03.01.2019

Statistical Methods in AI (CSE/ECE 471)

Lecture-1: Intro and Administrivia

Ravi Kiran (ravi.kiran@iiit.ac.in)

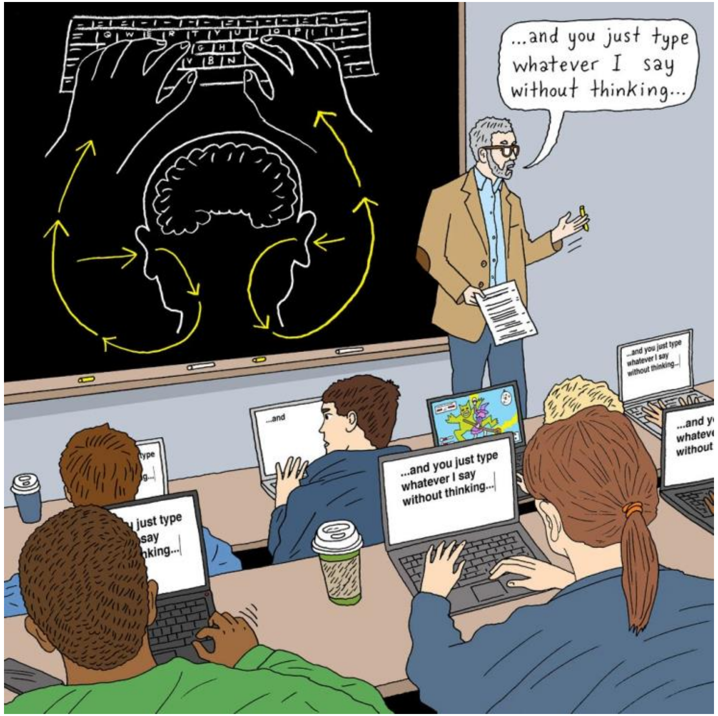
Vineet Gandhi (vineet@iiit.ac.in)

Center for Visual Information Technology (CVIT)

IIIT Hyderabad

No laptops

In a series of experiments at Princeton University and the University of California, Los Angeles, students were randomly assigned either laptops or pen and paper for note-taking at a lecture. Those who had used laptops had substantially worse understanding of the lecture, as measured by a standardized test, than those who did not.



SMAI (Statistical Methods in AI)

- SMAI ~ Introduction to Machine Learning

Machine Learning

Study of Algorithmic methods that use data to improve their knowledge of a task

Machine Learning: Examples

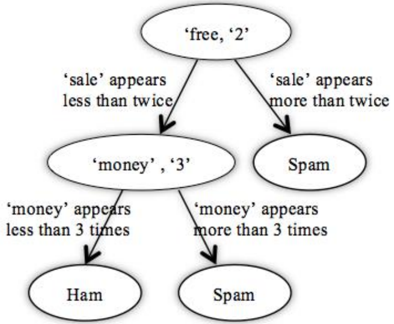
DATA → **LEARNING ALGORITHMS** → **KNOWLEDGE**

Algorithmic methods that use data to improve their knowledge of a task

Task: Detect spam email

Data: Labelled emails (in inboxes of other users as well !)

Knowledge:



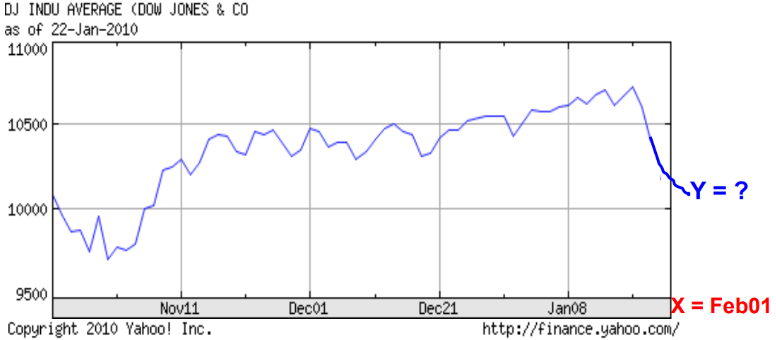
Improve → 85% reduction of spam emails in Inbox over 3 months

Algorithmic method: Decision Tree

Machine Learning: Examples

Algorithmic methods that use data to improve their knowledge of a task

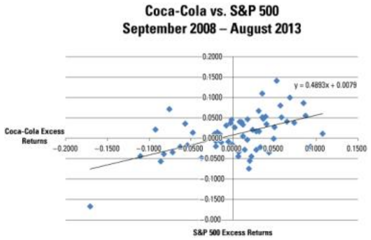
Task: Predict value of a stock (GOOG)



Data: Historical stock value (time, price/share)

Knowledge: Model coefficients

Improve → Predict stock to 95% of its value



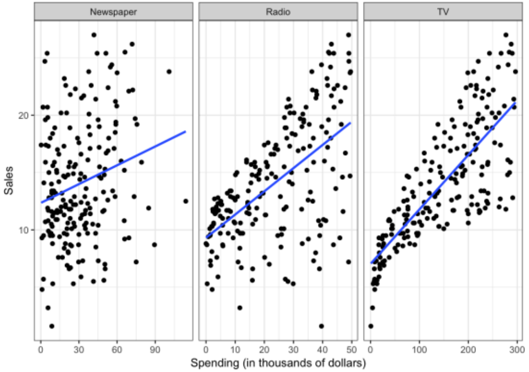
Algorithmic method: Linear Regression

Machine Learning: Examples

Algorithmic methods that use data to improve their knowledge of a task

Task: Predict effect of advertising on 'furniture' sales

Algorithmic method: Linear Regression



Data: Amount spent on ad spots in TV, radio, newspaper

	TV	Radio	Newspaper	Sales
0	230.1	37.8	69.2	22.1
1	44.5	39.3	45.1	10.4
2	17.2	45.9	69.3	9.3
3	151.5	41.3	58.5	18.5
4	180.8	10.8	58.4	12.9

Knowledge: For a given amount of TV and newspaper advertising, spending additional 10,000 rupees on FM radio leads to an additional sale of 150 units

Machine Learning: Examples

Algorithmic methods that use data to improve their knowledge of a task

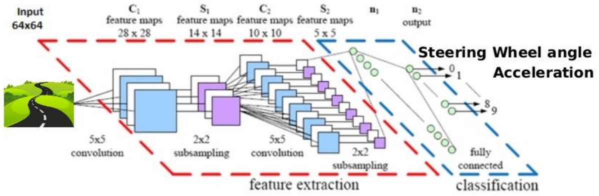
Task: Drive car ‘safely’ without human intervention

Algorithmic method: Deep + Rule-Based Learning

Data: Camera, Laser , GPS data ; Synthetic data

Knowledge: Model coefficients

Improve → Drive 160,000 miles without accident/human intervention

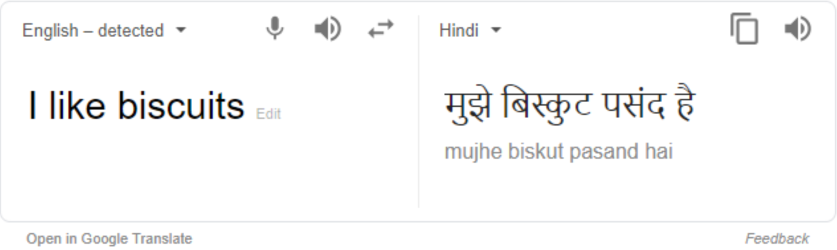


Machine Learning: Examples

DATA → LEARNING ALGORITHMS → KNOWLEDGE

Algorithmic methods that use data to improve their knowledge of a task

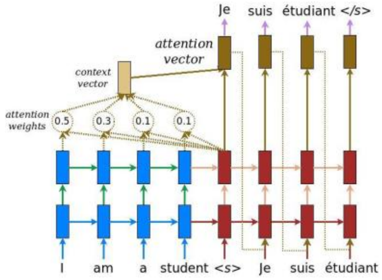
Task: Translate text from one language to another



Data: Paired sentences from source and target languages

Knowledge: Model coefficients

Improve → Reduce number of mistakes by 78%



Algorithmic method: Deep Recurrent Neural Networks

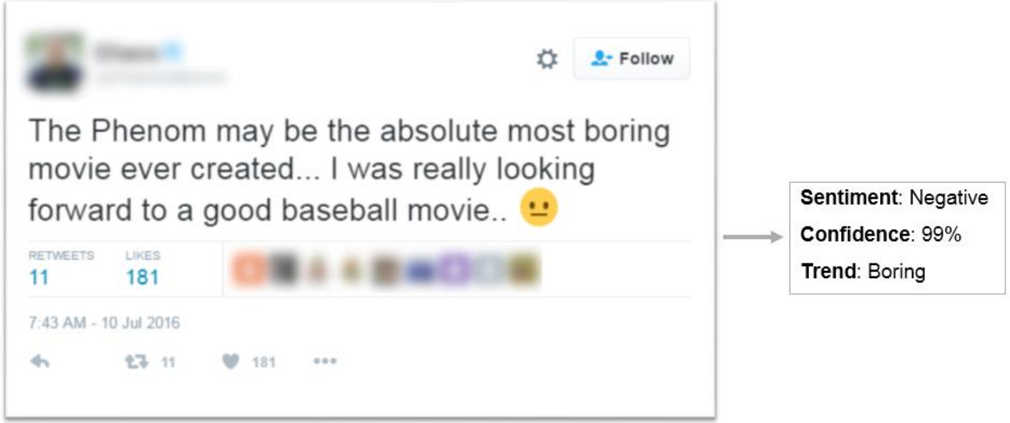
Machine Learning: Examples

DATA → LEARNING ALGORITHMS → KNOWLEDGE

Algorithmic methods that use data to improve their knowledge of a task

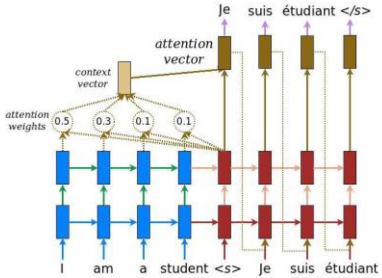
Task: Sentiment Analysis

Data: Text and ‘Sentiment’ label



Knowledge: Model coefficients

Improve → Reduce number of sentiment mislabelings by 80%



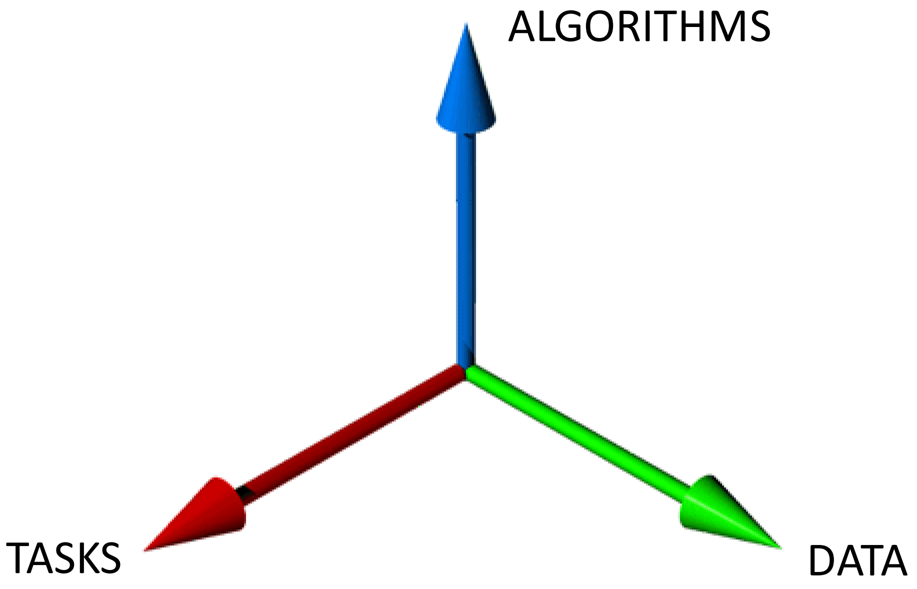
Algorithmic method: Deep Recurrent Neural Networks

What is ML ? (alternate definitions)

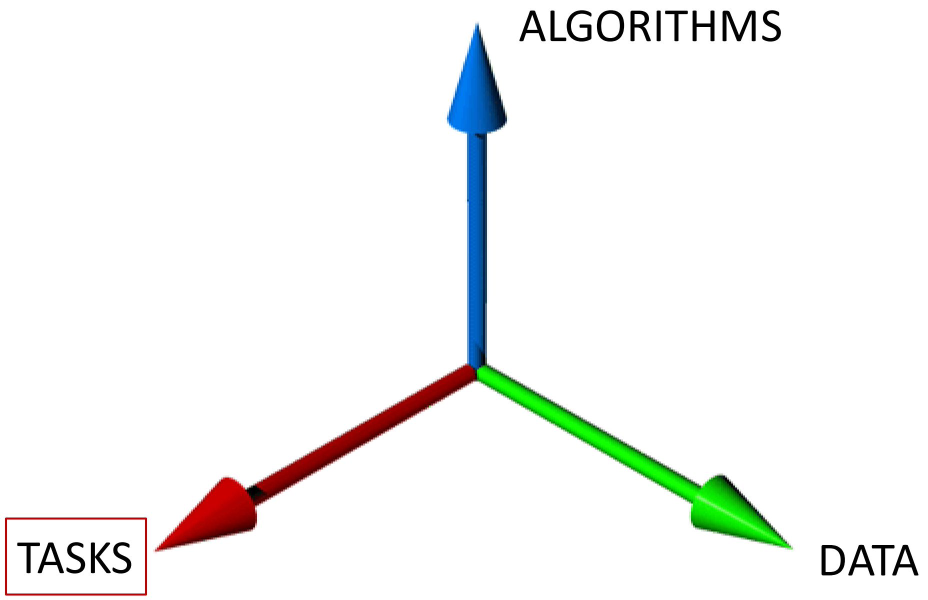
- Computer program whose behavior evolve based on empirical data (Wikipedia)
- Computer program that learns from experience E in order to improve its performance P on a task T (Tom Mitchell)

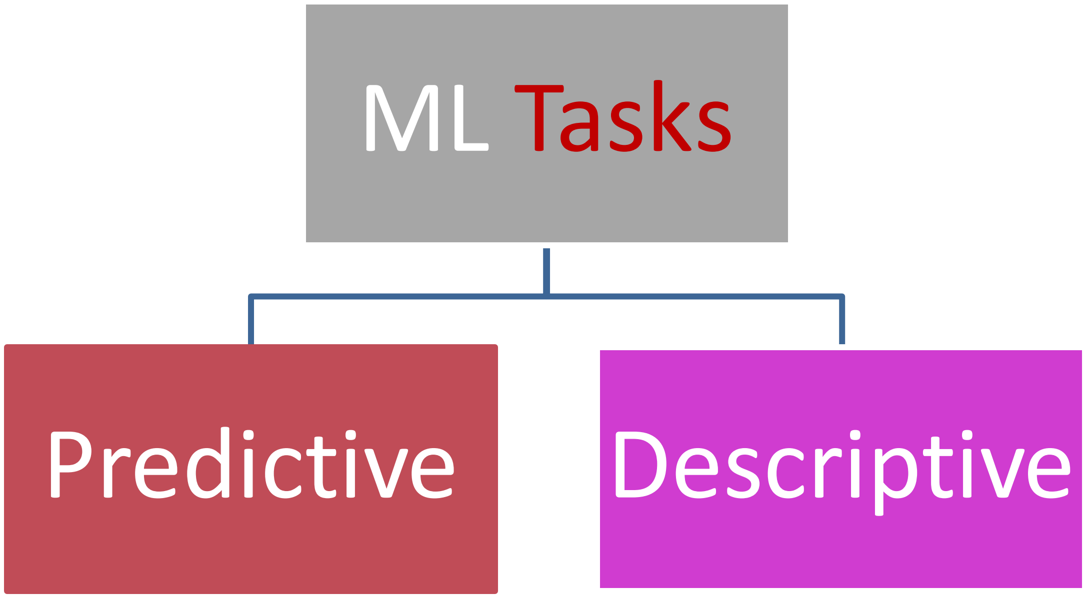
experience E : images, text, sensor measurements, biological data
task T : estimating probabilities, predicting object label, dimensionality reduction, clustering
performance P : probability of success, money/time saved,

3 axes of ML



3 axes of ML





Given an input,
estimate output

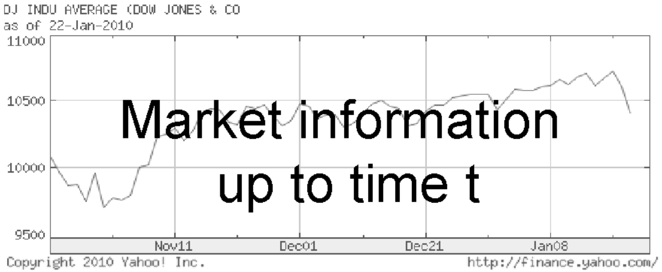
ML::Tasks → Predictive

Feature Space $\mathcal{X}$

Label Space $\mathcal{Y}$

Words in a document

“Sports”
“News”
“Science”
…



Market information up to time t

Share Price
“$ 24.50”

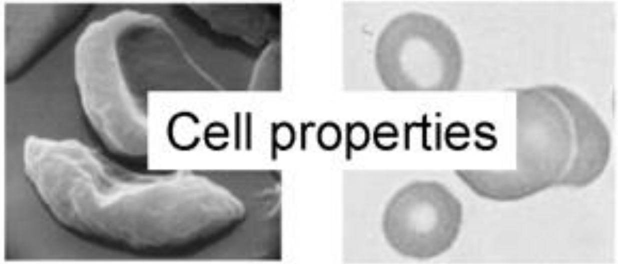
Task: Given $X \in \mathcal{X}$, predict $Y \in \mathcal{Y}$.

ML::Tasks → Predictive → Classification

Feature Space $\mathcal{X}$

Label Space $\mathcal{Y}$

"Sports"
"News"
"Science"
...



"Anemic cell"
"Healthy cell"

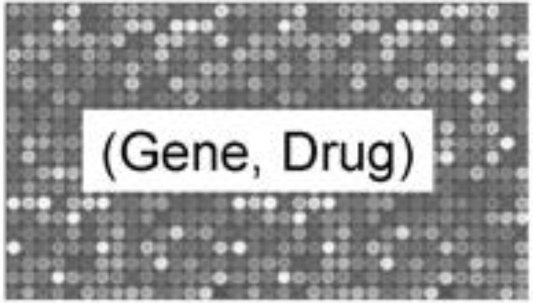
Task: Given $X \in \mathcal{X}$, predict $Y \in \mathcal{Y}$.

Discrete Labels

ML::Tasks → Predictive → Regression

Feature Space $\mathcal{X}$ **Label Space** $\mathcal{Y}$

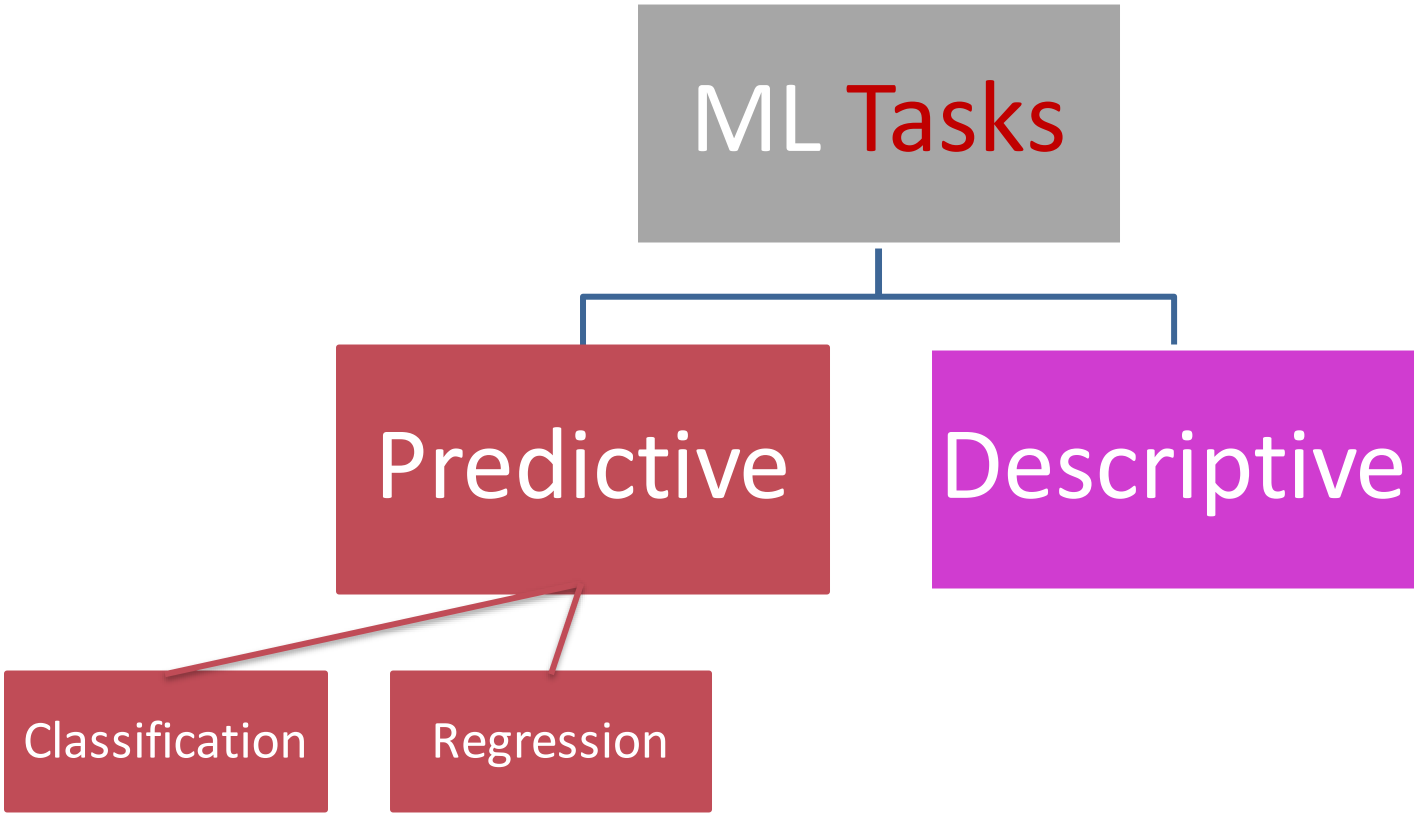
Share Price
"$ 24.577"



Expression level
"6.88"

Task: Given $X \in \mathcal{X}$, predict $Y \in \mathcal{Y}$.

Continuous Labels


ML Tasks
Predictive
Descriptive
Classification
Regression

ML Tasks

Predictive

Descriptive

Given an input,
study its ‘structure’

ML::Tasks → Descriptive

- Study/Exploit the ‘structure’ of data
  - Density Estimation
  - Clustering
  - Dimensionality Reduction
- Also studied as ‘Unsupervised Learning’
  - ‘Input’ data without paired ‘Output’

Unsupervised Learning → Density Estimation

Aka "learning without a teacher"

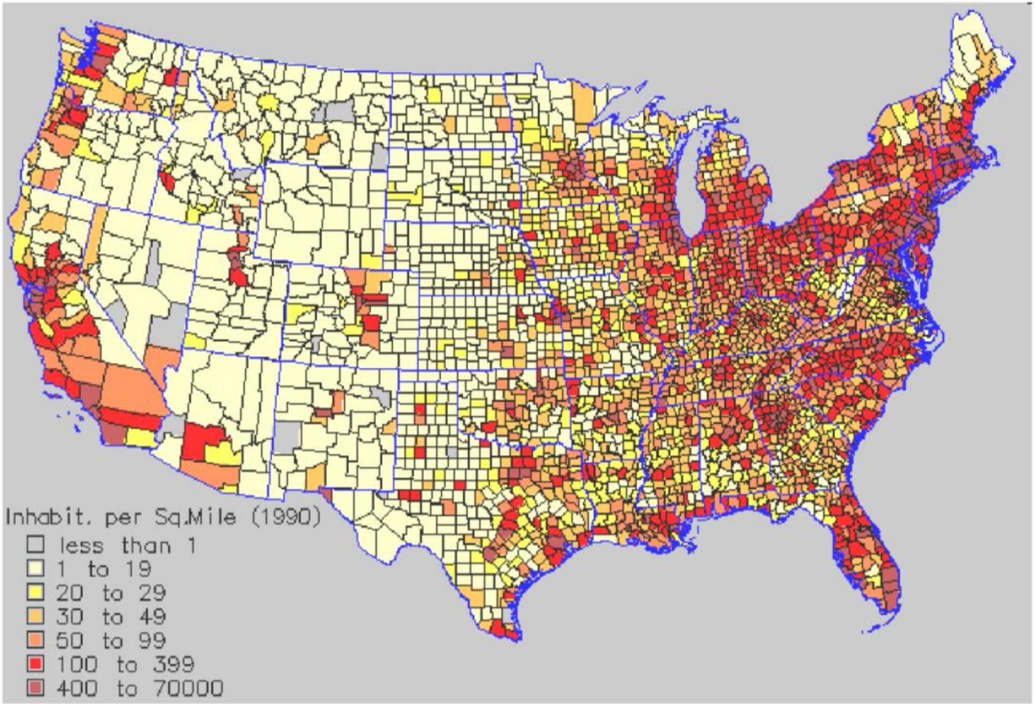
Feature Space $\mathcal{X}$

Word distribution
(Probability of a word)

Task: Given $X \in \mathcal{X}$, learn $f(X)$.

Unsupervised Learning → Density Estimation

Population density

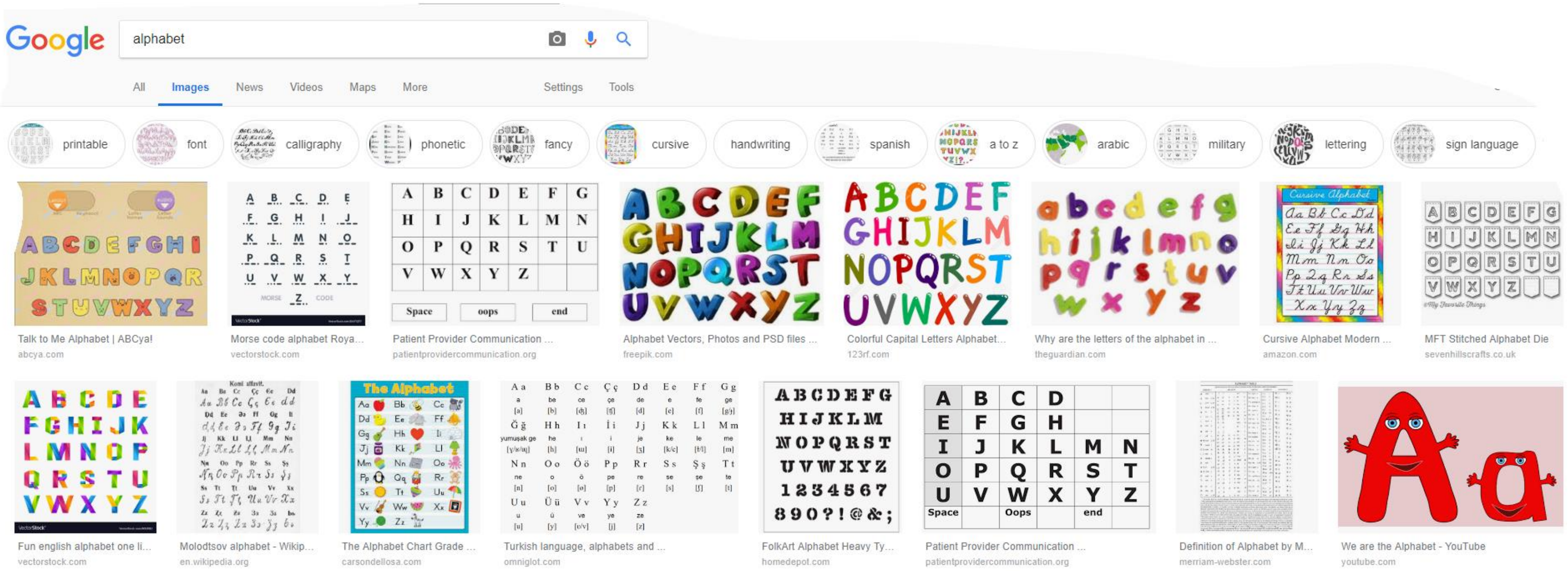


Unsupervised Learning → Clustering

Group similar things e.g. images

[Goldberger et al.]

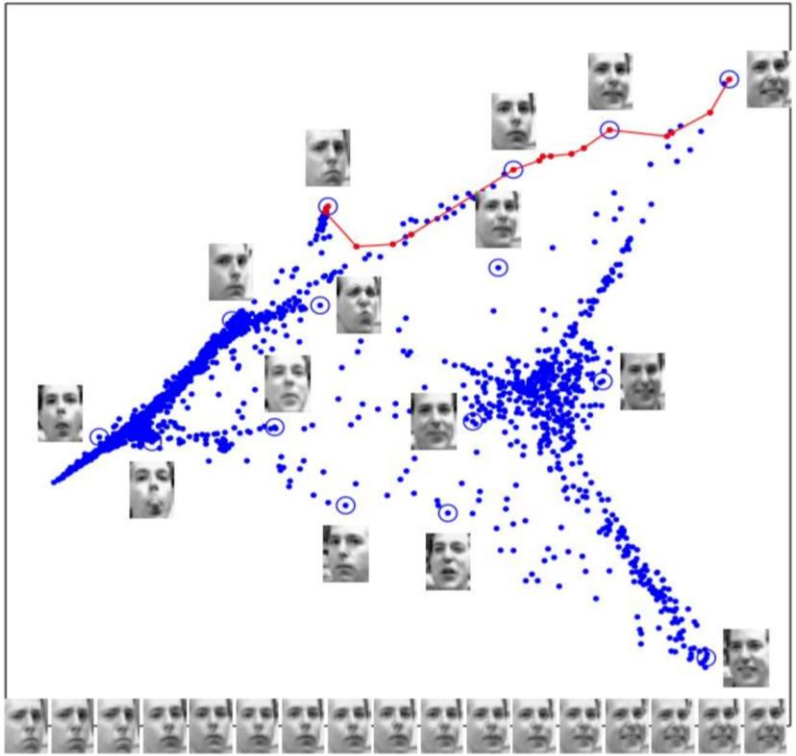
Unsupervised Learning → Web Search



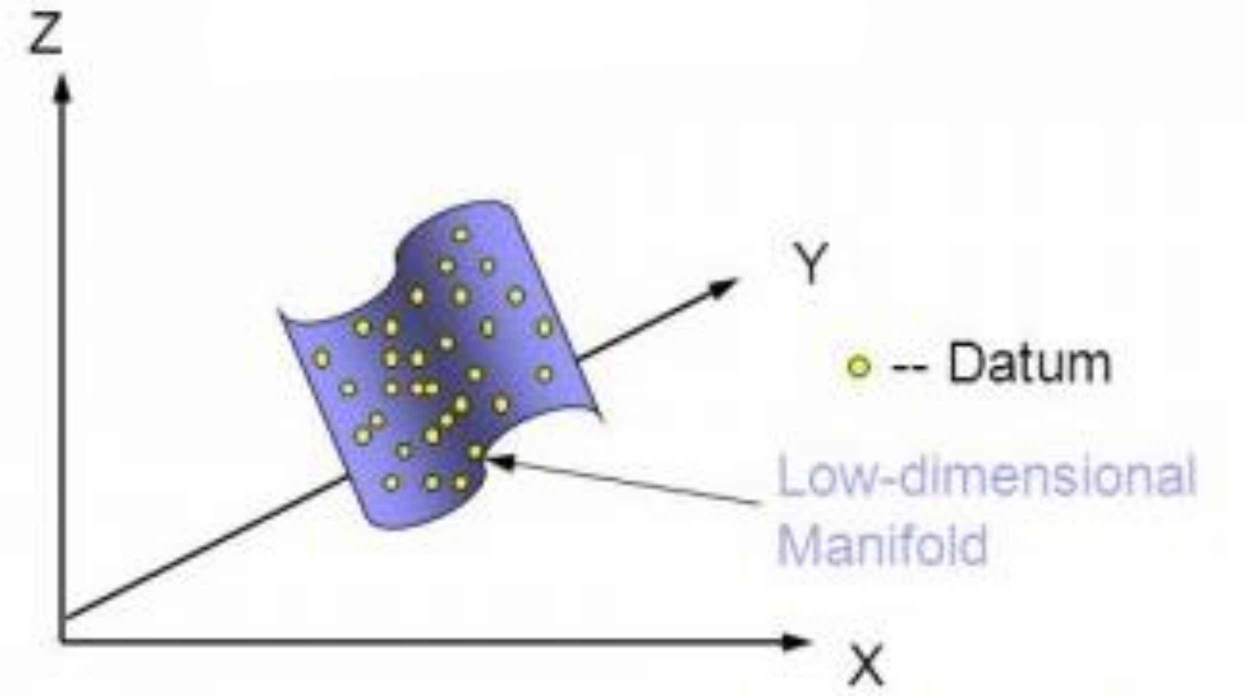
Unsupervised Learning → Dimensionality Reduction + Visualization

Images have thousands or millions of pixels.

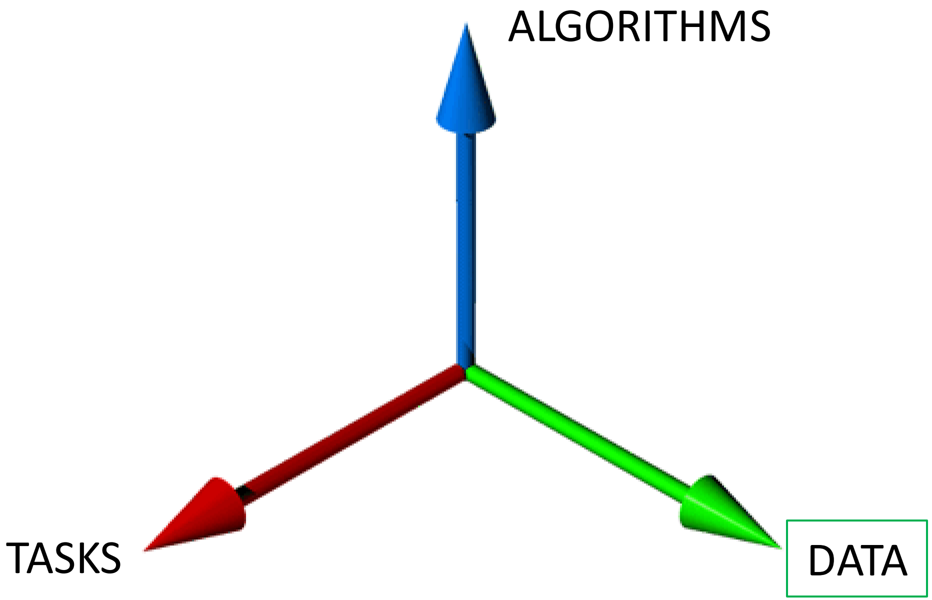
Can we give each image a coordinate, such that similar images are near each other?



Unsupervised Learning → Dimensionality Reduction

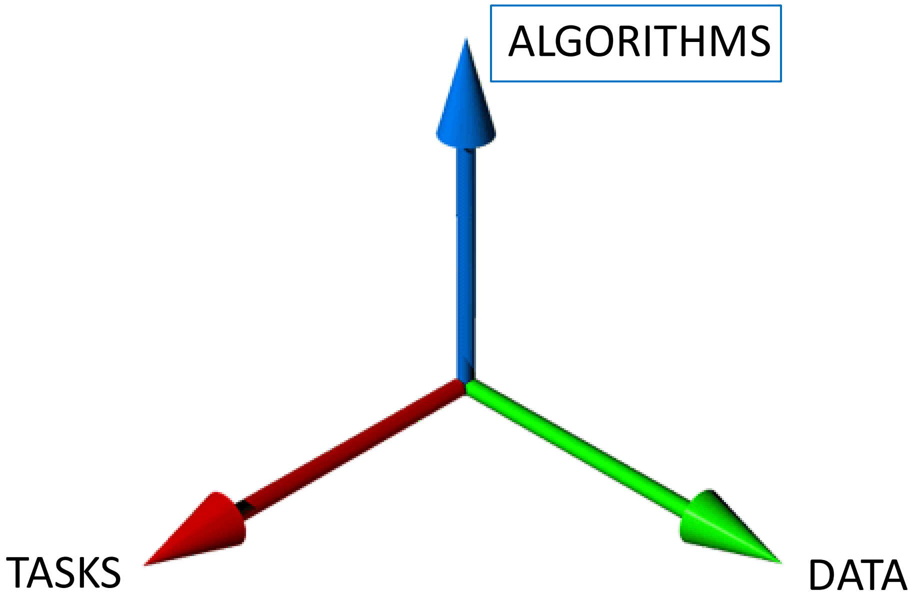


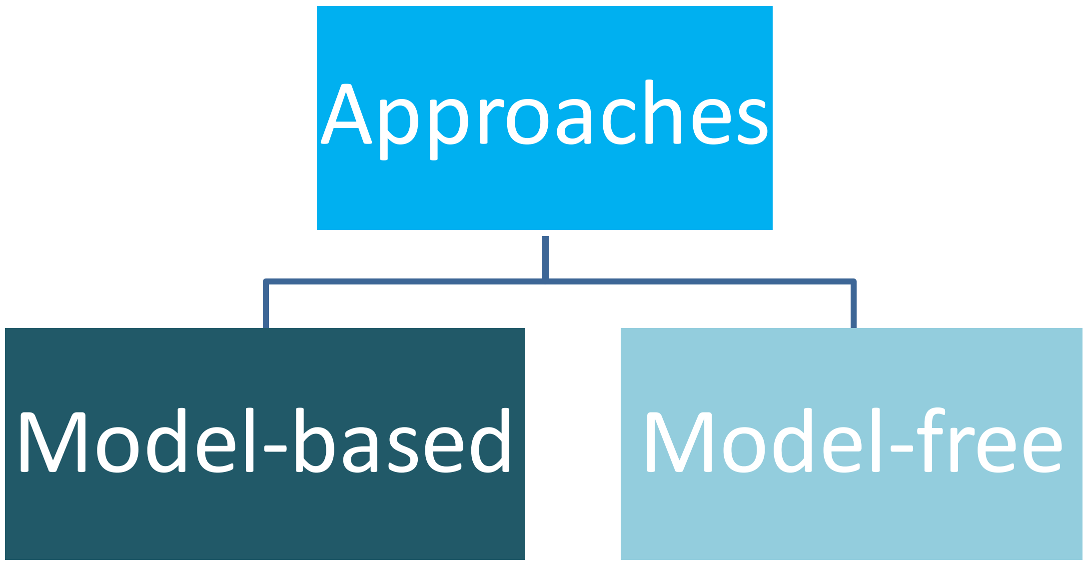
3 axes of ML



- Fully Observed
- Partially Observed
  - Some variables systematically not observed (e.g. 'topic' of a document)
  - Some variables missing some of the time (e.g. 'faulty sensor' readings)

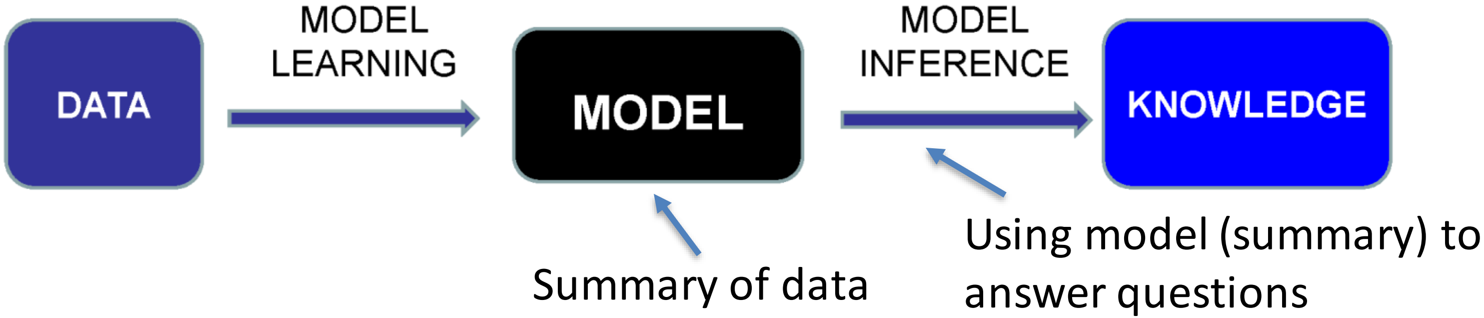
3 axes of ML

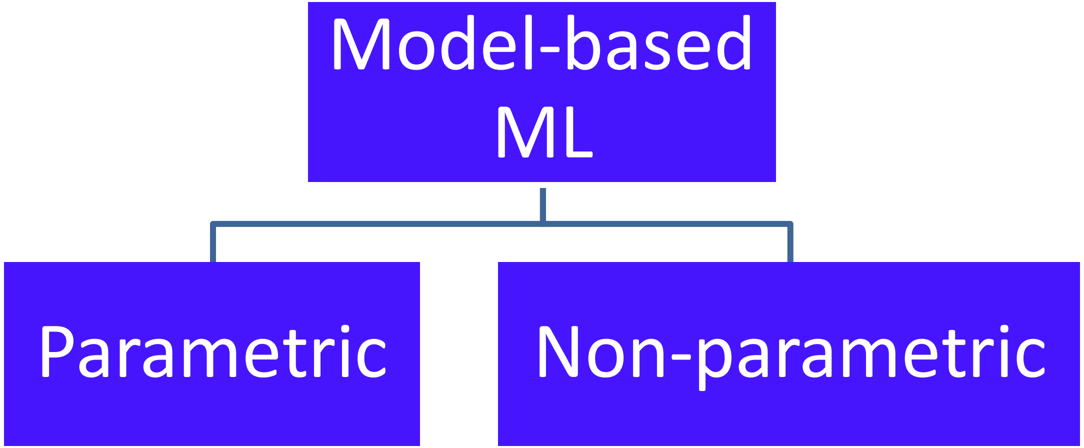



Approaches
Model-based
Model-free

Model-based ML

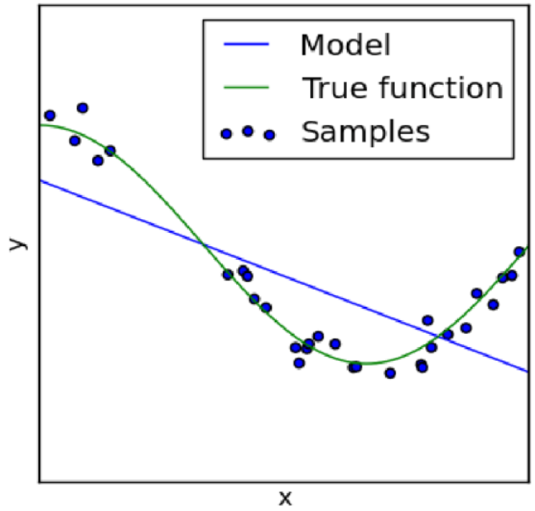
Algorithmic methods that use data to improve their knowledge of a task




Model-based ML
Parametric
Non-parametric

Parametric Models

- “Fixed-size” models that do not “grow” with the data
- More data just means you learn/fit the model better

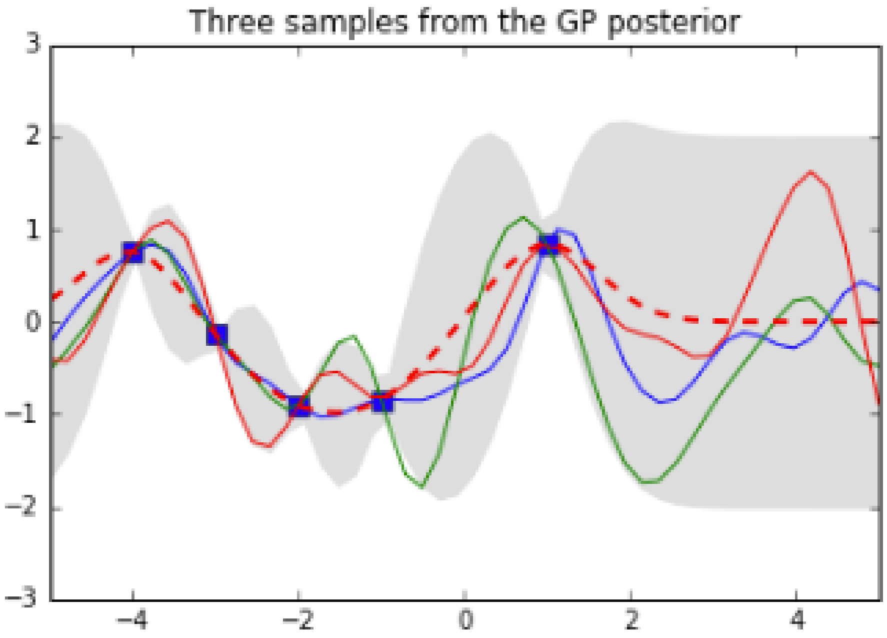


Fitting a simple line (2 params) to a bunch of one-dim. samples

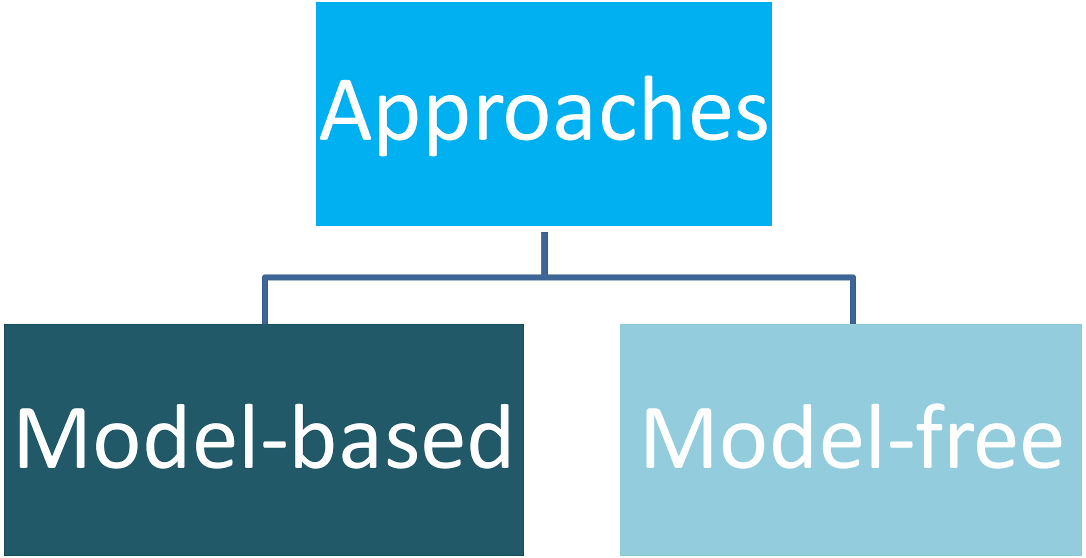
Model: data = point on line + noise

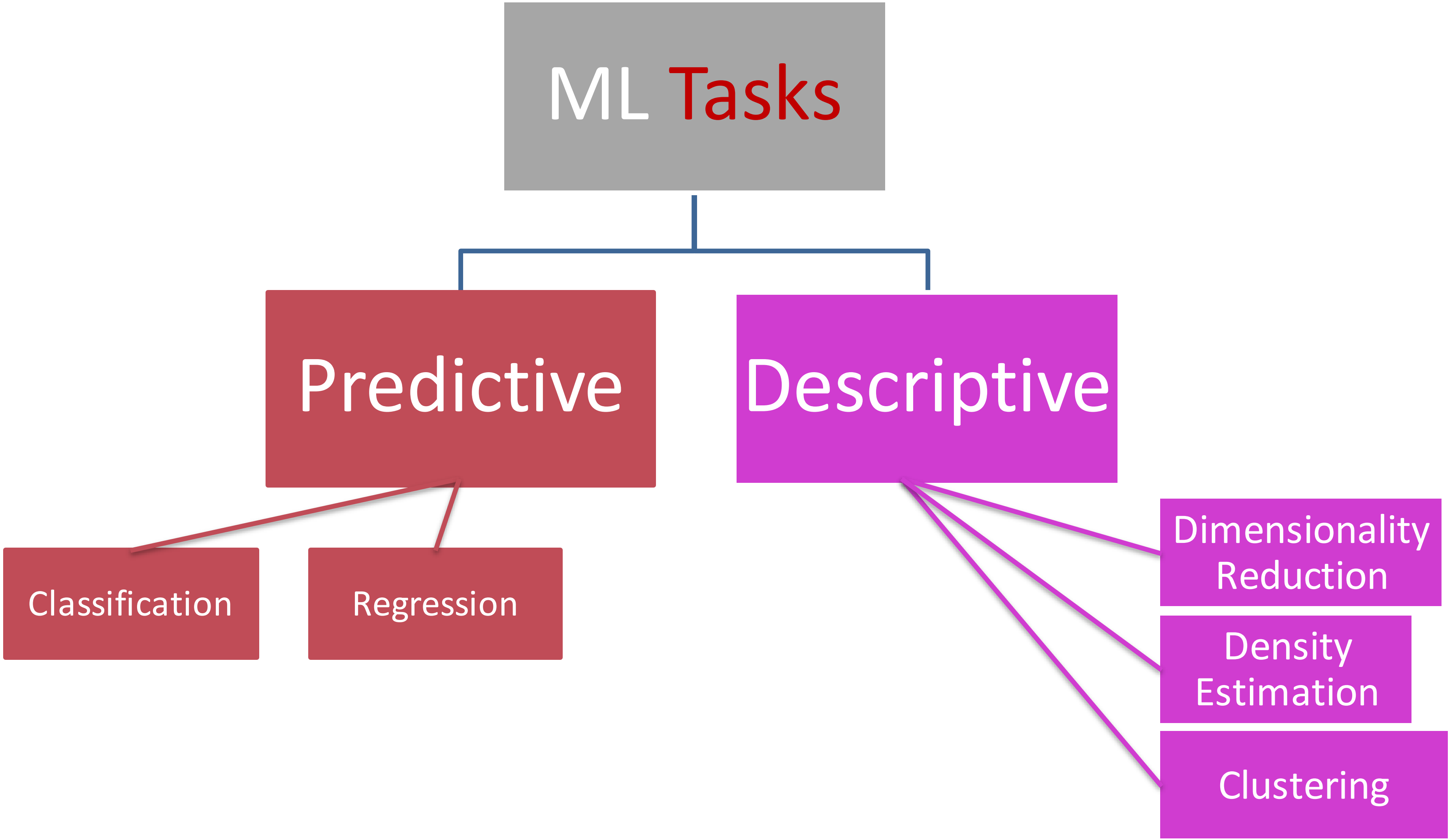
Nonparametric Models

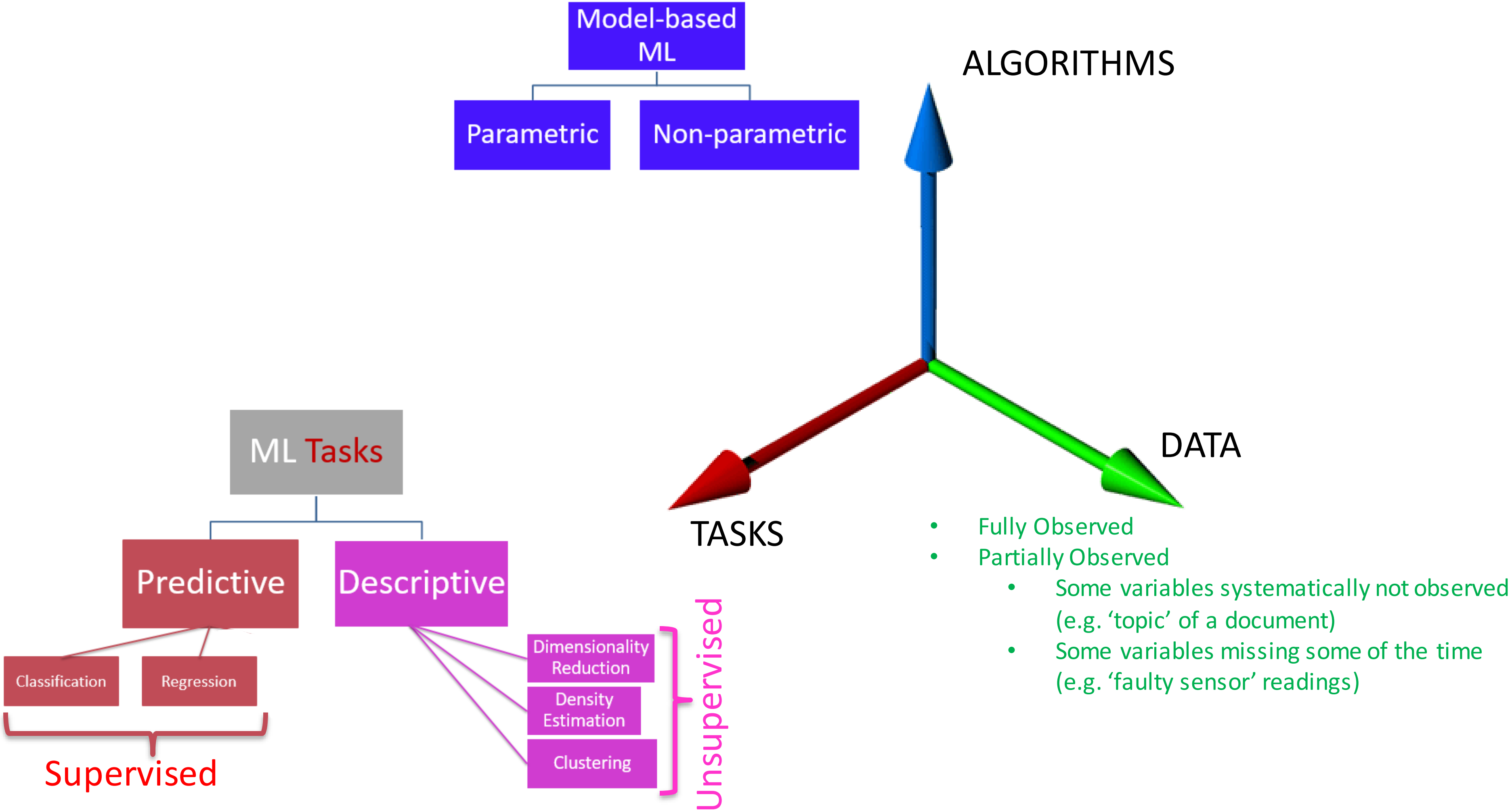
- Models that grow with the data
- More data means a more complex model

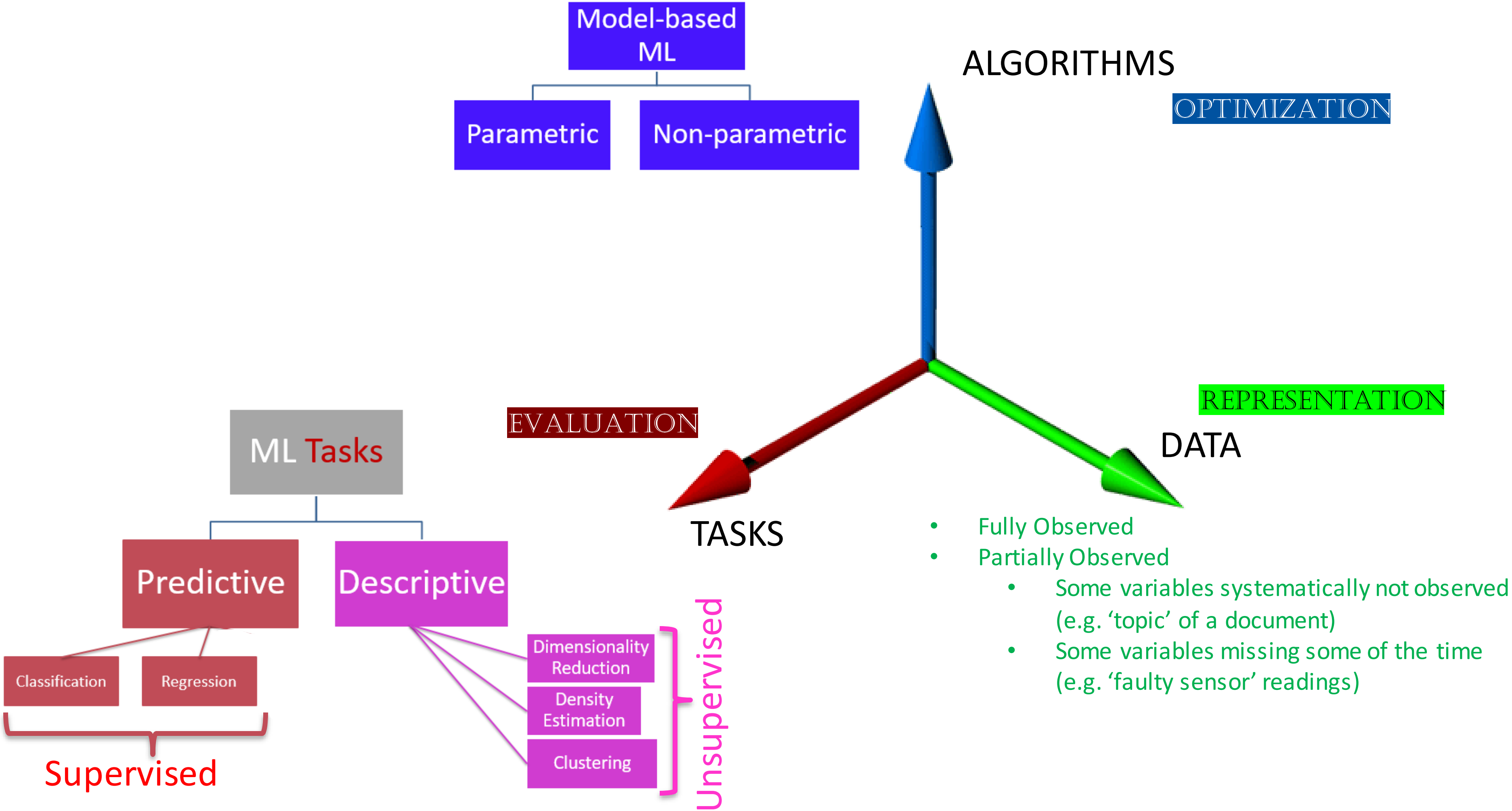


Gaussian Process

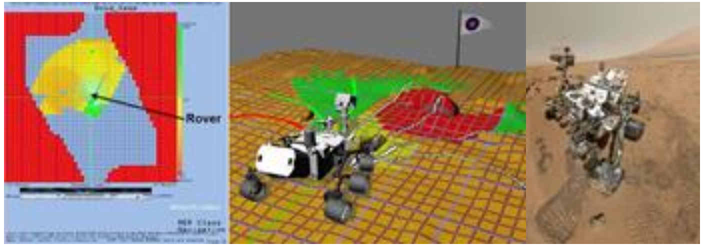

Approaches
Model-based
Model-free


ML Tasks
Predictive
Descriptive
Classification
Regression
Dimensionality Reduction
Density Estimation
Clustering


Model-based ML
Parametric
Non-parametric
ALGORITHMS
DATA
TASKS
ML Tasks
Predictive
Descriptive
Classification
Regression
Supervised
Dimensionality Reduction
Density Estimation
Clustering
Unsupervised
Fully Observed
Partially Observed
Some variables systematically not observed (e.g. ‘topic’ of a document)
Some variables missing some of the time (e.g. ‘faulty sensor’ readings)


Model-based ML
Parametric
Non-parametric
ALGORITHMS
OPTIMIZATION
REPRESENTATION
DATA
EVALUATION
TASKS
ML Tasks
Predictive
Descriptive
Classification
Regression
Supervised
Dimensionality Reduction
Density Estimation
Clustering
Unsupervised
Fully Observed
Partially Observed
Some variables systematically not observed (e.g. ‘topic’ of a document)
Some variables missing some of the time (e.g. ‘faulty sensor’ readings)

When to “Learn”

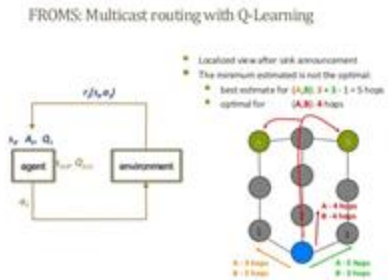


Human expertise does not exist

(‘learning’ to navigate on Mars)

Humans unable to explain their expertise

(‘learning’ to understand speech)

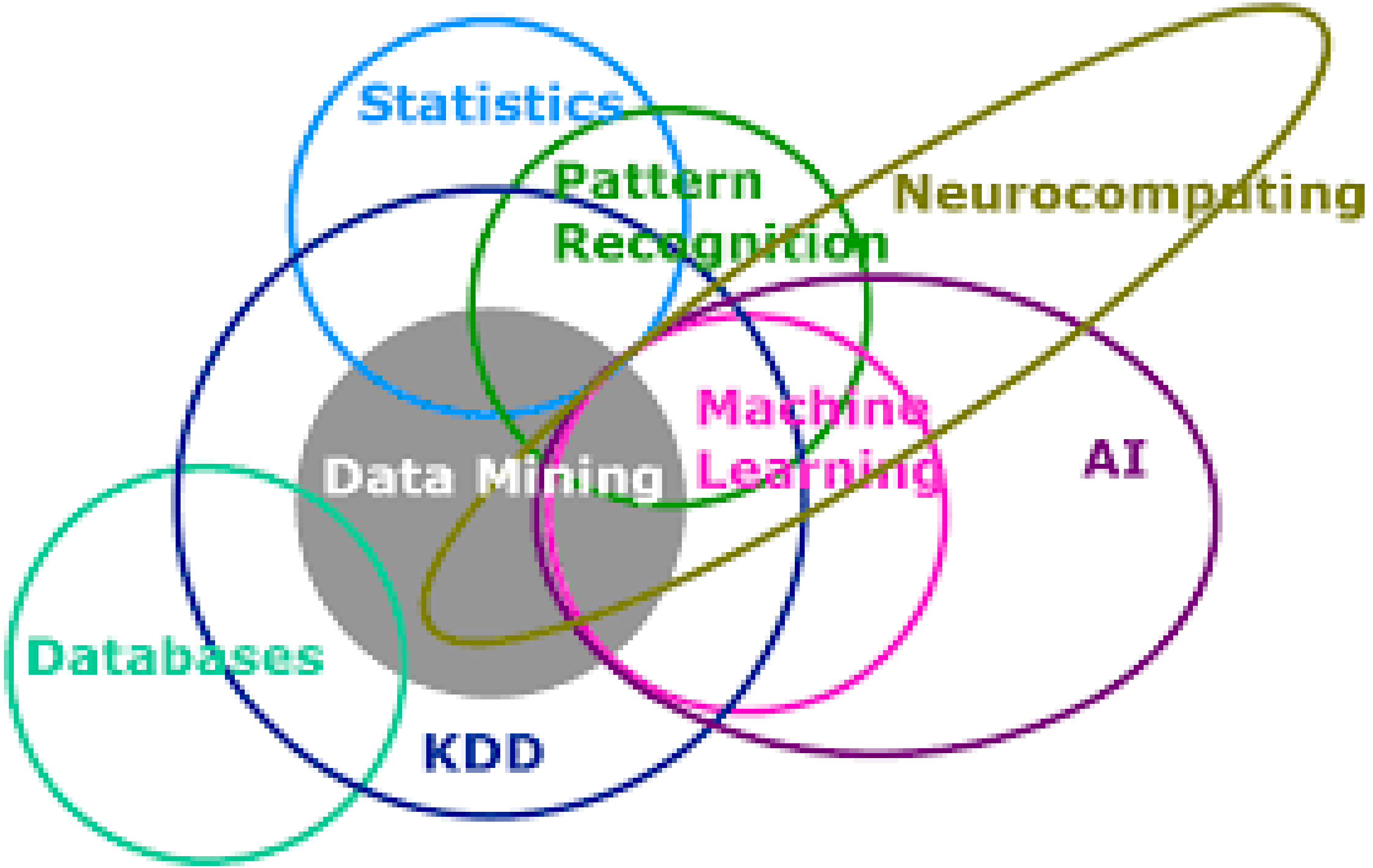


Solution changes over time

(‘learning’ to route network packet traffic)

Solution needs to be adapted to particular cases

(user-specific ‘learning’)



https://www.analyticsvidhya.com/blog/2015/07/difference-machine-learning-statistical-modeling/

ML v/s Statistics

- Statistics:
  - Common assumption: Data is generated by a model
  - Cares about: How well does data fit the model ?
- ML
  - Cares about: How well does model fit the data ?

About the course (471)

- Timings: Tue, Fri (Himalaya 205, 5.00p – 6.30p)

- Tutorial: Sat, Himalaya 205, 3.30p – 4.30p (tentative)

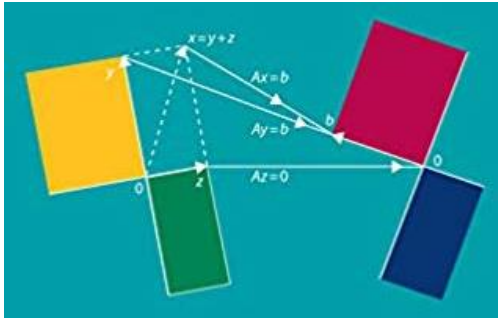
Course Overview

- Part-1 : Supervised Learning
- Part-2 : Unsupervised Learning
- Part-3 : Feature Selection, Ensemble Learning
- Part-4 : Neural Networks
- Part-5 : ML for sequential data
- Part-6 : Model Selection and Statistical Estimation
- Part-7 : Ranking and Retrieval

Pre-requisites

- CS
  - Programming
  - Data Structures (lists, trees, queues)
  - Algorithms (sort, search)

Pre-requisites

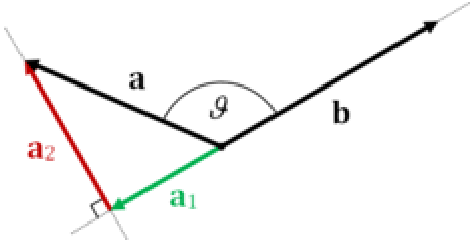


- Mathematics
  - Linear Algebra
    - Matrix, Vector operations
    - Systems of equations, Matrix Form (Ax = b), Conditions for existence of solution
    - Rank
    - Invertibility of matrix
    - Eigenvectors, Eigenvalues,
    - Semi-definiteness of matrix
    - Decompositions (Singular Value Decomposition, Eigendecomposition)
    - Properties of symmetric matrices

[Linear Algebra in 4 pages: https://courses.engr.illinois.edu/ece498rc3/fa2016/material/linearAlgebra_4pgs.pdf](https://courses.engr.illinois.edu/ece498rc3/fa2016/material/linearAlgebra_4pgs.pdf)

Pre-requisites

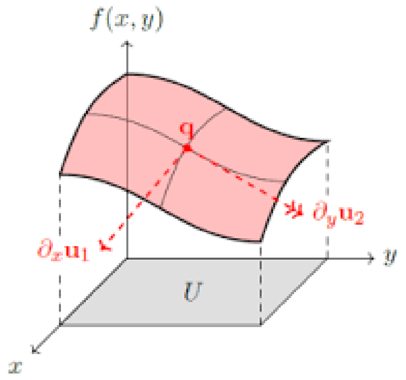
- Mathematics
  - Coordinate Geometry
    - Distance of point from a line
    - Distance between two parallel lines
  - Vector Calculus
    - Dot product , Projections



http://studyphysicswithme.com/blog/2016/11/07/vectors-vector-spaces/

Pre-requisites

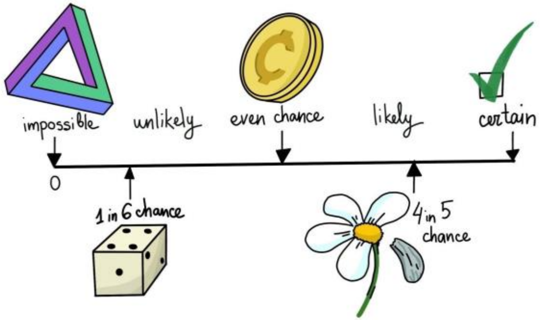
– Calculus
  - Derivative of single variable, y = f(x)
  - Partial derivative
  - Chain Rule
  - Gradient



http://tutorial.math.lamar.edu/getfile.aspx?file=B,41,N

Pre-requisites

- Probability
  - Axioms of probability
  - Sample Space, Event
  - Discrete, Continuous distributions
    - Uniform, Bernoulli, Geometric
    - Gaussian
  - Expectation of a random variable



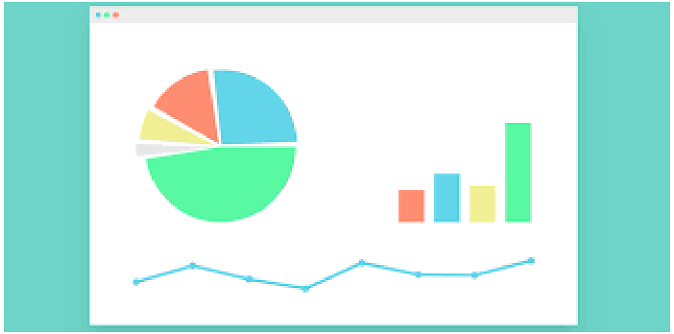
[Cheat-sheet: https://stanford.edu/~shervine/teaching/cme-106/](https://stanford.edu/~shervine/teaching/cme-106/)

http://www.wzchen.com/s/probability_cheatsheet.pdf

https://miro.medium.com/max/1200/1*YVL0hT1qsR8KPpkFDV2-wA.jpeg

Pre-requisites

- Statistics
  - Mean, Median, Mode
  - Standard Deviation



[Cheat-sheet: https://stanford.edu/~shervine/teaching/cme-106/](https://stanford.edu/~shervine/teaching/cme-106/)

Course Objectives

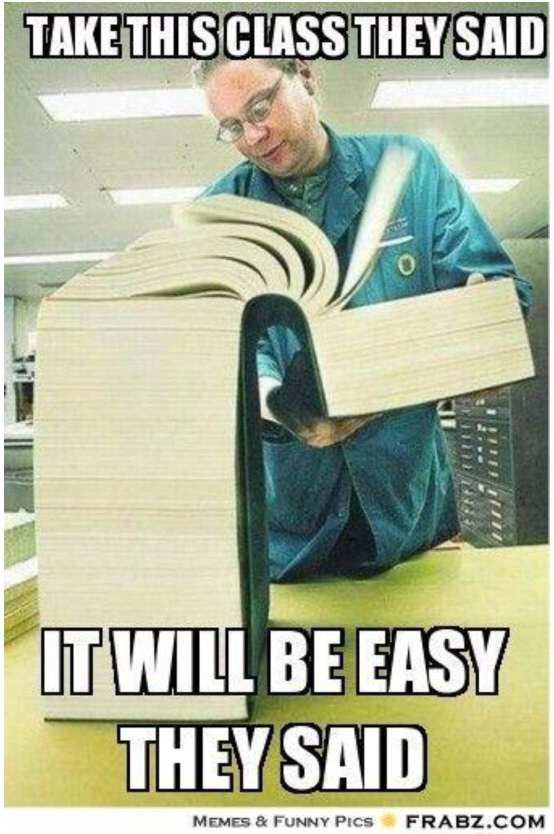
- Determine whether ML is suitable for a problem
- Formulate a problem as a ML problem (data ,representations, tasks, algorithms)
- **Understand** and apply ML method(s)
- Be aware of ML pitfalls, follow best practices
- Be ready to dive deeper (into ML theory or applied areas)

About the course - TAs

- TBA

About the course – Grading Policy

- Assessment
  - 1 Final Exam (35 %)
  - Assignments (35%)
  - 1 mid semester exam (25 %)
  - Scribe Class Notes (5%)



About the course - assignments

- Code
  - **MATLAB**
  - * Python (scikit-learn + jupyter notebook)
  - Neural Networks: TF, Pytorch, Keras

About the course – collaboration policy

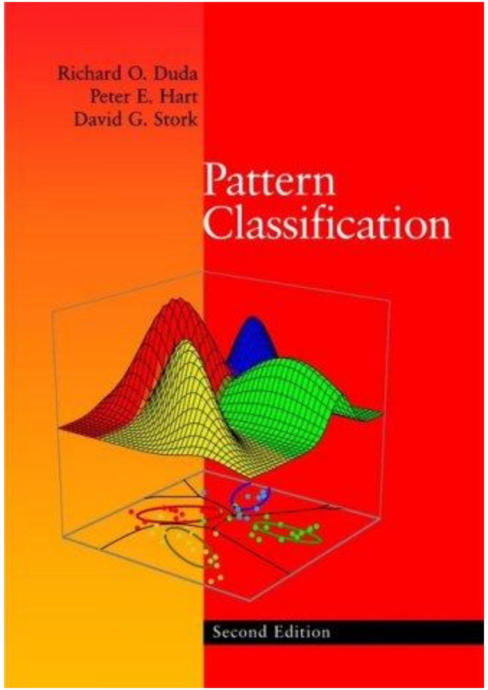
- OK to discuss assignment questions and approaches
- But work must be your own (no copying – partially or fully)
- If you worked with someone, mention their name(s)
- We will be checking for copying/plagiarism
- Better to own up than be caught !

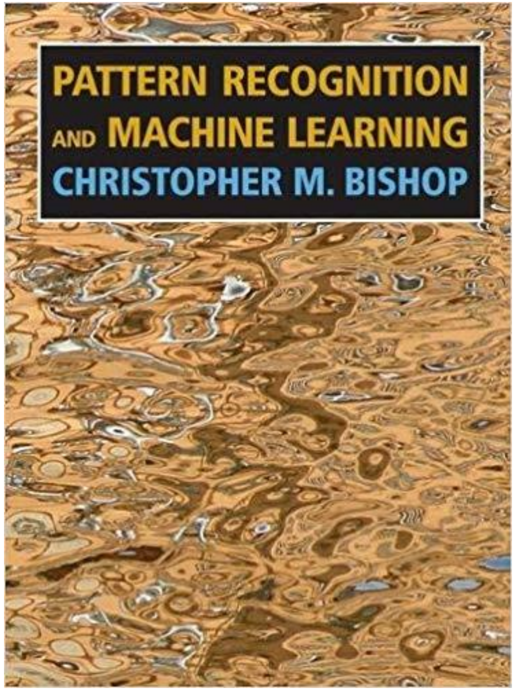
https://twitter.com/terago_networks/status/707283921323429889

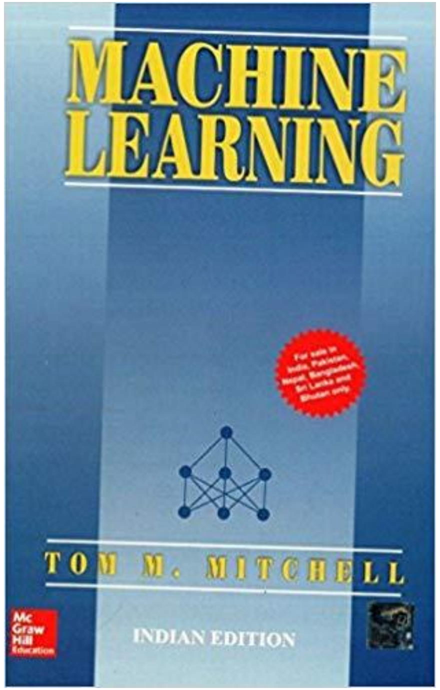
About the course – Grading Policy

- **Assignment Late Policy:** 50% if one day late; zero percent if more than one day late
- **A one-time late submission bonus:** With maximum of three days delay. You must adhere to standard late submission policy after using your bonus. No exceptions will be made. You'll need to inform TAs before assignment deadline if you wish to use the late submission bonus.

About the course - Textbooks







About the course - Material

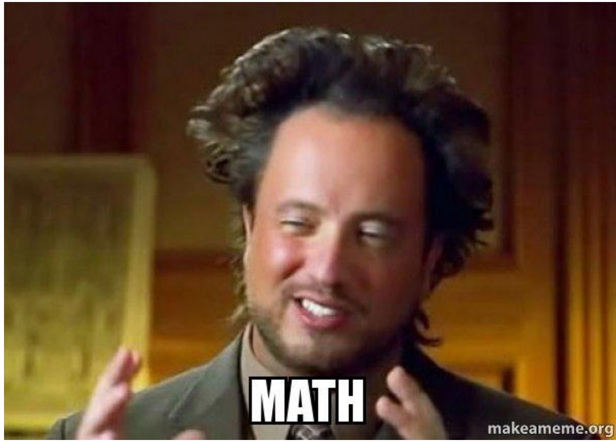
- Will be provided on a per lecture basis
- Scattered Resources across Internet

Survey

- For those **seriously** planning to take the course …

- Take the anonymous survey: https://forms.gle/dwJJdBuoeQXsDHCt5
- Deadline to submit survey: Monday 6th Jan 2020

- … Understand your background
- ... Will help tailor the course content



Additionally …

- **Understand**, don’t just memorize
- Love the math, not the toolbox !
- Capture the broad ideas and insights (useful years down the line)
- Implement ! No substitute for experience.
- Just the beginning ….

A tale of two airplanes

[“The Gimli Glider – 30 years later”](https://www.youtube.com/watch?v=3ffryZAd4Nw)
https://www.youtube.com/watch?v=3ffryZAd4Nw

[“Fatal Flight 447:Chaos in the Cockpit”](https://youtu.be/jM3CwBYX-ms)
https://youtu.be/jM3CwBYX-ms