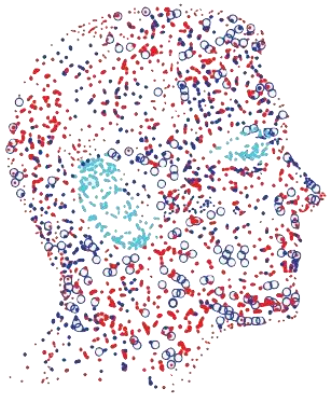
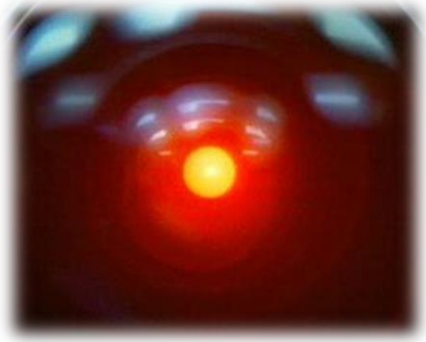
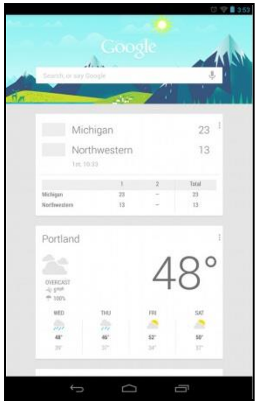




Machine Learning

CS 165B



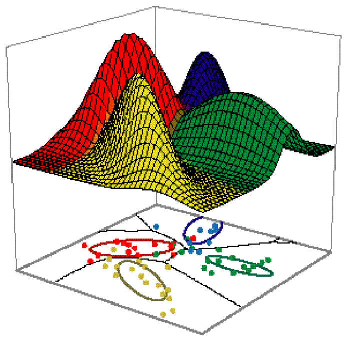
Spring Quarter 2016

AlphaGo

Matthew Turk

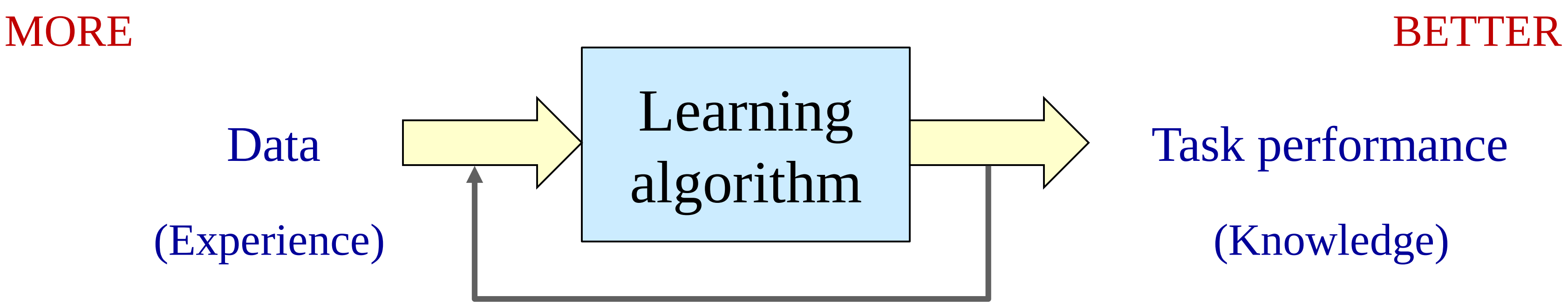
Professor, Computer Science and

Media Arts and Technology



Introduction to machine learning

Machine learning is *the design and analysis of algorithms that improve their performance at some task with experience*



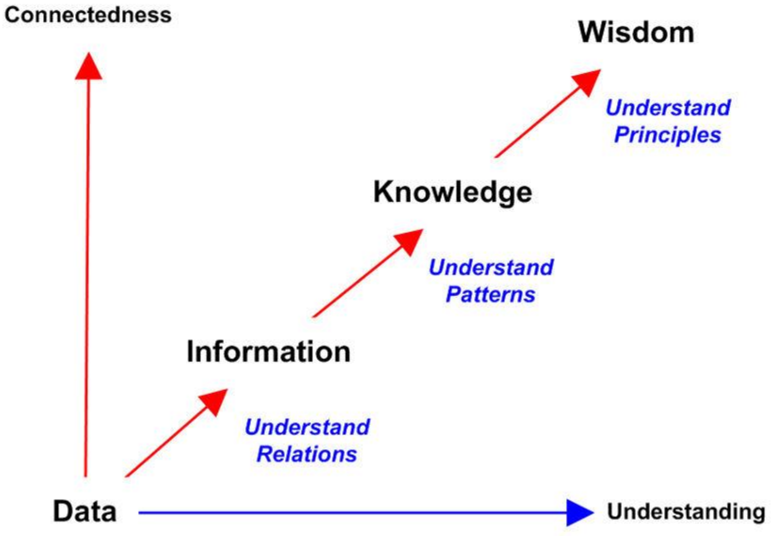
The learning may take place offline (before the execution of the task) – e.g., face detection, spam detection

It may take place online (as a task progresses) – e.g., an adaptive interface

It may do both – e.g., speech recognition systems

The big picture

- We wish to develop methods and tools for building learning machines that can solve problems in combination with available data sets of training examples
- Ultimately, to move up the "DIKW pyramid"
  - To be able to reveal principles, reveal directions, answer questions, make decisions, determine action
  - From understanding the past to empowering the future



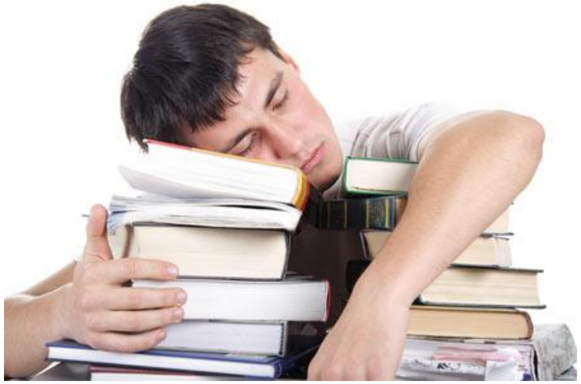
Big Data!

Some applications of machine learning

- Search engines
- Machine perception
- Computer vision, including object recognition
- Natural language processing
- Bioinformatics
- Syntactic pattern recognition
- Medical diagnosis
- Brain-machine interfaces
- Cheminformatics
- Detecting credit card fraud
- Stock market analysis
- Classifying DNA sequences
- Sequence mining
- Robot locomotion
- Speech and handwriting recognition
- Game playing
- Software engineering
- Adaptive websites
- Computational advertising
- Computational finance
- Structural health monitoring
- Sentiment analysis (or opinion mining)
- Affective computing
- Information retrieval
- Recommender systems

How do we learn?

When a machine improves its performance at a given task over time, without reprogramming, it can be said to have learned something



Some ways of learning for people:

- Rote learning, i.e., memorization or “muscle learning”
- Conditioning (associative learning)
- Learning from specific instructions
- Learning from explanations
- Learning from advice
- Learning from examples (episodic learning)
- Learning by doing
- Learning by exploration and discovery
- Learning by analogy

The typical machine learning problem

- Given a collection of examples (the training data), we want to build a model to predict something about novel examples

- For example:
  - Spam filtering
    - Is this email spam or ham?
  - Medical diagnosis
    - Do these test results indicate malignant or benign?
  - Natural language processing
    - What does the sentence ask for?
  - Face recognition
    - Who is this a picture of?
  - Chess
    - What move to make now?

Machine learning?

The output depends on the learning task!

- What's the next number in the sequence:
  - 1, 2, 4, 8, … **20**
  - "A big dog had the run of the land and fiercely protected it and the animals on it."

Best Friends:
A Bedtime Story

- How about this sequence:
  - 3, 4, 5, 2, … **5**
  - 3.14159265358979**3**
- Is this image in the same class as these images?

No, it's indoors



- What's the outlier in this group:
  - California, Virginia, Connecticut, France, Mississippi, Washington, Colorado, Pennsylvania, Massachusetts
  - Mississippi – I've lived in the other locations!

What is a learning problem?

- Learning involves improving performance
  - at some task ***T***
  - with experience ***E***
  - evaluated in terms of performance measure ***P***

- Example: learn to play checkers
  - Task ***T***: playing checkers well
  - Experience ***E***: playing against itself
  - Performance ***P***: percent of games won against humans

- What exactly should be learned?
  - How might this be represented?
  - What specific algorithm(s) should be used?

Components of a learning problem

- Task: the behavior or task that’s being improved; e.g., classification, object recognition, acting in an environment
- Data: the experiences that are being used to improve performance in the task
- Measure of performance: How can the improvement be measured? Examples:
  - Provide more accurate solutions (e.g., increasing the accuracy in prediction)
  - Cover a wider range of problems
  - Obtain answers more economically (e.g., improved speed)
  - Simplify codified knowledge
  - New skills that were not presented initially

Machine learning ingredients

- Prior assumptions
  - What do we know a priori about the problem?
- Data
  - What kind of data do we have?
- Representation
  - How do we represent the data?
- Model / hypothesis space
  - What hypotheses are we willing to entertain to explain the data?
- Feedback / learning signal
  - What kind of learning signal do we have (labels, delayed)?
- Learning algorithm
  - How do we update the model (or set of hypotheses) from feedback?
- Evaluation
  - How well did we do? Should we change the model?

Key types of machine learning

- Supervised learning
  - Provide *labeled* training data
  - Give the correct answers – input/output pairs
- Semi-supervised learning
  - Provide *some* labeled training data, other data unlabeled
  - Give *some* correct answers, others unknown
- Reinforcement learning
  - Provide occasional, usually delayed, information or reward
  - E.g., win or lose game (but no feedback on individual moves)
- Unsupervised learning
  - No direct learning signal or labels
  - The task is typically to find structure in the data (e.g., clustering, dimensionality reduction)

Some machine learning problems

Training data includes desired outputs (input/output pairs)

Training data does not include desired outputs

Outputs		Supervised learning	Unsupervised learning
	Discrete	Classification / labeling Binary / multi-class	Clustering
	Continuous	Regression Real-valued prediction	Dimensionality reduction

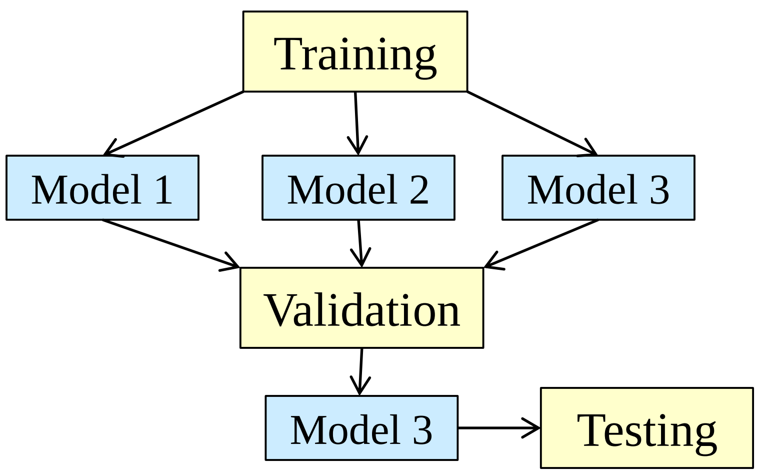
Also <u>semi-supervised learning</u>, <u>reinforcement learning</u>, etc.

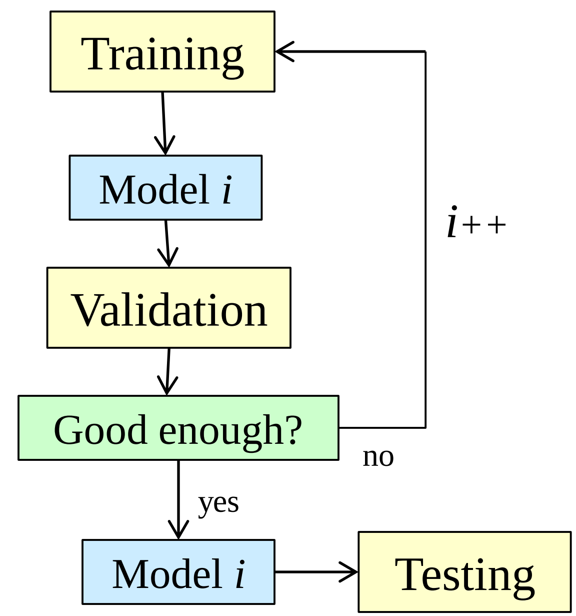
Training data includes *some* desired outputs

Learning through trial-and-error interactions with the environment

Training, validation, and test datasets

- We typically divide the dataset into three subsets:
  - Training data is used for learning the parameters of the models
  - Validation data is used to decide which model to employ
  - Test data is used to get a final, unbiased estimate of how well the model works

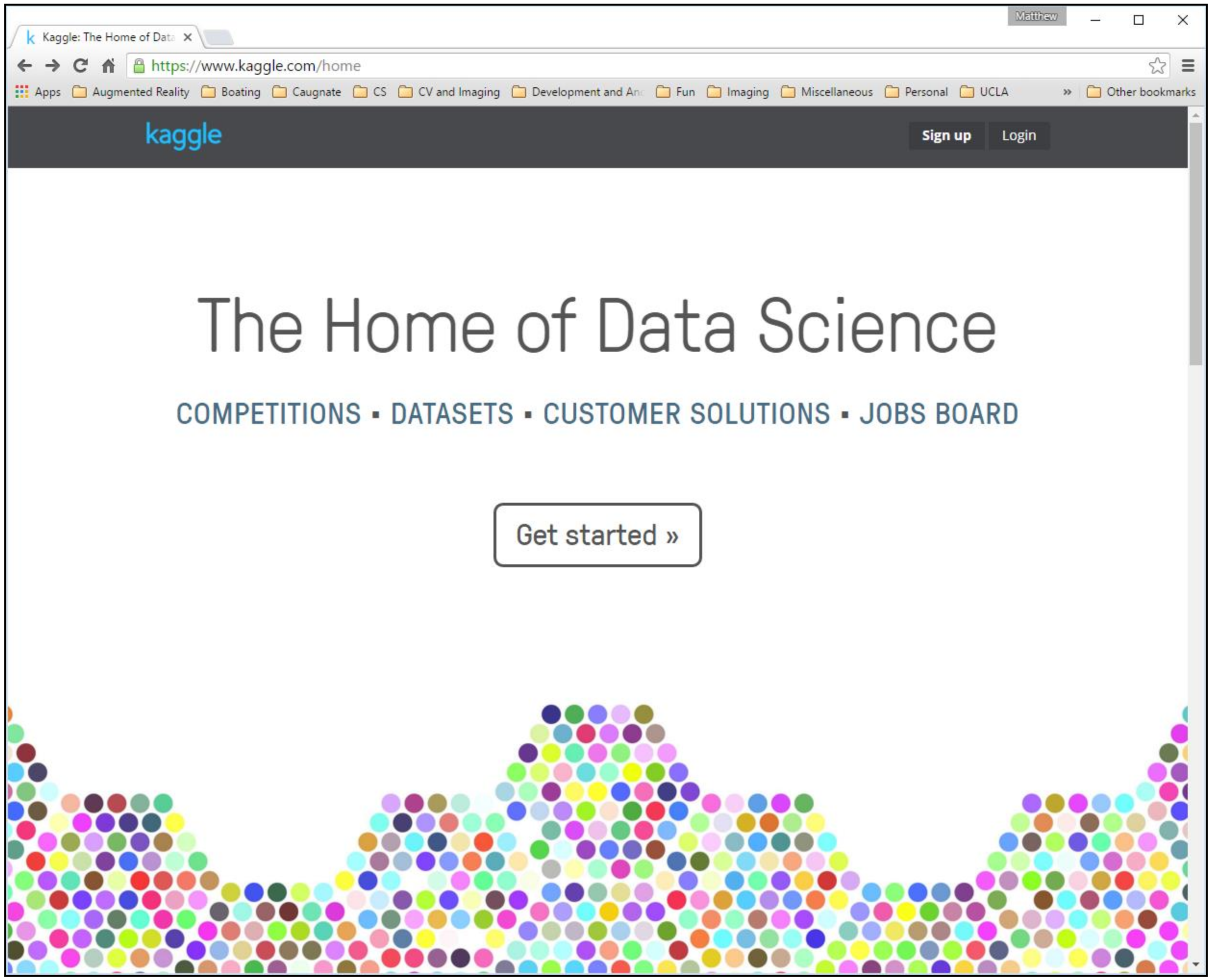




Sometimes reduced to training and testing
a single mode (no validation step)

Kaggle

Data science competitions



The Netflix Prize

- The Netflix Prize sought to substantially improve the accuracy of predictions about how much someone is going to enjoy a movie based on their movie preferences
- $1M grand prize awarded in 2009
- Provided teams with:
  - Anonymous rating data
  - A prediction accuracy bar that was 10% better than what Netflix could do on the same training data set

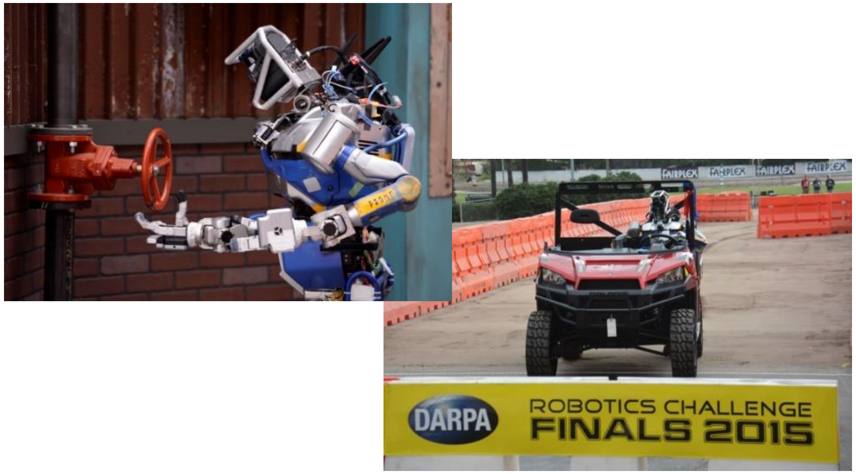
Winner: BellKor's Pragmatic Chaos

DARPA Grand Challenges

2004, 2005 Grand Challenges

2007 Urban Challenge

2012-2015 Robotics Challenge

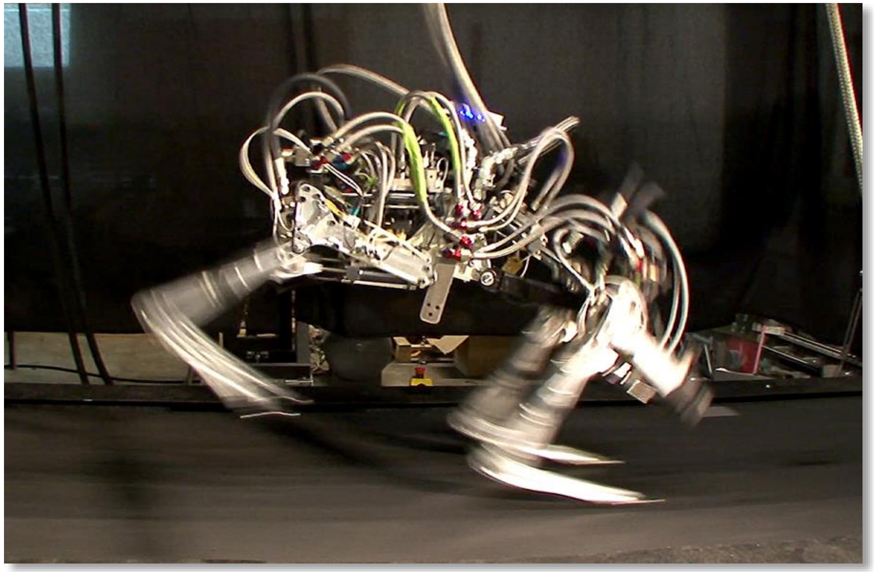


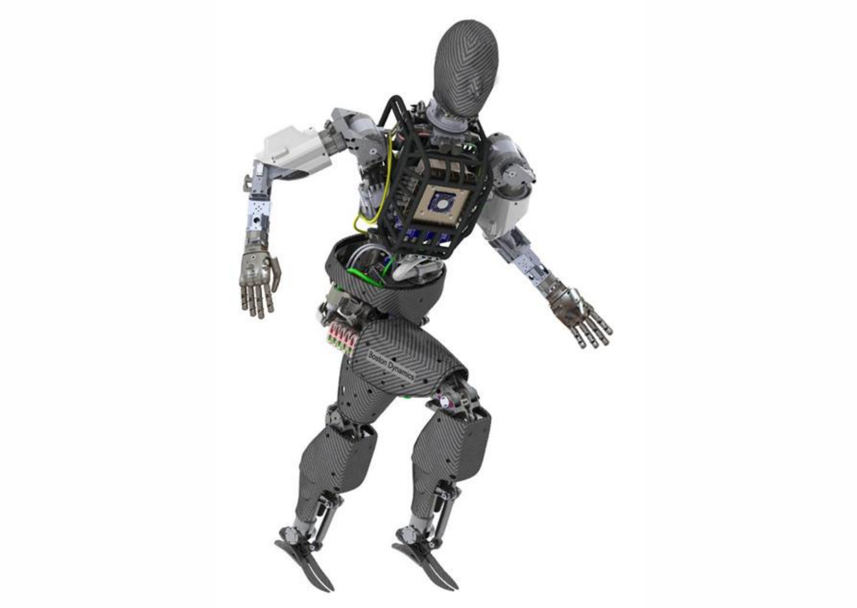
2013 FANG Challenge

Fast Adaptable Next-Generation Ground Vehicle

Boston Dynamics

http://www.youtube.com/BostonDynamics





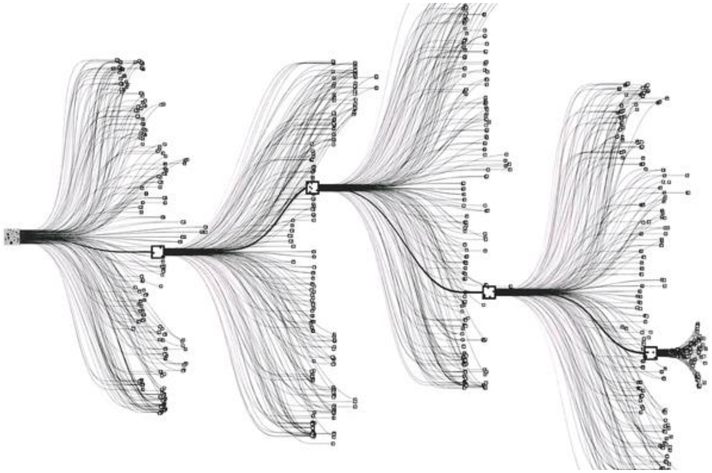
Google AlphaGo

AlphaGo

- The first computer program to beat a professional player at the game of Go

- Defeated top Go player in the world, Lee Sedol, in March 2016, 4-1

- Uses deep neural networks that are trained by a novel combination of supervised learning from human expert games, and reinforcement learning from games of self-play.
  - DeepMind Technologies acquired by Google in 2014



165B course web site

- http://www.cs.ucsb.edu/~mturk/ML
  - Home page: info and announcements
  - Syllabus
  - Schedule
    - Lecture notes
      - Need userid and password: **`machine | learning`**
  - Assignments
  - Links

- Bookmark this and visit often!
- Will also use the course GauchoSpace page
  - Primarily for assignments and grading
  - Piazza also

How to succeed in 165B

- Do the assigned reading
  - The lectures will overlap somewhat with the reading, but nowhere near 100% – both lecture and assigned reading are critical
  - The textbook supplements the lecture (providing additional depth, examples, and topics not mentioned in class)
  - The lecture supplements the textbook (providing additional depth, examples, and topics not mentioned in the textbook)
- Attend the discussion sessions
  - Various topics: review material (like this week), more details on lecture material, algorithms examples, homework help, questions, etc.
- Be engaged in class
  - Ask questions, offer feedback, laugh at my jokes….
  - Review the lecture notes after class (posted soon <u>after</u> class ends)
- Get started on the homework assignments early

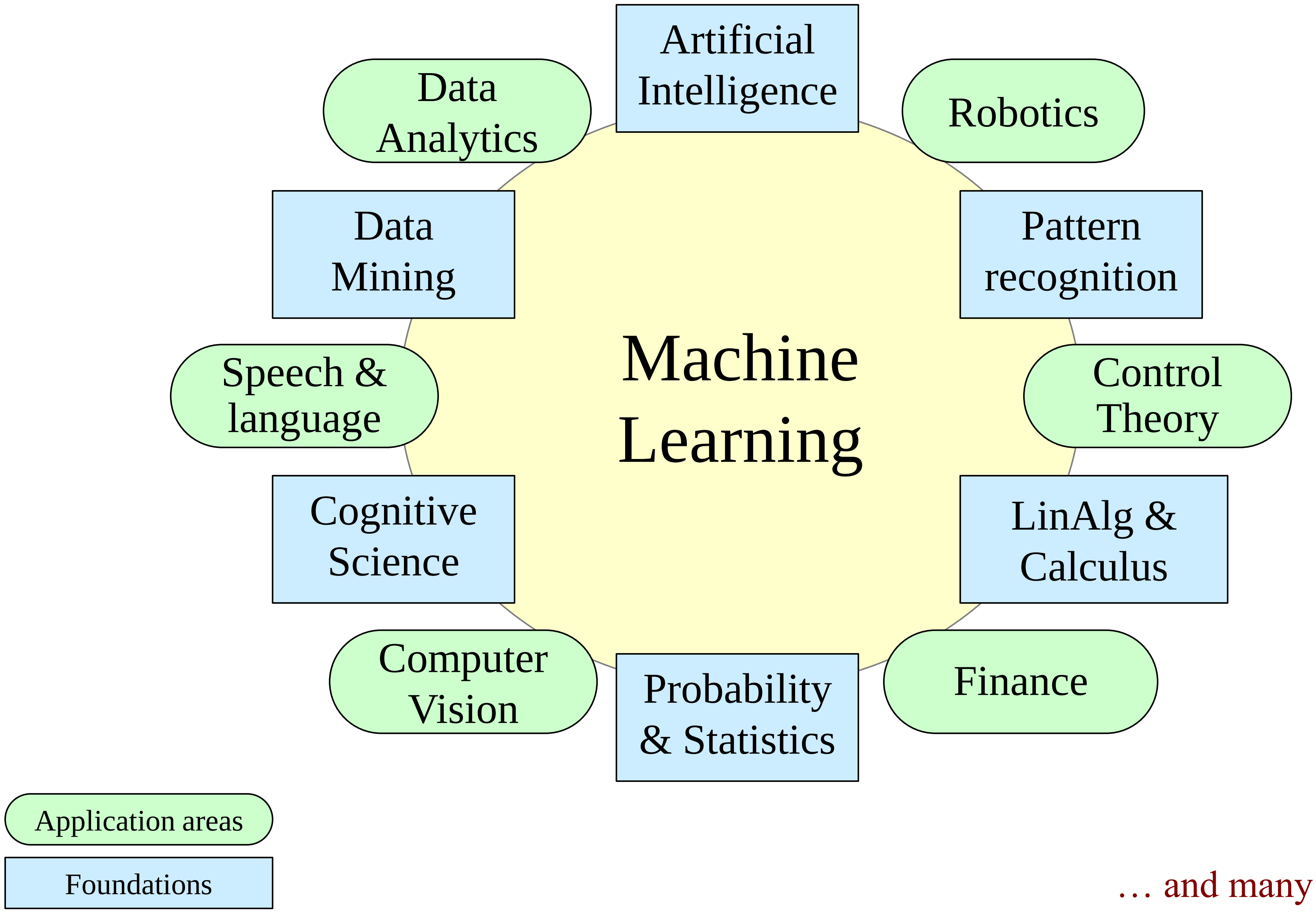
Machine learning is the systematic study of algorithms and systems that improve their knowledge or performance with experience.

What is machine learning?

“machine” + “learning”

Algorithms Programs Systems	…that…	Make sense of data Learn from data Improve with experience Adapt to the user/situation

Related topics



… and many more!

Classical AI vs. machine learning approach

Classical AI

- Think about the world
- Write down rules that encapsulate intelligent behavior
- Hope that these adequately cover the range of real-world situations

Deductive reasoning

Machine learning

- Collect massive amounts of data
- Provide an architecture for learning (primed with some goals)
- Hope that the data can be abstracted into meaningful general concepts

Inductive reasoning

Deductive and inductive learning

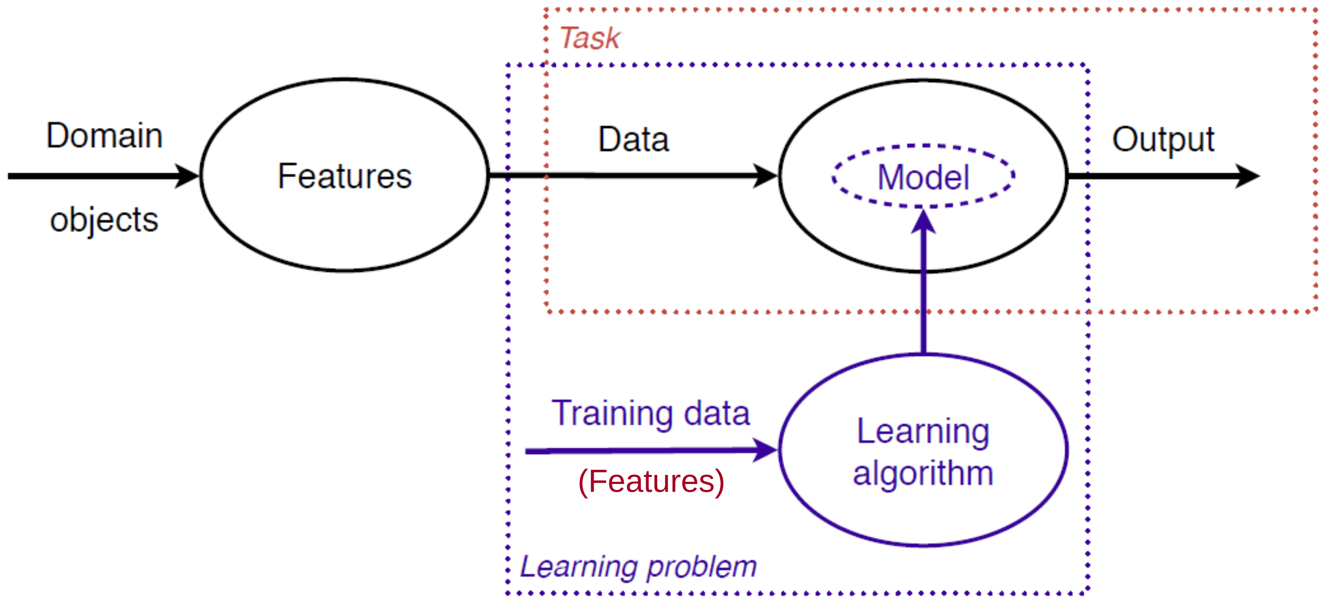
- Deductive Learning: Learning from a set of known facts and rules to produce additional rules or conclusions that are guaranteed to be true.
  - From general knowledge to specific knowledge
  - E.g., modus ponens

- Inductive Learning: Learning from a set of examples to produce a general rules. The rules should be applicable to new examples, but there is no guarantee that the result will be correct.
  - Generalizing from specific examples
  - I.e., learning from experience

Some key machine learning concepts

Textbook prologue

A machine learning system



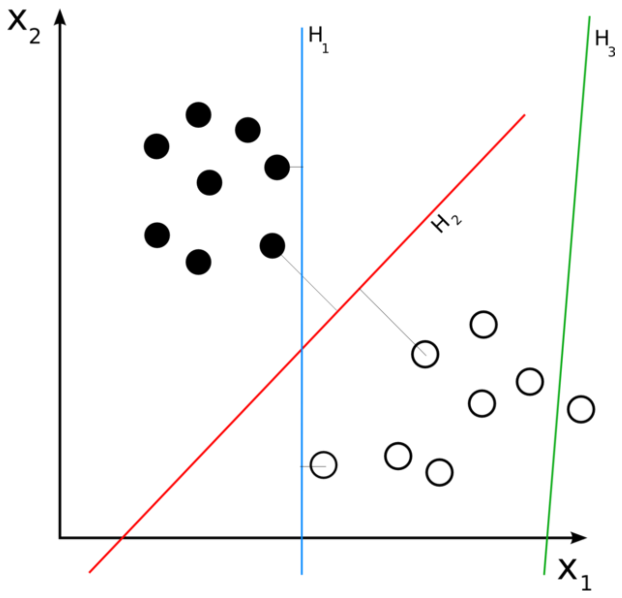
A **task** requires an appropriate mapping – a **model** – from data described by **features** to outputs. Obtaining such a model from training data is what constitutes a **learning problem**.

Tasks are addressed by **models**.

Learning problems are solved by **learning algorithms** that produce **models**.

Linear classification

- Outputs a classification (one of N possible classes) based on the value of a linear combination of the characteristics (features)

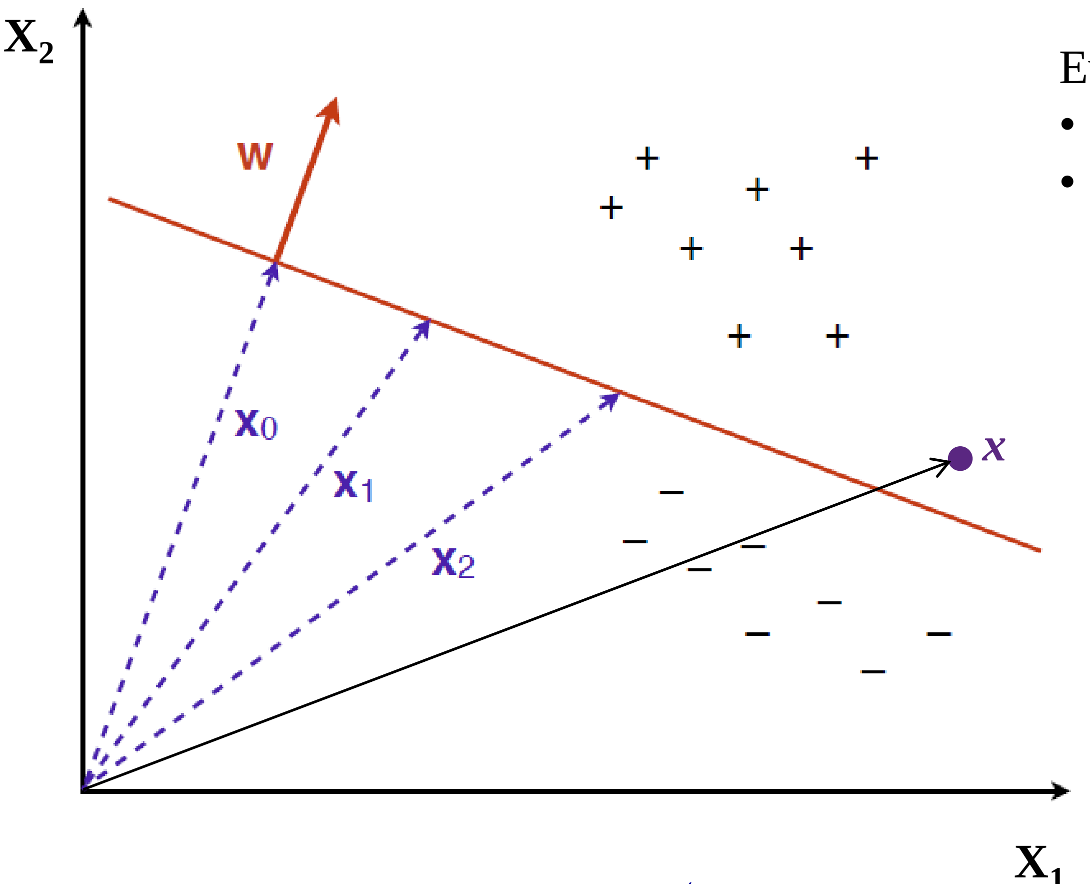


Example with 2 features – 2D feature vector (x_1, x_2):

- Linear classifiers H_1 and H_2 successfully partition the two classes of dots
- H3 does not

Which is better, H_1 or H_2? Why?

Linear classification



$$x_0 \cdot w = x_1 \cdot w = x_2 \cdot w = t$$

$$x \cdot w = x^T w$$

How to determine if a feature vector ***x*** is on the + or – side of the line?

Evaluate the dot product of ***x*** and ***w***:

- If $x \cdot w > t$, then +
- Otherwise –

2 features means 2D classification and a 1D classification boundary

N features means N dimensional classification and an N-1 dimensional classification boundary

Dimensions	Linear boundary
1	Point
2	Line
3	Plane
>3	Hyperplane