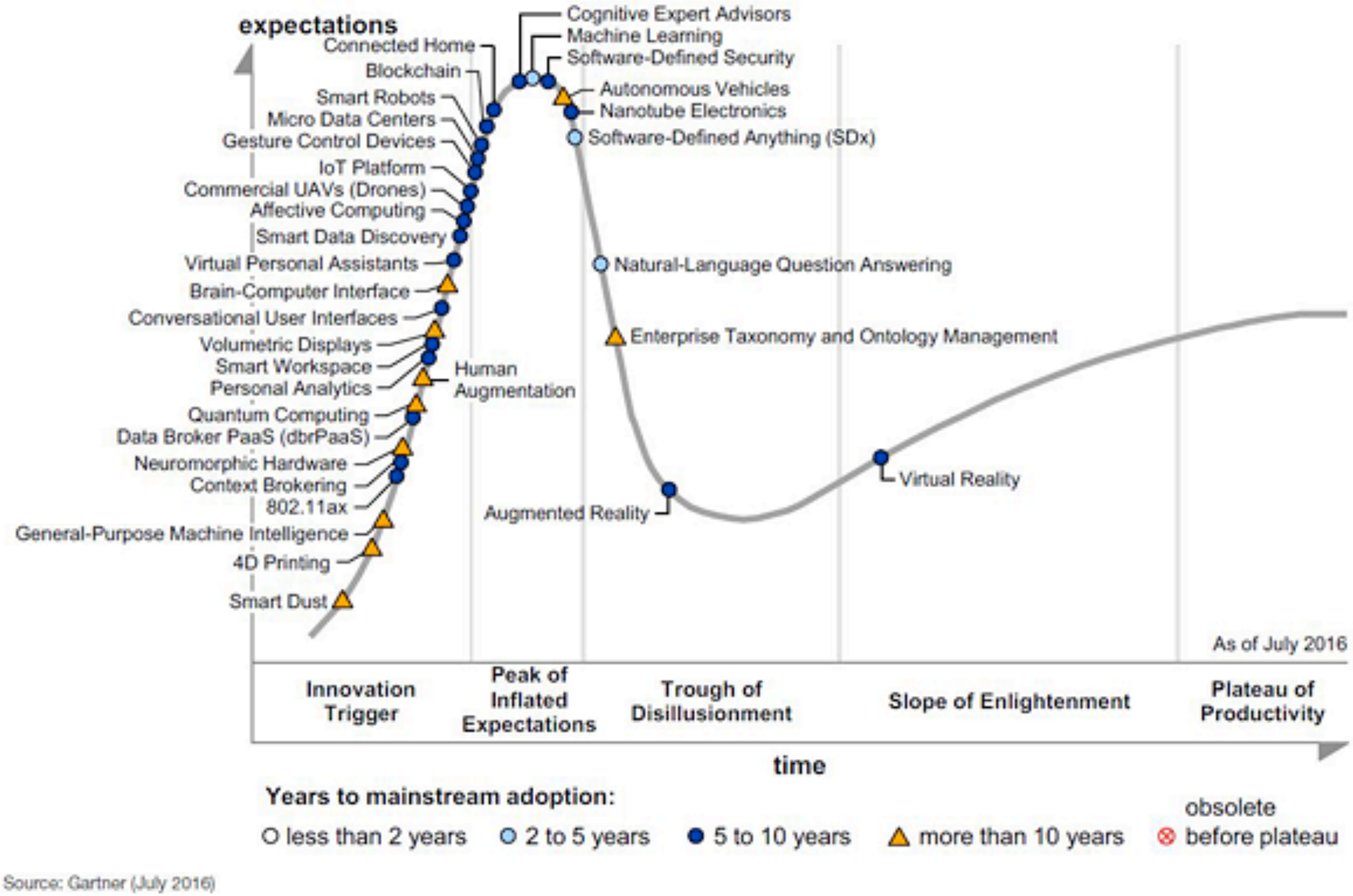
- 3-4 people (exact number TBA)
- Topic of your choice
  - Can be implementing a paper
  - Extension of a homework
  - Project for other courses with an additional machine learning component
  - Your current research (with additional scope)
  - Or work on a new application
  - Must already have existing data! No data collection!
- Topics need to be pre-approved
  - Details about the procedure TBA

The machine learning trend

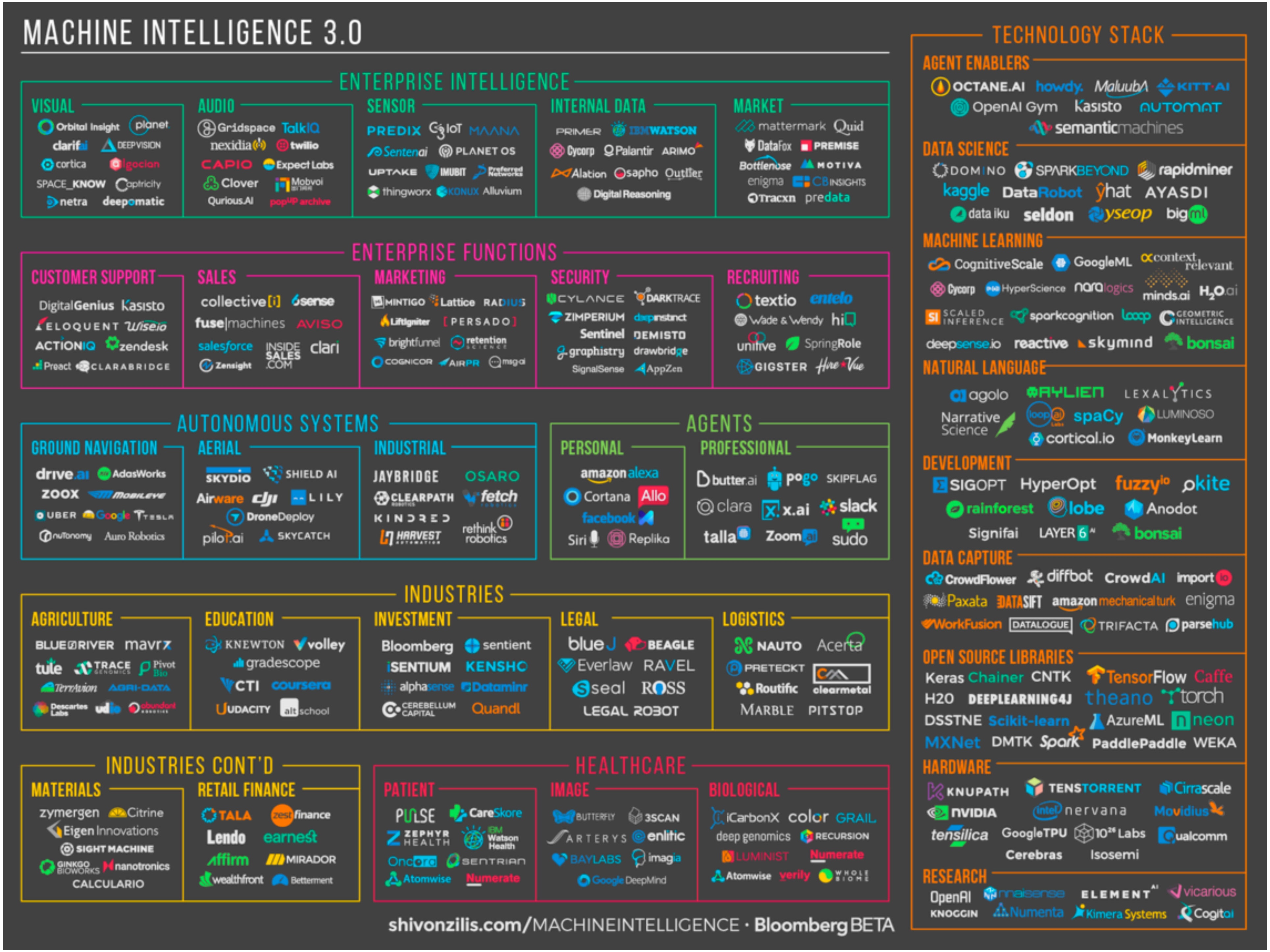


http://www.gartner.com/newsroom/id/3114217

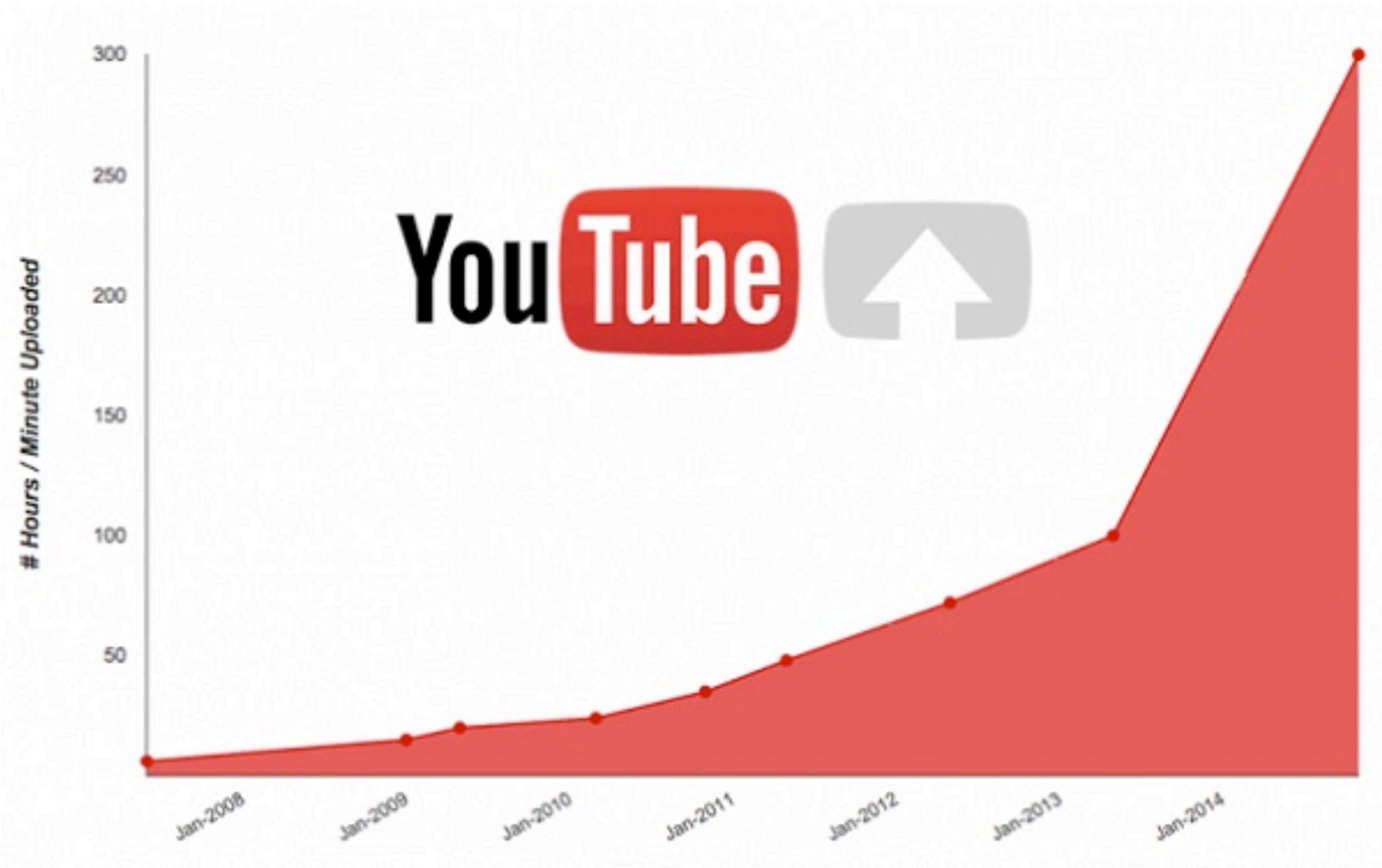
The machine learning trend



http://www.gartner.com/newsroom/id/3412017


MACHINE INTELLIGENCE 3.0
ENTERPRISE INTELLIGENCE
VISUAL
Orbital Insight
planet
clarifai
DEEP VISION
cortica
gigocion
SPACE_KNOW
Captricity
netra
deepomatic
AUDIO
Gridspace
TalkIQ
nexidia
twilio
CAPIO
Expect Labs
Clover
Mobvoi
Qurious.AI
popUP archive
SENSOR
PREDIX
GE IoT
MAANA
Sentenai
PLANET OS
UPTAKE
IMUBIT
Preferred Networks
thingworx
KONUX
Alluvium
INTERNAL DATA
PRIMER
IBM WATSON
Cycorp
Palantir
ARIMO
Alation
sapho
Outlier
Digital Reasoning
MARKET
mattermark
Quid
Datafox
PREMISE
Bottlenose
MOTIVA
enigma
CB INSIGHTS
Tracxn
predata
ENTERPRISE FUNCTIONS
CUSTOMER SUPPORT
DigitalGenius
Kasisto
ELOQUENT
Wiseio
ACTIONIQ
zendesk
Preact
CLARABRIDGE
SALES
collective[i]
6sense
fuse|machines
AVISO
salesforce
INSIDE SALES .COM
clari
Zensight
MARKETING
MINTIGO
Lattice
RADIUS
Liftigniter
[PERSADO]
brightfunnel
retention science
COGNICOR
AIRPR
msg.ai
SECURITY
CYLANCE
DARKTRACE
ZIMPERIUM
deep instinct
Sentinel
DEMISTO
graphistry
drawbridge
SignalSense
AppZen
RECRUITING
textio
entelo
Wade & Wendy
hiQ
unitive
SpringRole
GIGSTER
Hire Vue
AUTONOMOUS SYSTEMS
GROUND NAVIGATION
drive.ai
AdasWorks
ZOOX
MOBILEYE
UBER
Google
TESLA
nuTonomy
Auro Robotics
AERIAL
SKYDIO
SHIELD AI
Airware
dji
LILY
DroneDeploy
pilot.ai
SKYCATCH
INDUSTRIAL
JAYBRIDGE
OSARO
CLEARPATH ROBOTICS
fetch robotics
KINDRED
HARVEST AUTOMATION
rethink robotics
AGENTS
PERSONAL
amazon alexa
Cortana
Allo
facebook M
Siri
Replika
PROFESSIONAL
butter.ai
pogo
SKIPFLAG
clara
x.ai
slack
talla
Zoom.ai
sudo
INDUSTRIES
AGRICULTURE
BLUE RIVER
mavrx
tule
TRACE GENOMICS
Pivot Bio
TerrAvion
AGRI-DATA
Descartes Labs
udio
abundant robotics
EDUCATION
KNEWTON
volley
gradescope
CTI
coursera
UDACITY
altschool
INVESTMENT
Bloomberg
sentient
iSENTIUM
KENSHO
alphasense
Dataminr
CEREBELLUM CAPITAL
Quandl
LEGAL
blue J
BEAGLE
Everlaw
RAVEL
seal
ROSS
LEGAL ROBOT
LOGISTICS
NAUTO
Acerta
PRETECKT
Routific
clearmetal
MARBLE
PITSTOP
INDUSTRIES CONT'D
MATERIALS
zymergen
Citrine
Eigen Innovations
SIGHT MACHINE
GINKGO BIOWORKS
nanotronics
CALCULARIO
RETAIL FINANCE
TALA
zestfinance
Lendo
earnest
affirm
MIRADOR
wealthfront
Betterment
HEALTHCARE
PATIENT
PULSE
CareSkore
ZEPHYR HEALTH
IBM Watson Health
Oncora
SENTRIAN
Atomwise
Numerate
IMAGE
BUTTERFLY
3SCAN
ARTERYS
enlitic
BAYLABS
imagia
Google DeepMind
BIOLOGICAL
iCarbonX
color
GRAIL
deep genomics
RECURSION
LUMINIST
Numerate
Atomwise
verily
WHOLE BIOME
TECHNOLOGY STACK
AGENT ENABLERS
OCTANE.AI
howdy.
Maluuba
KITT.AI
OpenAI Gym
Kasisto
AUTOMAT
semantic machines
DATA SCIENCE
DOMINO
SPARKBEYOND
rapidminer
kaggle
DataRobot
yhat
AYASDI
data iku
seldon
yseop
bigml
MACHINE LEARNING
CognitiveScale
GoogleML
context relevant
Cycorp
HyperScience
nara logics
minds.ai
H2O.ai
SCALED INFERENCE
sparkcognition
Loop
GEOMETRIC INTELLIGENCE
deepsense.io
reactive
skymind
bonsai
NATURAL LANGUAGE
agolo
AYLIEN
LEXALYTICS
Narrative Science
loop ai labs
spaCy
LUMINOSO
cortical.io
MonkeyLearn
DEVELOPMENT
SIGOPT
HyperOpt
fuzzy.io
kite
rainforest
lobe
Anodot
Signifai
LAYER 6
bonsai
DATA CAPTURE
CrowdFlower
diffbot
CrowdAI
import.io
Paxata
DATASIFT
amazon mechanical turk
enigma
WorkFusion
DATALOGUE
TRIFACTA
parsehub
OPEN SOURCE LIBRARIES
Keras
Chainer
CNTK
TensorFlow
Caffe
H2O
DEEPLEARNING4J
theano
torch
DSSTNE
Scikit-learn
AzureML
neon
MXNet
DMTK
Spark
PaddlePaddle
WEKA
HARDWARE
KNUPATH
TENSTORRENT
Cirrascale
NVIDIA
intel nervana
Movidius
tensilica
GoogleTPU
10^{26} Labs
Qualcomm
Cerebras
Isosemi
RESEARCH
OpenAI
maluuba
ELEMENT AI
vicarious
KNOGGIN
Numenta
Kimera Systems
Cogitai
shivonzilis.com/MACHINEINTELLIGENCE · Bloomberg BETA

The data era

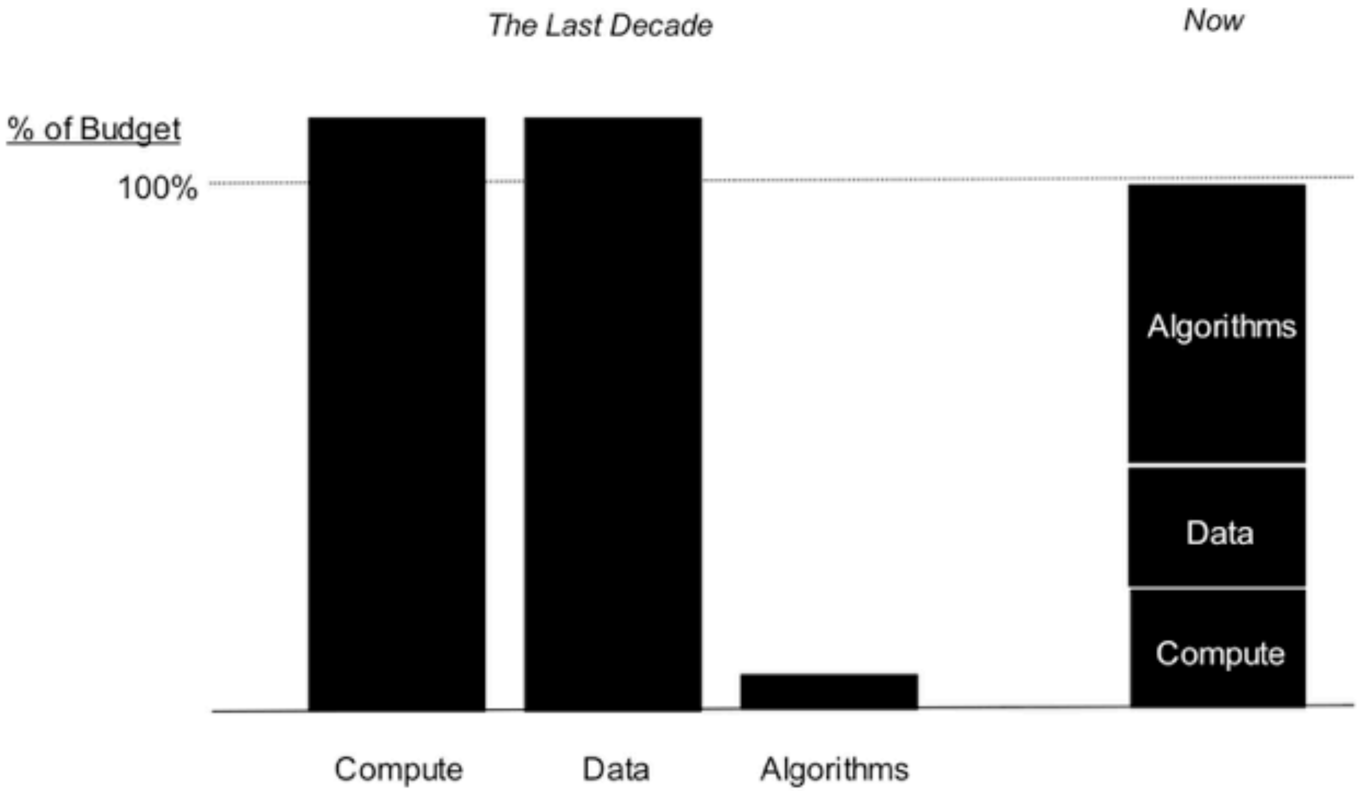


2017 numbers = 400 hours/min

http://www.tubefilter.com/2014/12/01/youtube-300-hours-video-per-minute/

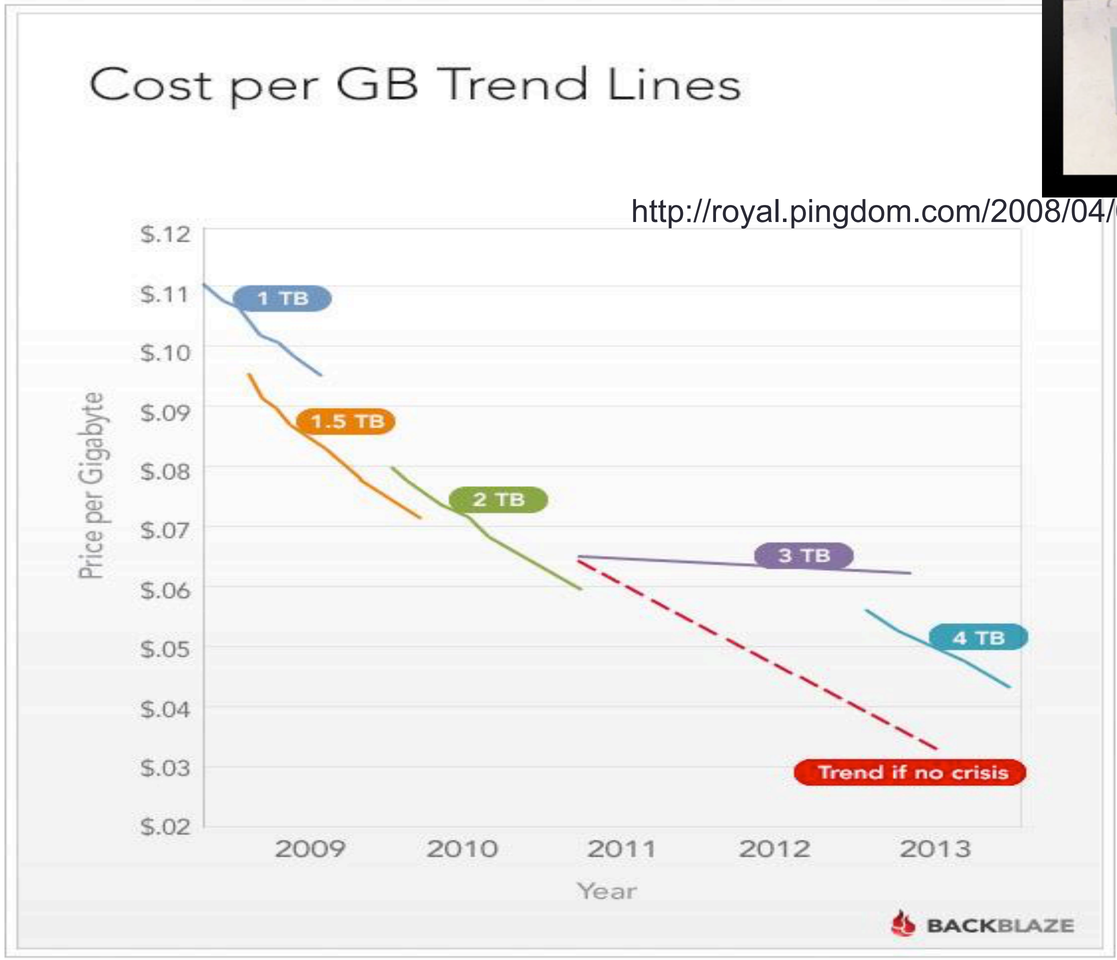
Factors for ML

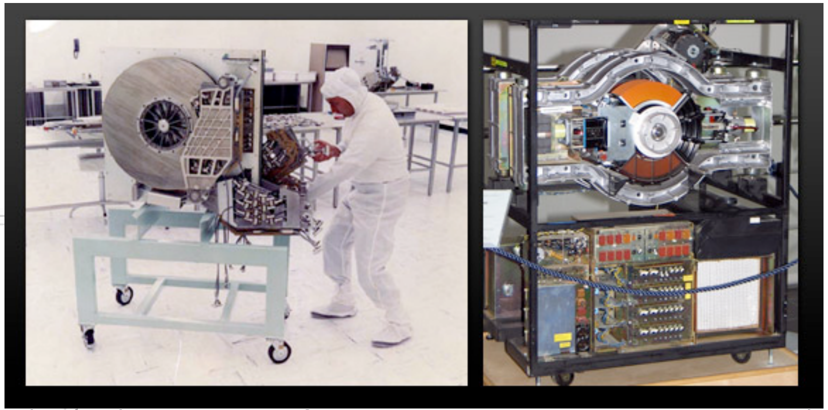
- Data
- Compute



http://www.kdnuggets.com/2017/06/practical-guide-machine-learning-understand-differentiate-apply.html

The cost of storage



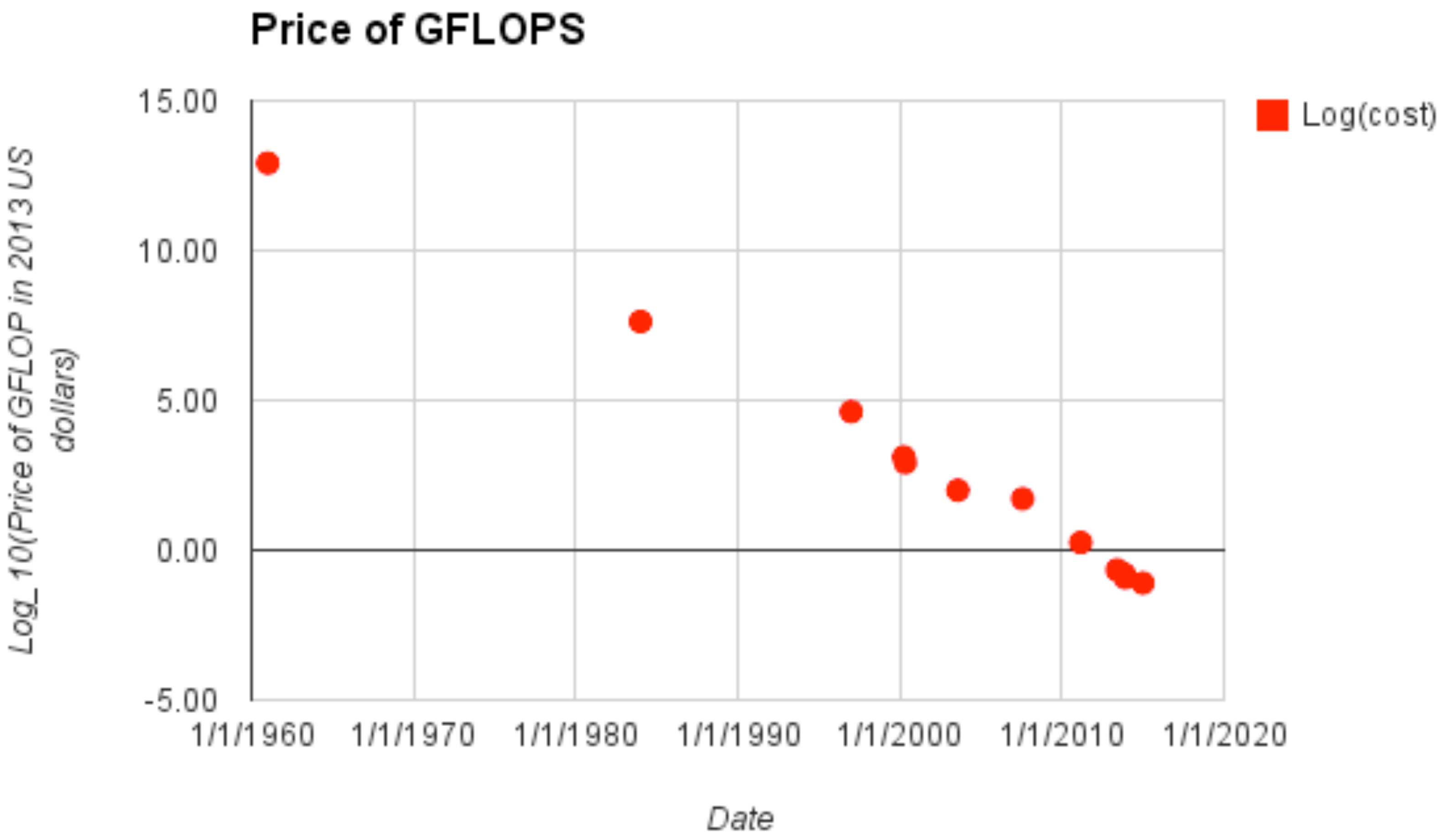


http://royal.pingdom.com/2008/04/08/the-history-of-computer-data-storage-in-pictures/

1980 250MB hard disk drive
250 kg 100k USD (300k USD in today's dollar)

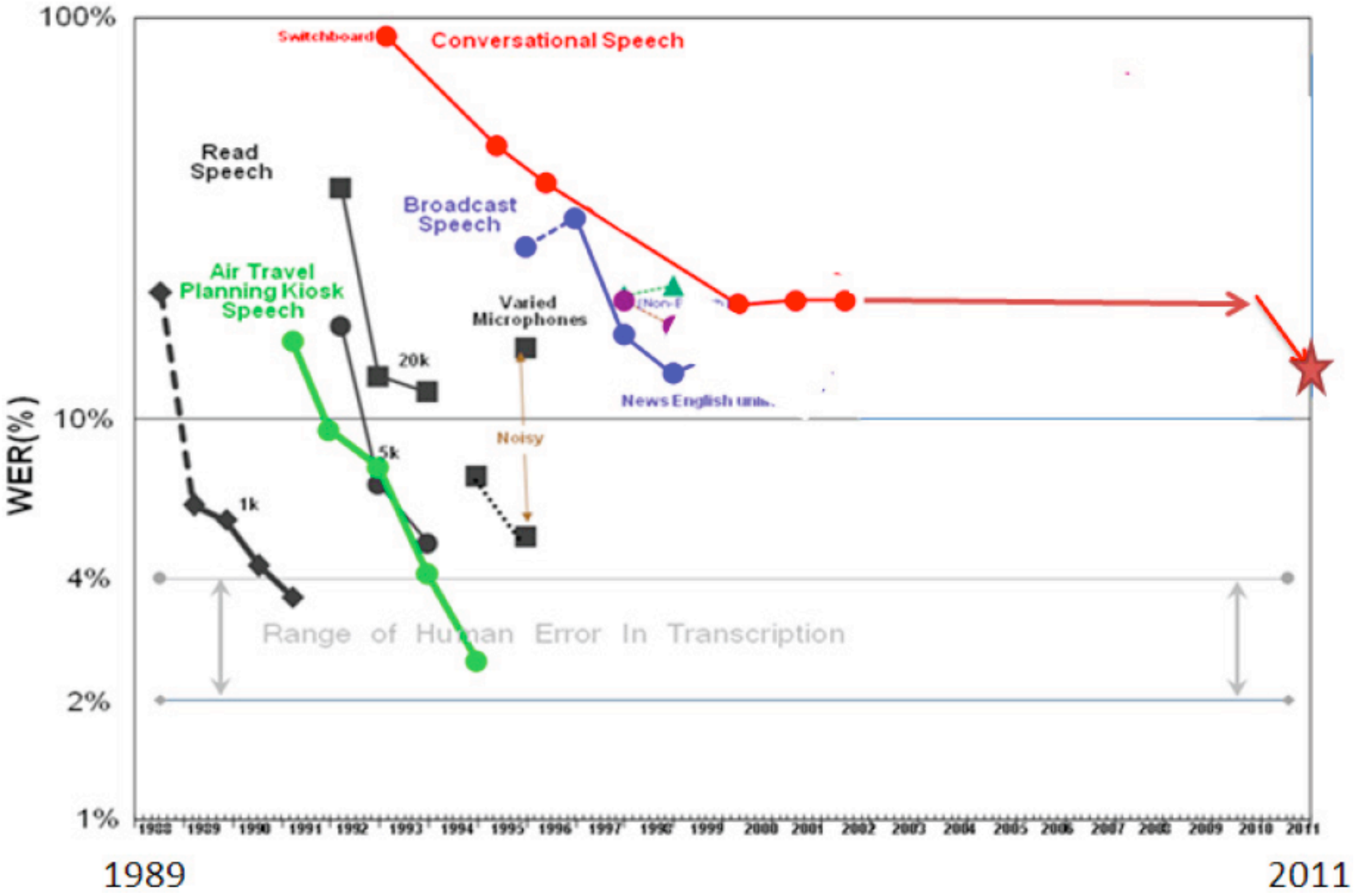
https://www.backblaze.com/blog/farming-hard-drives-2-years-and-1m-later/

The cost of compute



http://aiimpacts.org/trends-in-the-cost-of-computing/

Hitting the sweet spot on performance



Hitting the sweet spot in performance

PRESS RELEASE
OCTOBER 4, 2011

Apple Launches iPhone 4S, iOS 5 & iCloud

iPhone 4S Features Dual-Core A5 Chip, All New Camera, Full 1080p HD Video Recording & Introduces Siri

CUPERTINO, California—October 4, 2011—Apple® today announced iPhone® 4S, the most amazing iPhone yet, packed with incredible new features including Apple's dual-core A5 chip for blazing fast performance and stunning graphics; an all new camera with advanced optics; full 1080p HD resolution video recording; and Siri™, an intelligent assistant that helps you get things done just by asking. With the launch of iPhone 4S

Now time for a video

https://www.youtube.com/watch?v=wiOopO9jTZw

Darren Cunningham @dcunni · 6h

Zuckerberg blasts @elonmusk warnings against artificial intelligence as 'pretty irresponsible' bizjournals.com/sanjose/news/2… @svbizjournal #ai

Facebook CEO Mark Zuckerberg blasts Tesla CEO Elon Musk's warn…

"People who are naysayers and try to drum up these doomsday scenarios — I just, I don't understand it," the Facebook CEO said. "It's really negative

bizjournals.com

30 296 566

Elon Musk
@elonmusk

Following

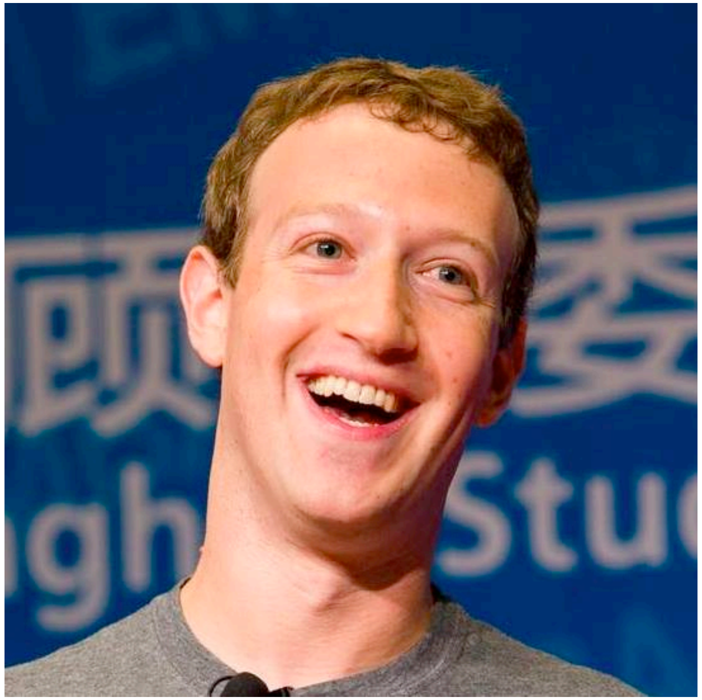
Replying to @dcunni @SVbizjournal

I've talked to Mark about this. His understanding of the subject is limited.

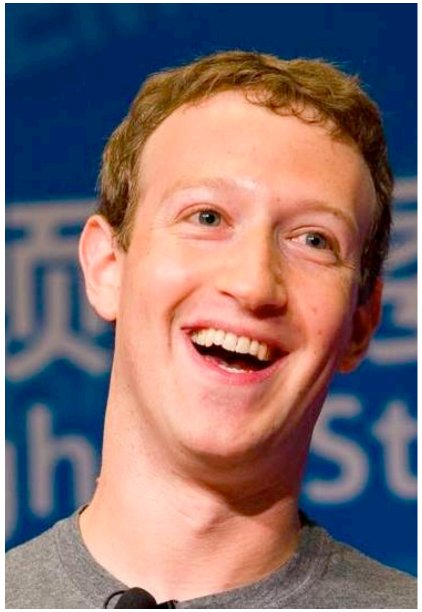
8:07 AM - 25 Jul 2017

- “If I were to guess like what our biggest existential threat is, it’s probably that. So we need to be very careful with the artificial intelligence. There should be some regulatory oversight maybe at the national and international level, just to make sure that we don’t do something very foolish.”

- “I think people who are naysayers and try to drum up these doomsday scenarios — I just, I don’t understand it. It’s really negative and in some ways I actually think it is pretty irresponsible”



Poll



What is Pattern Recognition?

- "Pattern recognition is a branch of machine learning that focuses on the recognition of patterns and regularities in data, <u>although it is in some cases considered to be nearly synonymous with machine learning</u>."

  wikipedia

- What about
  - Data mining
  - Knowledge Discovery in Databases (KDD)
  - Statistics

ML vs PR vs DM vs KDD

- "The short answer is: None. They are … concerned with the same question: how do we learn from data?"

Larry Wasserman – CMU Professor

- Nearly identical tools and subject matter

History

- Pattern Recognition started from the engineering community (mainly Electrical Engineering and Computer Vision)
- Machine learning comes out of AI and mostly considered a Computer Science subject
- Data mining starts from the database community

Different community viewpoints

- A screw looking for a screw driver
- A screw driver looking for a screw

Different applications

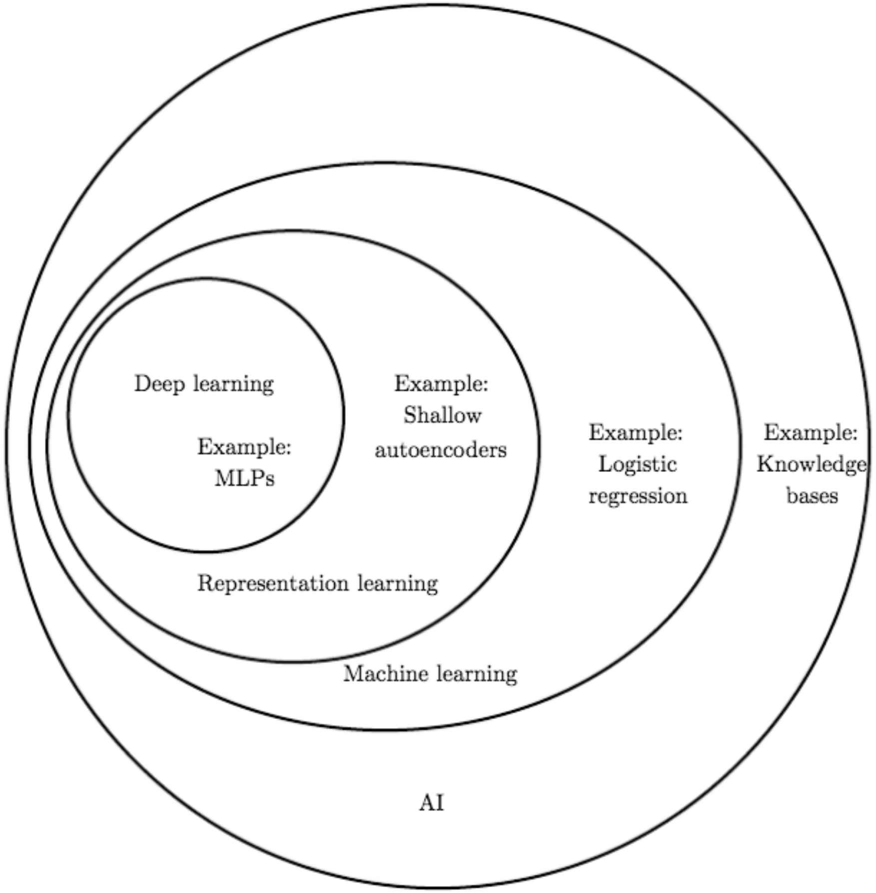
Different tools

The Screwdriver and the Screw

DM PR ML AI

Distinguishing things

- DM – Data warehouse, ETL
- AI – Artificial General Intelligence
- PR – Signal processing (feature engineering)



http://www.deeplearningbook.org/

Different terminologies

http://statweb.stanford.edu/~tibs/stat315a/glossary.pdf

Machine learning	Statistics
network, graphs	model
weights	parameters
learning	fitting
generalization	test set performance
supervised learning	regression/classification
unsupervised learning	density estimation, clustering
large grant = $1,000,000	large grant= $50,000

Merging communities and fields

- With the advent of Deep learning the fields are merging and the differences are becoming unclear

How do we learn from data?

- The typical workflow

Real world observations

Squire Trelawney, Dr. Livesey, and the rest of those gentlemen having asked me to write down the whole particulars about Treasure Island, from the beginning to the end, keeping nothing back but the bearings of the island, and that only because there is still treasure not yet lifted, I take up my pen in the year of grace 17— and go back to the time when my father kept the Admiral Benbow inn and the brown old seaman with the sabre cut first took up his lodging under our roof.

I remember him as if it were yesterday, as he came plodding to the inn door, his sea-chest following behind him in a hand-barrow; a tall, strong, heavy, nut brown man, his tarry pigtail falling over the shoulder of his soiled blue coat, his hands ragged and scarred, with black, broken nails, and the sabre cut across one cheek, a dirty, livid white. I remember him looking round the cover and whistling to himself as he did so, and then breaking out in that old sea-song that he sang so often afterwards:

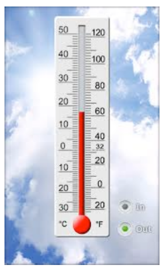
'*Fifteen men on the dead man's chest–Yo-ho-ho, and a bottle of rum!*' in the high, old tottering voice that seemed to have been tuned and broken at the capstan bars. Then he rapped on the door with a bit of stick like a handspike that he carried, and when my father appeared, called roughly for a glass of rum. This, when it was brought to him, he drank slowly, like a connoisseur, lingering on the taste and still looking about him at the cliffs and up at our signboard.

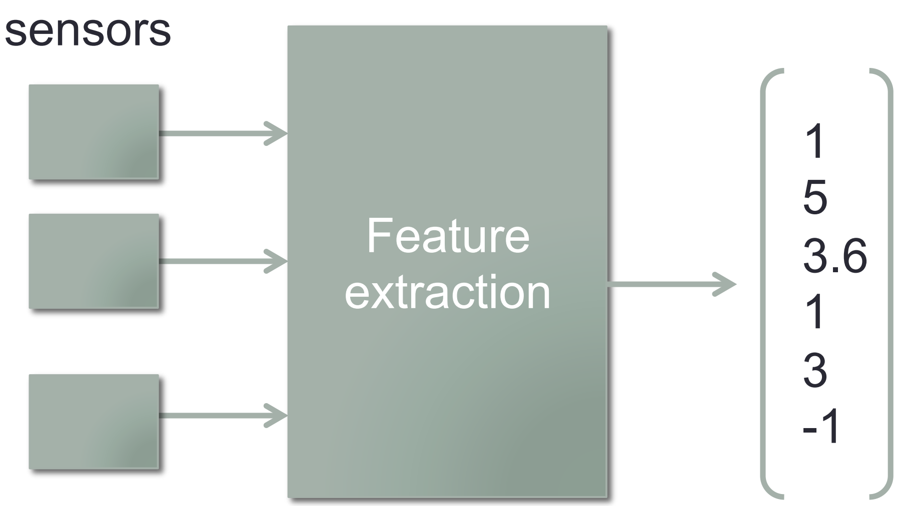
'This is a handy cove,' says he at length; 'and a pleasant sittyated grog-shop. Much company, mate?'

My father told him no, very little company, the more was the pity.

'Well, then,' said he, 'this is the berth for me. Here you, matey,' he cried to the man who trundled the barrow; 'bring up alongside and help up my chest. I'll stay here a bit,' he continued. 'I'm a plain man; rum and bacon and eggs is what I want, and that head up there for to watch ships off. What you mought call me? You mought call me captain. Oh, I see what you're at - there'; and he threw down three or four gold pieces on the threshold. 'You can tell me when I've worked through that,' says he, looking as fierce as a commander.

And indeed bad as his clothes were and coarsely as he spoke, he had none of the appearance of a





Feature vector
x

How do we learn from data?

1
5
3.6
1
3
-1

Training set

Learning algorithm

Model

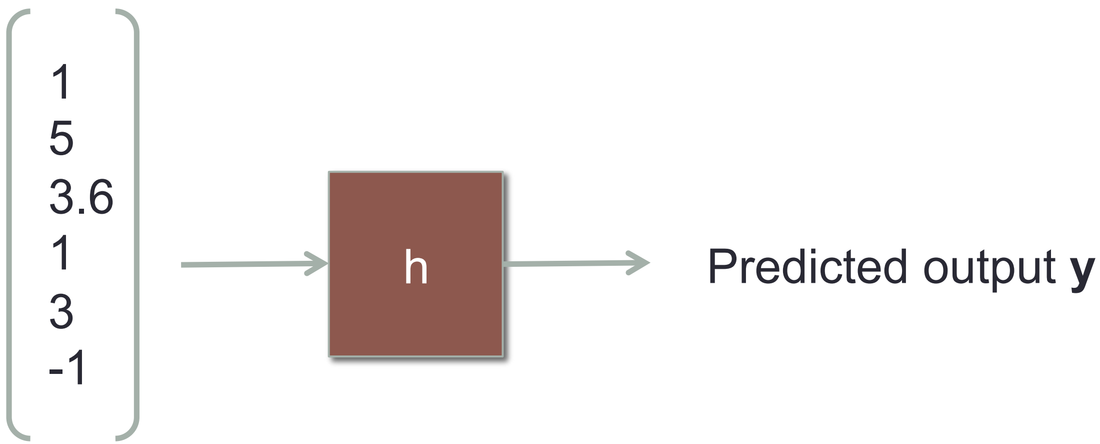
h

Desired output **y**

Training phase

How do we learn from data?

New input **X**



Testing phase

A task

The raw inputs and the desired output defines a machine learning task

Squire Trelawney, Dr. Livesey, and the rest of these gentlemen having asked me to write down the whole particulars about Treasure Island, from the beginning to the end, keeping nothing back but the bearings of the island, and that only because there is still treasure not yet lifted, I take up my pen in the year of grace 17— and go back to the time when my father kept the Admiral Benbow inn and the brown old seaman with the sabre cut first took up his lodging under our roof.

I remember him as if it were yesterday, as he came plodding to the inn door, his sea-chest following behind him in a hand-barrow; a tall, strong, heavy, nut brown man, his tarry pigtail falling over the shoulder of his soiled blue coat, his hands ragged and scarred, with black, broken nails, and the sabre cut across one cheek, a dirty, livid white. I remember him looking round the cover and whistling to himself as he did so, and then breaking out in that old sea-song that he sang so often afterwards:

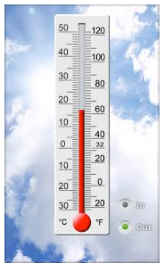
'Fifteen men on the dead man's chest–Yo-ho-ho, and a bottle of rum!' In the high, old tottering voice that seemed to have been tuned and broken at the capstan bars. Then he rapped on the door with a bit of stick like a handspike that he carried, and when my father appeared, called roughly for a glass of rum. This, when it was brought to him, he drank slowly, like a connoisseur, lingering on the taste and still looking about him at the cliffs and up at our signboard.

'This is a handy cove,' says he at length; 'and a pleasant sittyated grog-shop. Much company, mate?'

My father told him no, very little company, the more was the pity.

'Well, then,' said he, 'this is the berth for me. Here you, matey,' he cried to the man who trundled the barrow; 'bring up alongside and help up my chest. I'll stay here a bit,' he continued. 'I'm a plain man; rum and bacon and eggs is what I want, and that head up there for to watch ships off. What you mought call me? You mought call me captain. Oh, I see what you're at - there'; and he threw down three or four gold pieces on the threshold. 'You can tell me when I've worked through that,' says he, looking as fierce as a commander.

And indeed bad as his clothes were and coarsely as he spoke, he had none of the appearance of a



data1

data2

data3

Magic

Predicted output **y**

Predicting After You stock price with CCTV image, facebook posts, and daily temperature

Key concepts

- Feature extraction
- Evaluation

Feature extraction

- The process of extracting meaningful information related to the goal
- A distinctive characteristic or quality
- Example features

Squire Trelawney, Dr. Livesey, and the rest of these gentlemen having asked me to write down the whole particulars about Treasure Island, from the beginning to the end, keeping nothing back but the bearings of the island, and that only because there is still treasure not yet lifted, I take up my pen in the year of grace 17— and go back to the time when my father kept the Admiral Benbow inn and the brown old seaman with the sabre cut first took up his lodging under our roof.

I remember him as if it were yesterday, as he came plodding to the inn door, his sea-chest following behind him in a hand-barrow; a tall, strong, heavy, nut-brown man, his tarry pigtail falling over the shoulder of his soiled blue coat, his hands ragged and scarred, with black, broken nails, and the sabre cut across one cheek, a dirty, livid white. I remember him looking round the cover and whistling to himself as he did so, and then breaking out in that old sea-song that he sang so often afterwards:

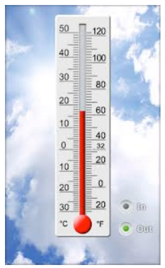
'*Fifteen men on the dead man's chest-Yo-ho-ho, and a bottle of rum!*' in the high, old tottering voice that seemed to have been tuned and broken at the capstan bars. Then he rapped on the door with a bit of stick like a handspike that he carried, and when my father appeared, called roughly for a glass of rum. This, when it was brought to him, he drank slowly, like a connoisseur, lingering on the taste and still looking about him at the cliffs and up at our signboard.

'This is a handy cove,' says he at length; 'and a pleasant sittyated grog-shop. Much company, mate?'

My father told him no, very little company, the more was the pity.

'Well, then,' said he, 'this is the berth for me. Here you, matey,' he cried to the man who trundled the barrow; 'bring up alongside and help up my chest. I'll stay here a bit,' he continued. 'I'm a plain man; rum and bacon and eggs is what I want, and that head up there for to watch ships off. What you mought call me? You mought call me captain. Oh, I see what you're at - there'; and he threw down three or four gold pieces on the threshold. 'You can tell me when I've worked through that,' says he, looking as fierce as a commander.

And indeed bad as his clothes were and coarsely as he spoke, he had none of the appearance of a



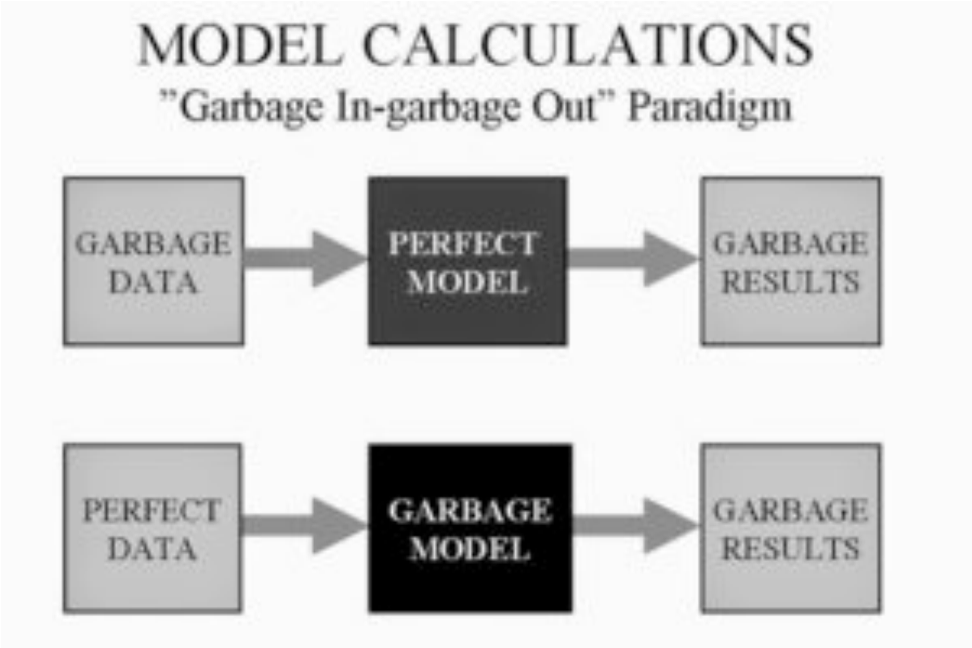
data1 →

data2 →

data3 →

Garbage in Garbage out

- The machine is as intelligent as the data/features we put in
- “Garbage in, Garbage out”
- Data cleaning is often done to reduce unwanted things



https://precisionchiroco.com/garbage-in-garbage-out/

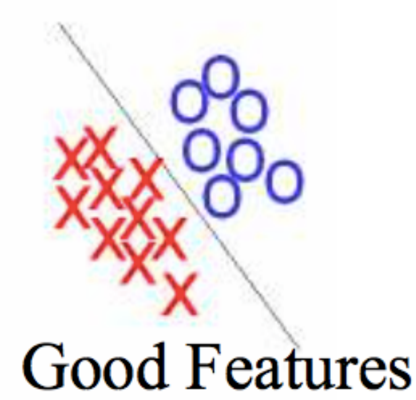
The need for data cleaning

However, good models should be able to handle some dirtiness!

https://www.linkedin.com/pulse/big-data-conundrum-garbage-out-other-challenges-business-platform

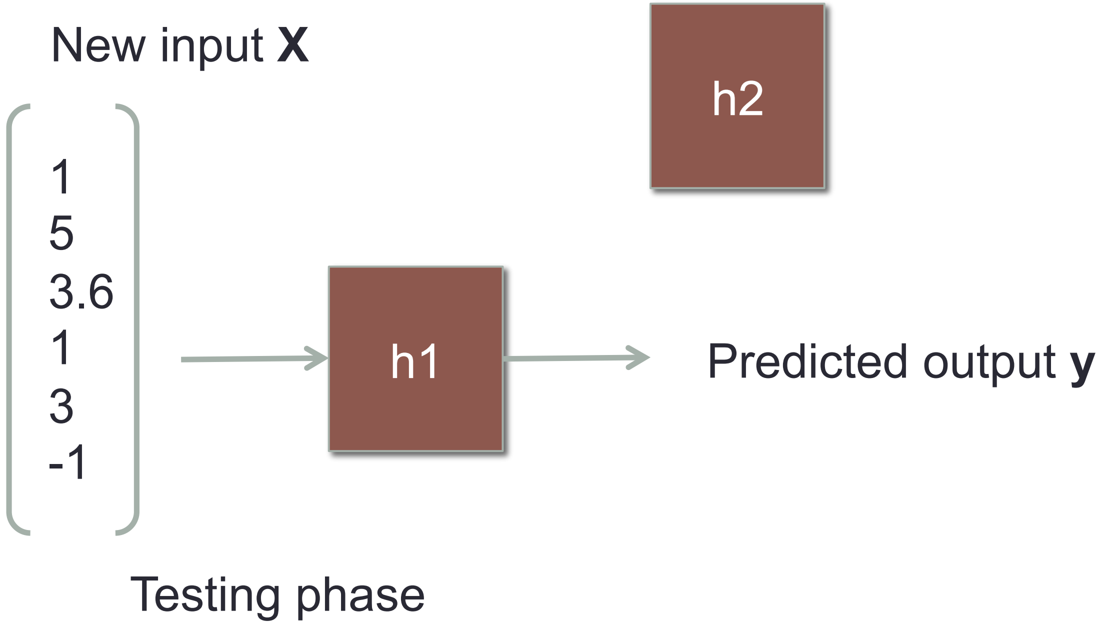
Feature properties

- The quality of the feature vector is related to its ability to discriminate samples from different classes



Model evaluation

How to compare h1 and h2?



Metrics

- Compare the output of the models
  - Errors/failures, accuracy/success
- We want to quantify the error/accuracy of the models
- How would you measure the error/accuracy of the following

Ground truths

- We usually compare the model predicted answer with the correct answer.
- What if there is no real answer?
  - How would you rate machine translation?

ไปไหน

Model A: Where are you going?
Model B: Where to?

Designing a metric can be tricky, especially when it’s subjective

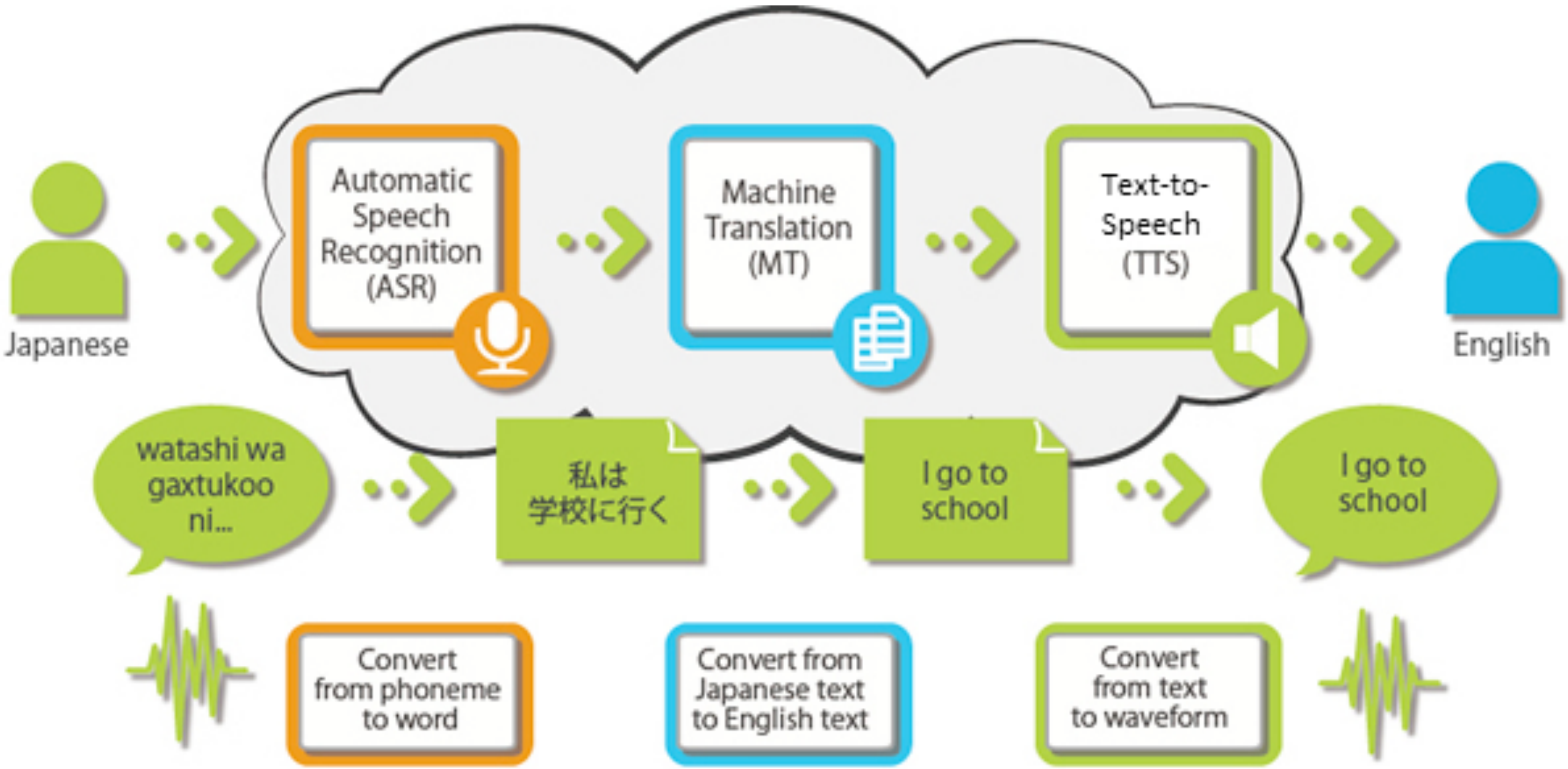
Metrics consideration 1

- Are there several metrics?

- Use the metric closest to your goal but never disregard other metrics.
  - May help identify possible improvements

Metrics consideration 2

- Are there sub-metrics?

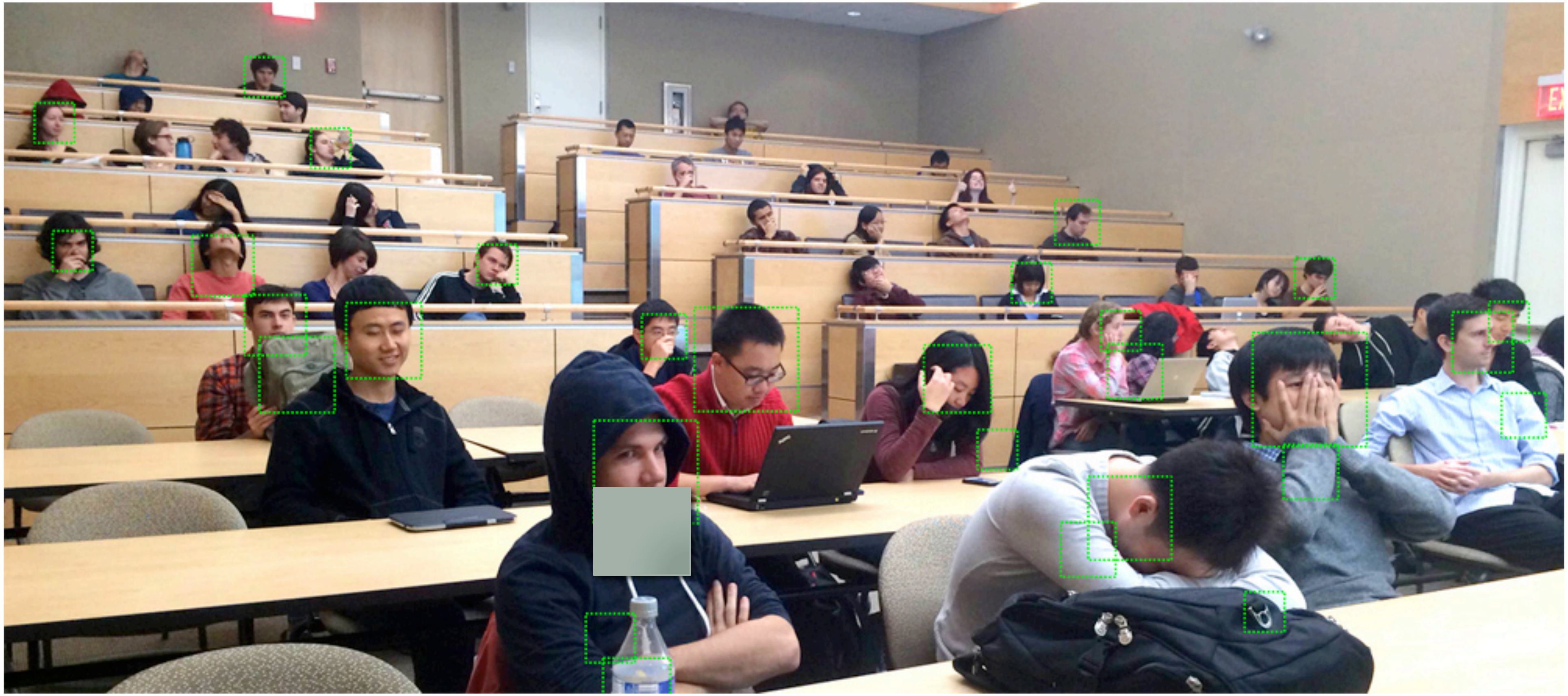


http://www.ustar-consortium.com/qws/slot/u50227/research.html

Metrics definition

- Defining a metric can be tricky when the answer is flexible

https://www.cc.gatech.edu/~hays/compvision/proj5/



Be clear about your definition of an error before hand!
Make sure that it can be easily calculated!
This will save you a lot of time.

Commonly used metrics

- Error rate
- Accuracy rate

- Precision
- True positive
- Recall
- False alarm
- F score

A detection problem

- Identify whether an event occur
- A yes/no question
- A binary classifier

Smoke detector

Hotdog detector

Evaluating a detection problem

- 4 possible scenarios

		Detector	
		Yes	No
Actual	Yes	True positive	False negative (Type II error)
	No	False Alarm (Type I error)	True negative

True positive + False negative = # of actual yes
False alarm + True negative = # of actual no

- False alarm and True positive carries all the information of the performance.

Definitions

- True positive rate (Recall, sensitivity)

  = # true positive / # of actual yes
- False positive rate (False alarm rate)

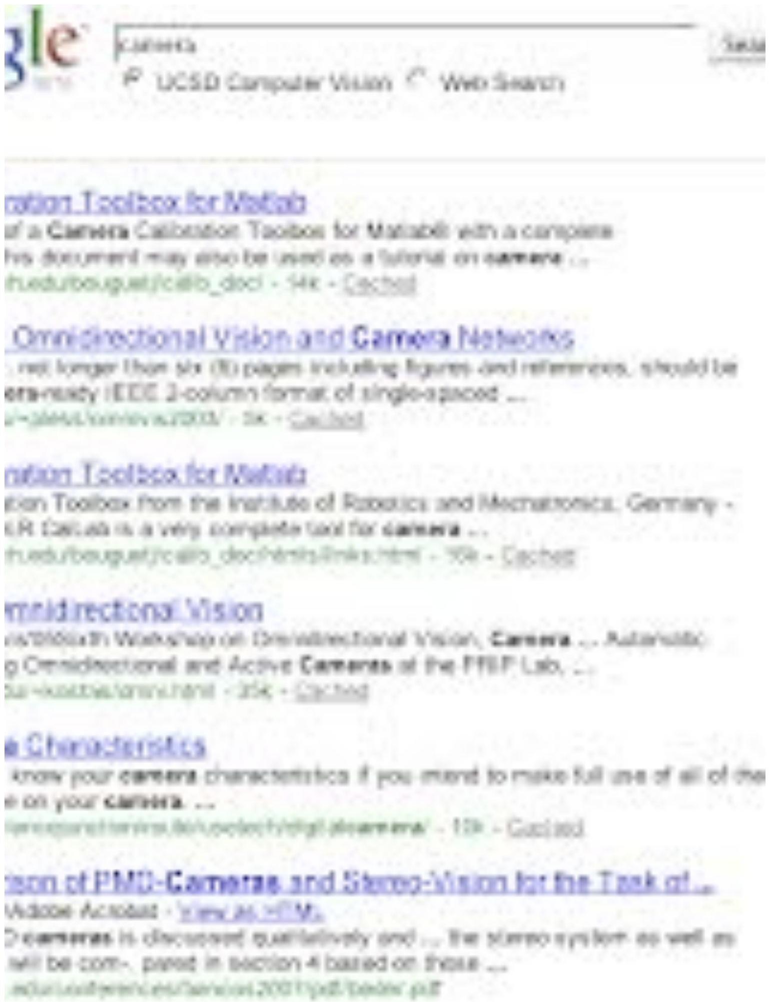
  = # false positive / # of actual no
- False negative rate (Miss rate)

  = # false negative / # of actual yes
- True negative rate (Specificity)

  = # true negative / # of actual no

- Precision = # true positive / # of predicted positive

Search engine example



A recall of 50% means?

A precision of 50% means?

When do you want high recall?
When do you want high precision?

Recall/precision

- When do you want high recall?
- When do you want high precision?

- Initial screening for cancer
- Face recognition system for authentication
- Detecting possible suicidal postings on social media

Usually there’s a trade off between precision and recall. We will re-visit this later

Definitions 2

- F score (F1 score, f-measure)

$$F_1 = 2 \cdot \frac{1}{\frac{1}{\mathrm{recall}} + \frac{1}{\mathrm{precision}}} = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}.$$

  - A single measure that combines both aspects
  - A harmonic mean between precision and recall (an average of rates)

Note that precision and recall says nothing about the true negative

Harmonic mean vs Arithmetic mean

- You travel for half an hour for 60 km/hr, then half an hour for 40 km/hr. What is your average speed?
  - Arithmetic mean = 50 km/hr
  - Harmonic mean

$$\frac{n}{\frac{1}{x_1}+\ldots+\frac{1}{x_n}} = \frac{2}{\frac{1}{40}+\frac{1}{60}} = \text{48 km/hr}$$

  - Total distance covered in 1 hour = 30+20 = 50

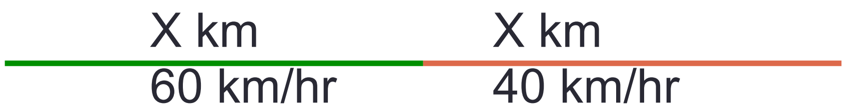
30 mins	30 mins
60 km/hr	40 km/hr

Harmonic mean vs Arithmetic mean

- You travel for distance X for 60 km/hr, then another X for 40 km/hr. What is your average speed?
  - Arithmetic mean = 50 km/hr
  - Harmonic mean

$$\frac{n}{\frac{1}{x_1}+\ldots+\frac{1}{x_n}} = \frac{2}{\frac{1}{40}+\frac{1}{60}} = \text{48 km/hr}$$

  - Total distance covered 2X



Harmonic mean vs Arithmetic mean

- For the arithmetic mean to be valid you need to compared over the same number of hours (denominator)

- For precision and recall, you have different denominators, but the same numerator, which fits the harmonic mean.

True positive rate (Recall, sensitivity)
= # true positive / # of actual yes

Precision = # true positive / # of predicted positive

Evaluating models

- We talked about the training set used to learn the model
- We use a different data set to test the accuracy/error of models – "test set"
- We can still compute the error and accuracy on the training set
- Training error vs Testing error
- We will discuss how we can use these to help guide us later

Other considerations when evaluating models

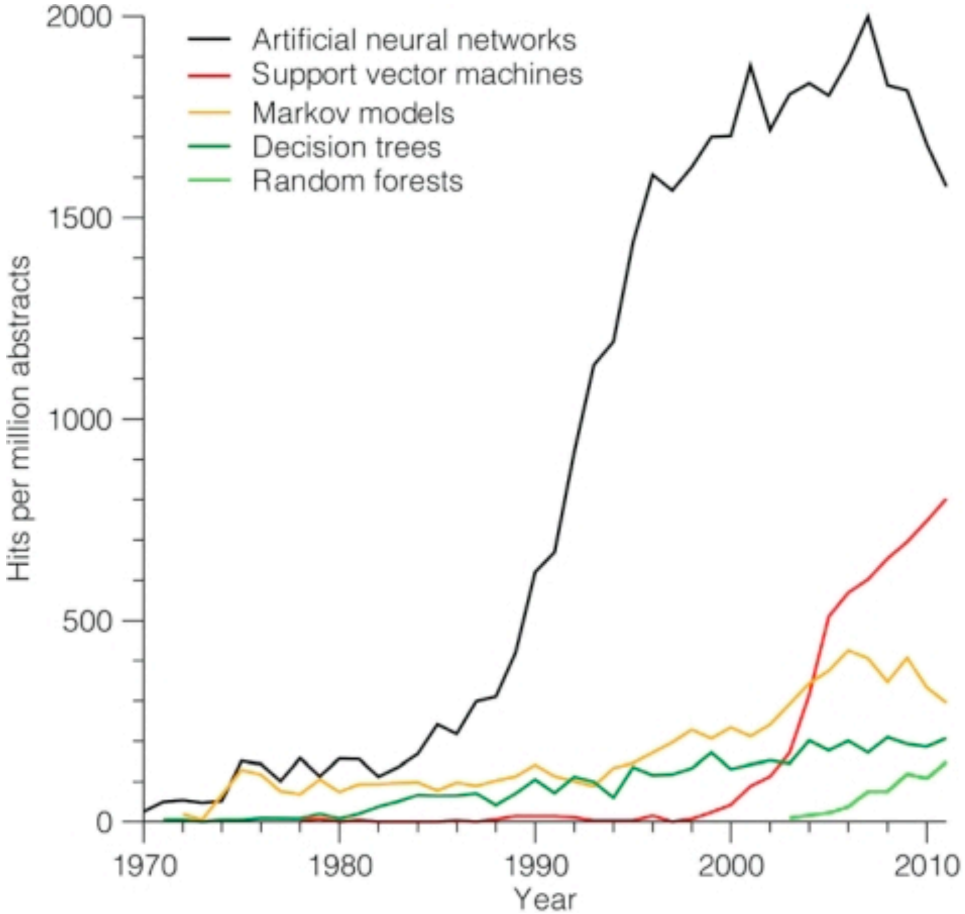
- Training time
- Testing time
- Memory requirement
- Parallelizability
- Latency

Course walkthrough

1 - 15/8	Introduction
2 - 22/8	Applications in Bioinformatics
3 - 29/8	Regression & Jupyter Notebook demo
4 - 5/9	Probability review
5 - 12/9	MLE, MAP, and Naive Bayes
6 - 19/9	GMM and EM
7 - 26/9	Dimensionality reduction and visualization
	Midterm week - No classes
8 - 9/10	SVM
9 - 16/10	Neural network basics & Gcloud, TensorFlow, and Keras demo
10 - 23/10	CNN, Recurrent architectures
11 - 30/10	Recent Advances in NN
12 - 7/11	Reinforcement Learning
13 - 14/11	Unsupervised methods
14 - 21/11	Tricks of the trade: machine learning in the real world
15 - 28/11	Project presentation

Why anything else besides deep learning

- The rise and fall of machine learning algorithms



Methods used in bioinformatics papers

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3232371/figure/F1/

What we will not cover

- Random forest
- Decision trees
- Boosting
- Graphical models