CE9010: Introduction to Data Science

Lecture 2: Learning Techniques

Semester 2 2017/18

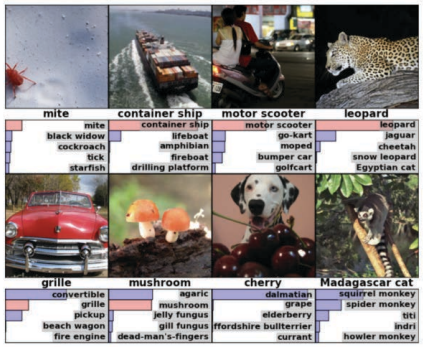
Xavier Bresson

School of Computer Science and Engineering
Data Science and AI Research Centre
Nanyang Technological University (NTU), Singapore

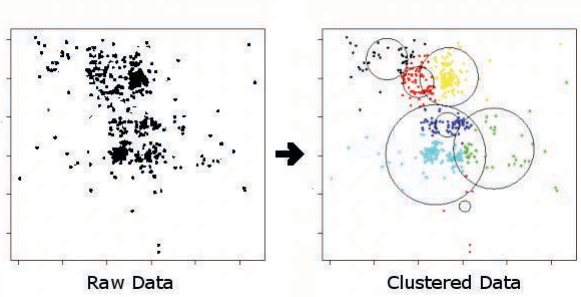
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Learning techniques

- Supervised learning (recent breakthrough 2012-)

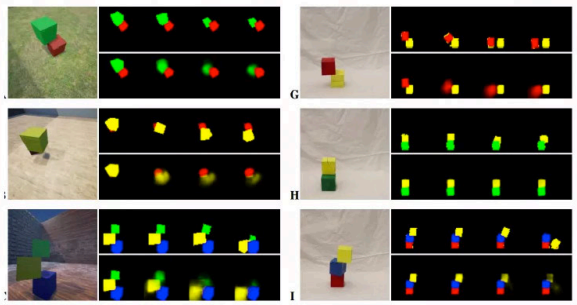


- Unsupervised learning (still incomplete)



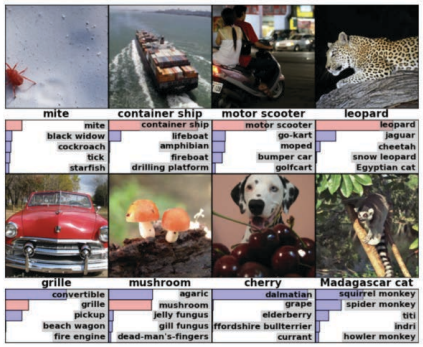
- Reinforcement learning (recent breakthrough 2015-)

- Predictive learning (main obstacle to general AI)

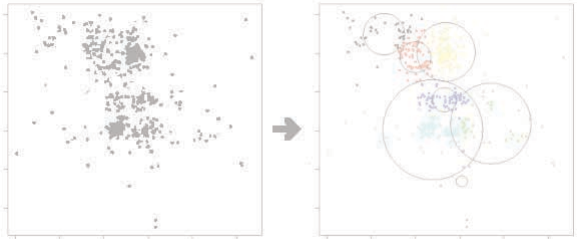


Learning techniques

- **Supervised learning (recent breakthrough 2012-)**

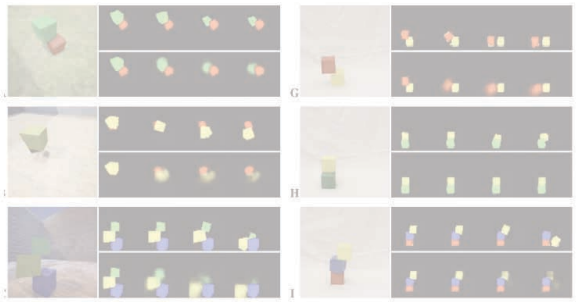


- Unsupervised learning (still incomplete)



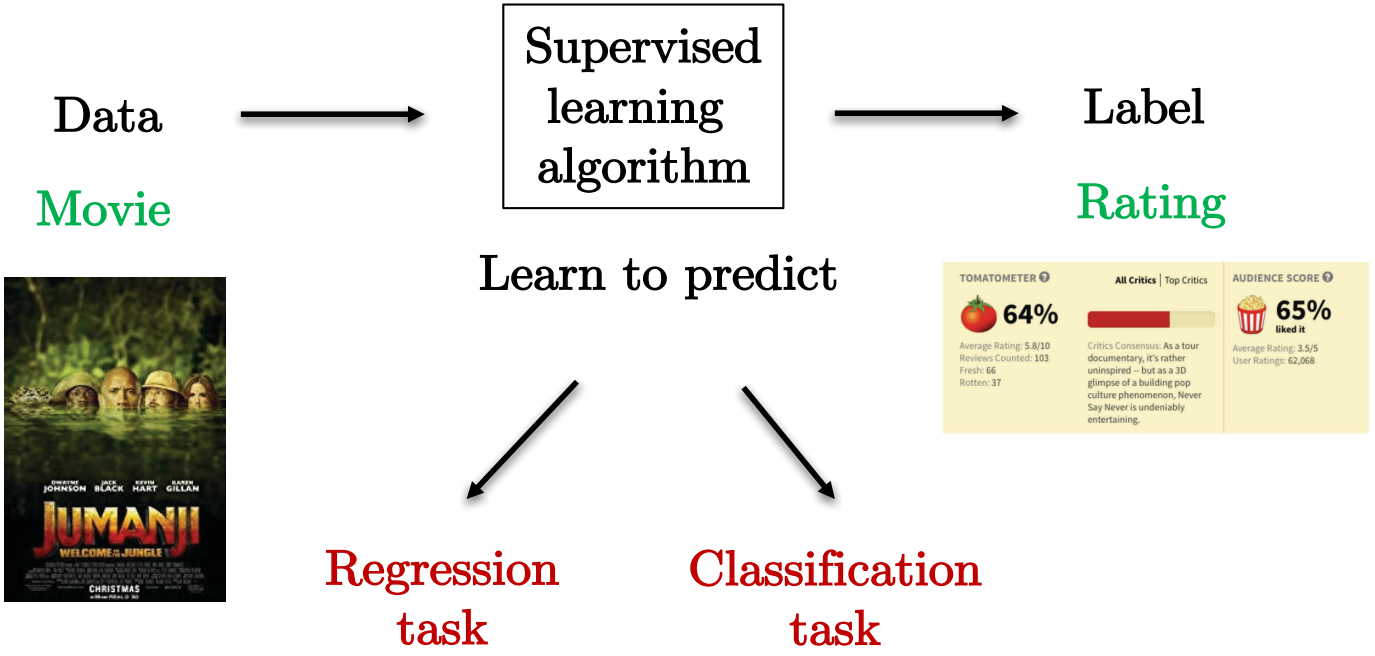
- Reinforcement learning (recent breakthrough 2015-)

- Predictive learning (main obstacle to general AI)



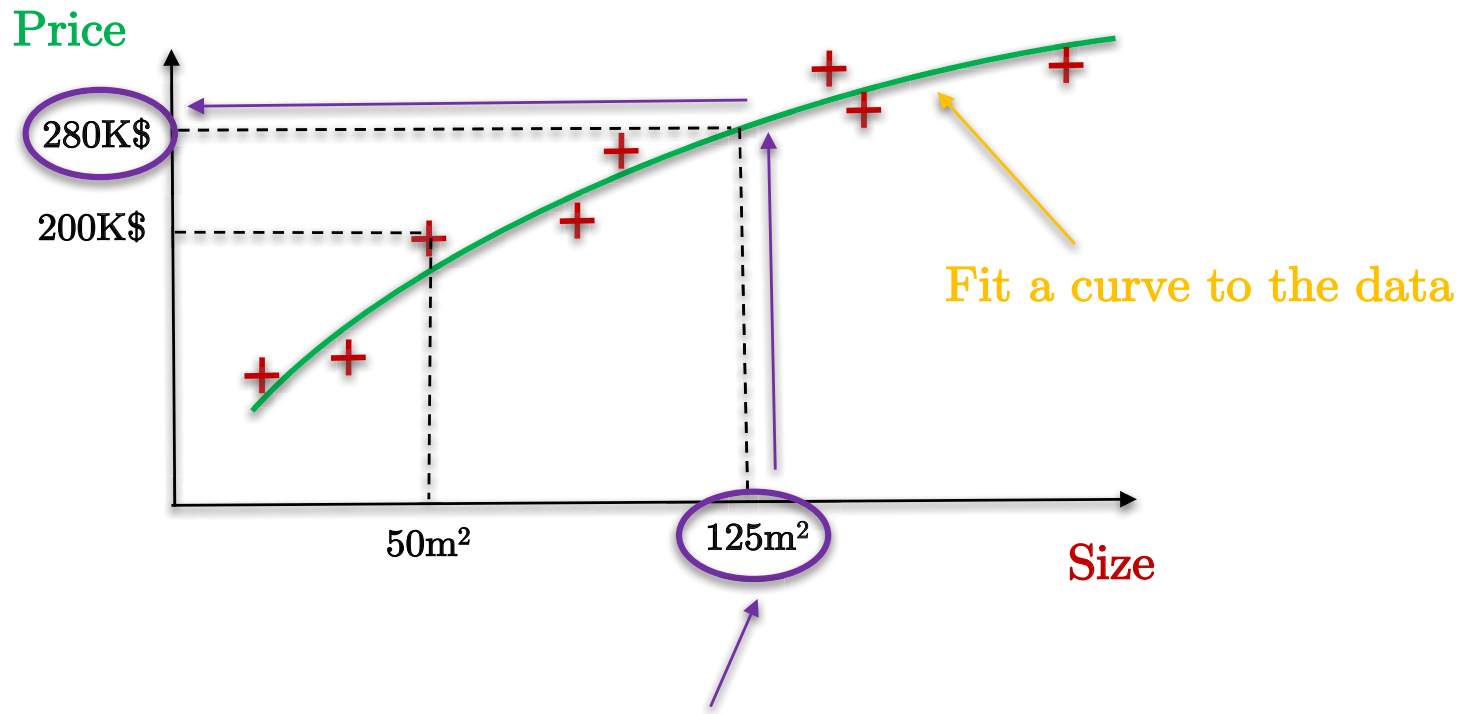
Supervised learning

- Definition: Given a set of data and labels (some properties of the data), design a learning algorithm to predict label for new data (never observed before).



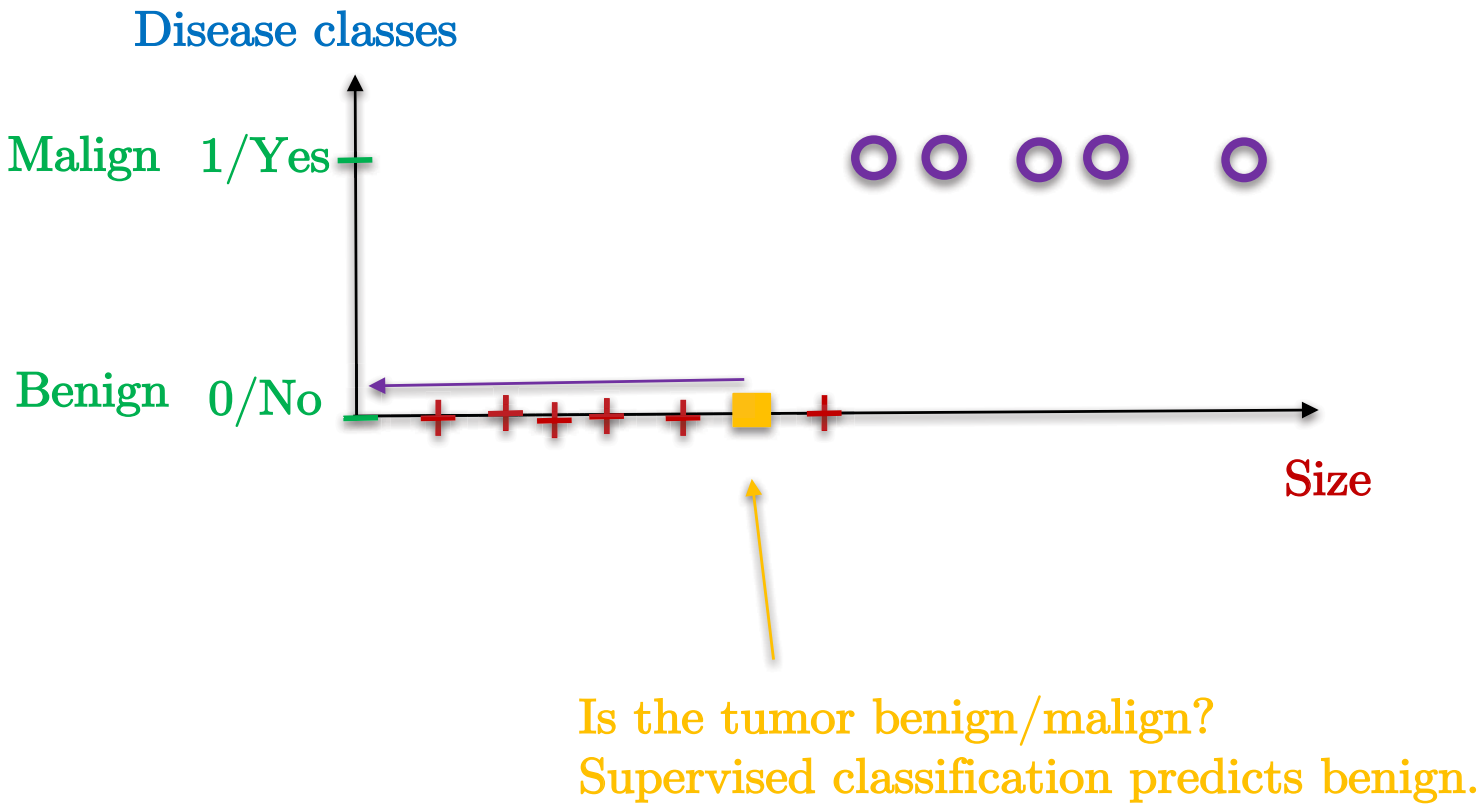
Regression: Housing price prediction

- Predict the price (continuous value) of houses given existing data label (house size).



Classification: Disease prediction

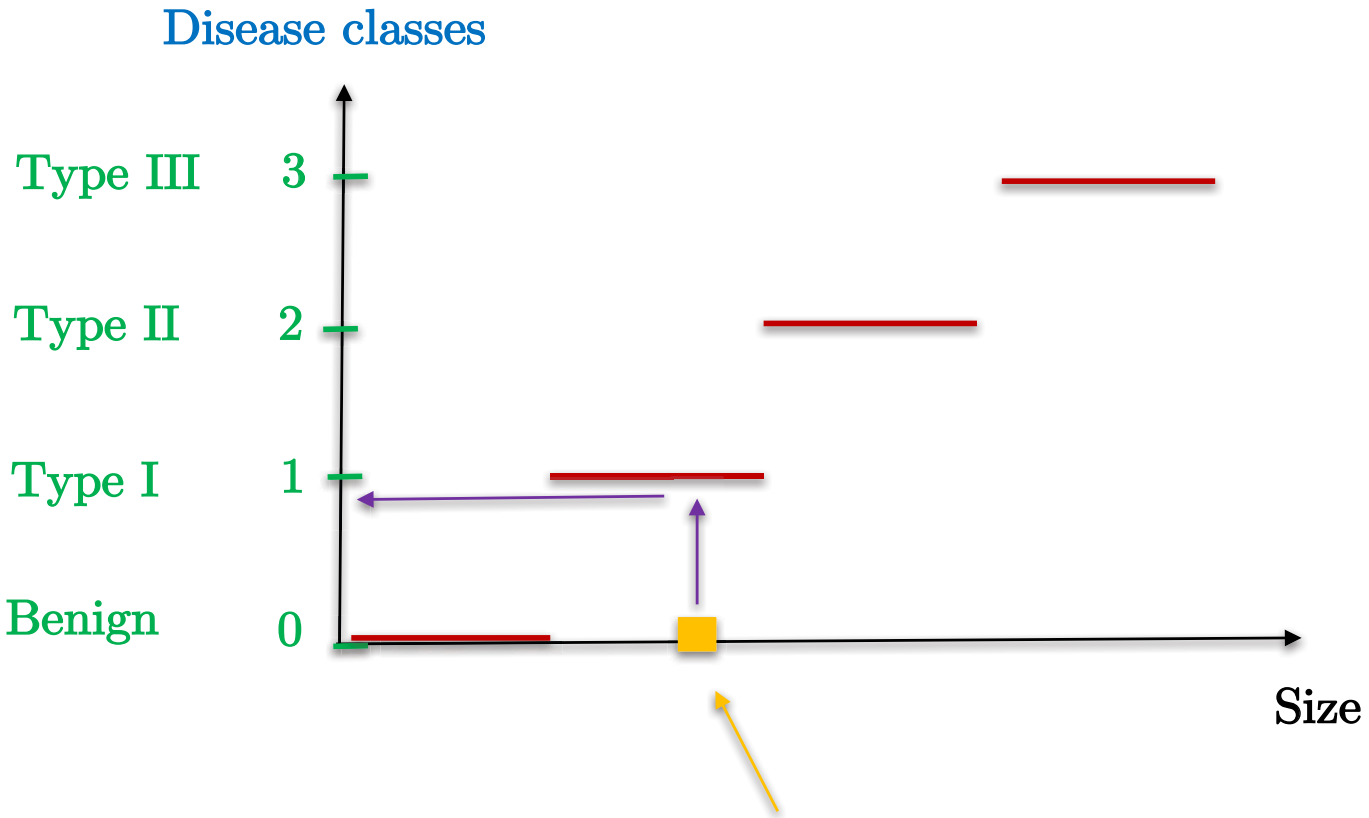
- Predict disease class (discrete value) of patient given existing medical data (tumor size).



Multiple classes

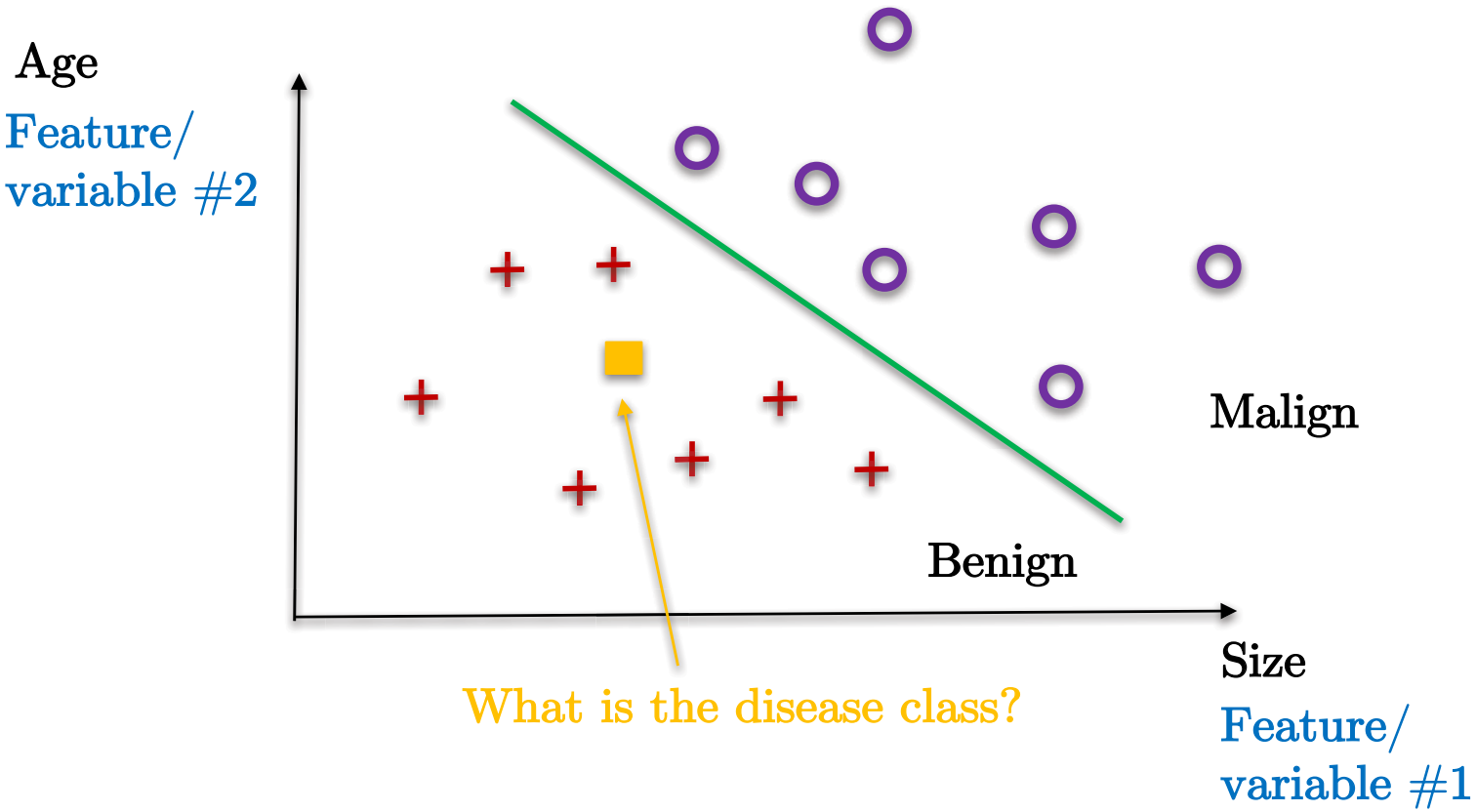
- From binary class to multi-class prediction

  Example: Disease classes can be {benign, type I, type II, type III}.



Multiple features

- Data have 2 features:



- Data dimension is the number of features (here dim(data)=2).

High-dimensional learning

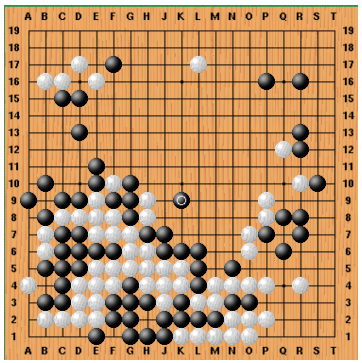
- Generalization: Data with N>2 features, for examples 1K, 1M features ⇒ Curse of dimensionality (too many variables)

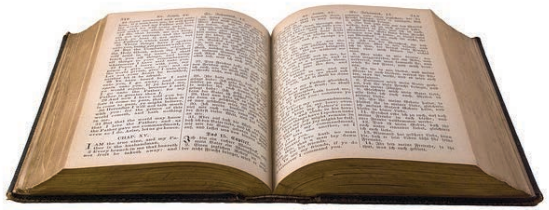
- Examples

  dim(Go) = 19 x 19 ≈ 10^3

  dim(Images) = 512 x 512 ≈ 10^6

  dim(Books) = 80,000 ≈ 10^5





Quiz

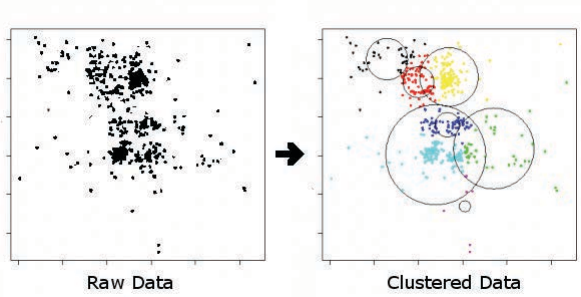
- You're running a company, and you want to develop learning algorithms to address each of two problems.
  - Problem 1: You have a large inventory of identical items. You want to predict how many of these items will sell over the next 3 months.
  - Problem 2: You'd like software to examine individual customer accounts, and for each account decide if it has been hacked/compromised.

- Should you treat these as classification or as regression problems?
  - Treat both as classification problems.
  - Treat problem 1 as a classification problem, problem 2 as a regression problem.
  - Treat problem 1 as a regression problem, problem 2 as a classification problem.
  - Treat both as regression problems.

Learning techniques

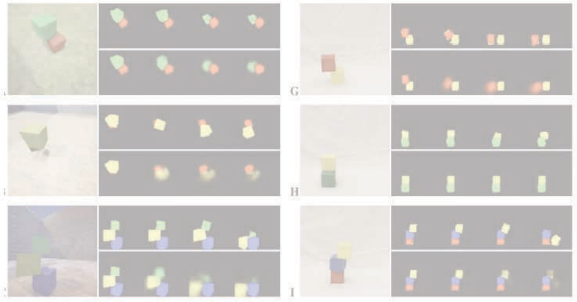
- Supervised learning (recent breakthrough 2012-)

- **Unsupervised learning (still incomplete)**



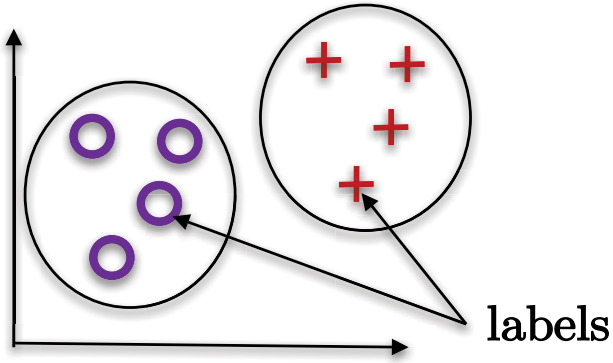
- Reinforcement learning (recent breakthrough 2015-)

- Predictive learning (main obstacle to general AI)

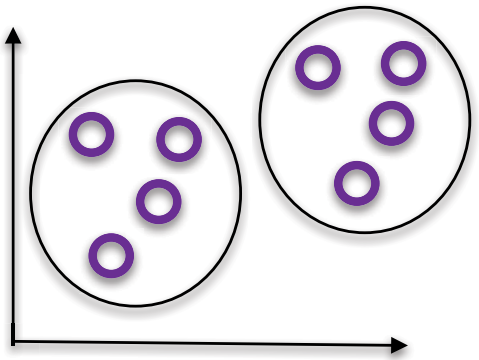


Unsupervised learning

- Difference between supervised and unsupervised learning:



Supervised learning
{data,label}



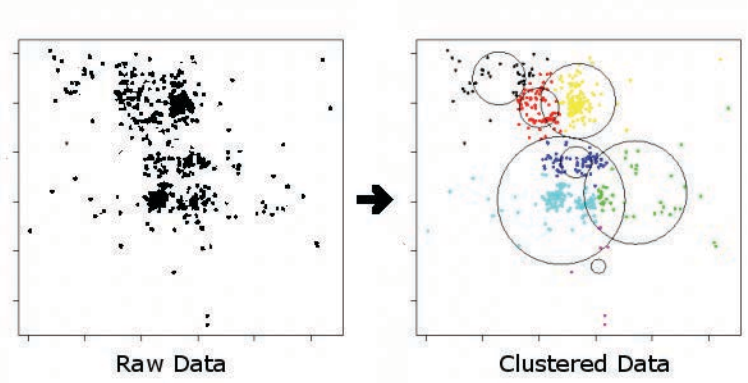
Unupervised learning
{data,-}

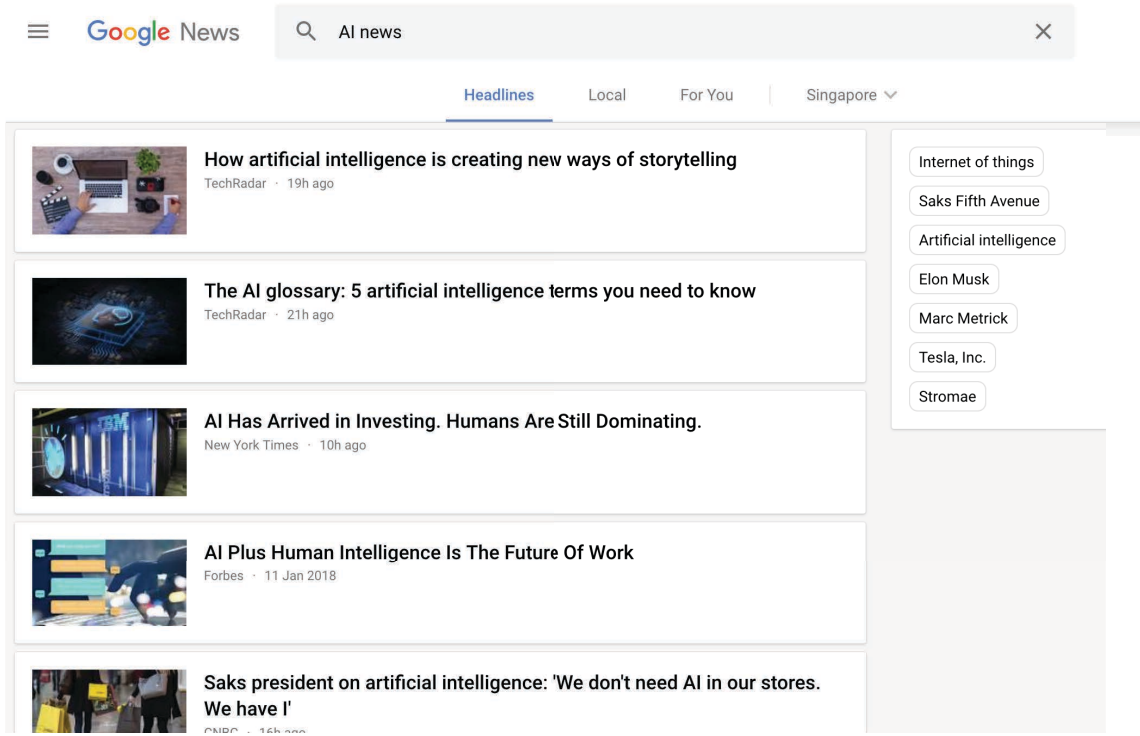
No label, no additional information

- Definition: Find structures in data that can solve tasks like clustering, classification, compression, generative model.

Example 1: Data clustering

- Definition: Group similar data into groups
- Example: News articles (Google News)





Example 2: Google engine

- Google PageRank recommender system:

Artificial intelligence - Wikipedia
https://en.wikipedia.org/wiki/Artificial_intelligence
Artificial intelligence is intelligence displayed by machines, in contrast with the natural intelligence (NI) displayed by humans and other animals. In computer science **AI** research is defined as the study of "intelligent agents": any device that perceives its environment and takes actions that maximize its chance of success at ...

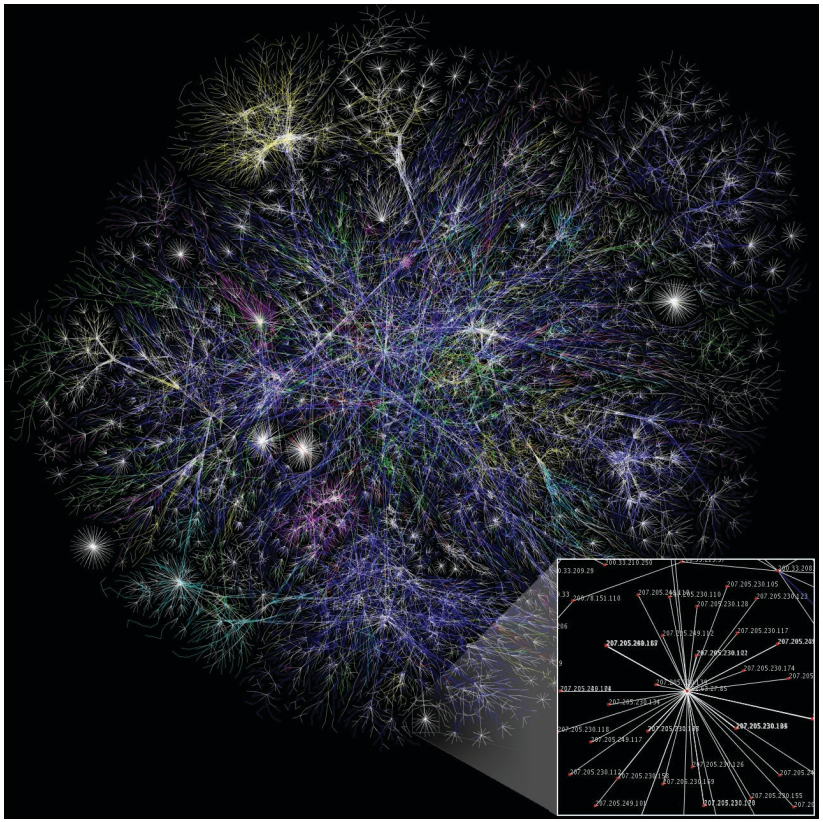
Google.ai
https://ai.google/
At Google, we think that **AI** can meaningfully improve people's lives and that the biggest impact will come when everyone can access it. Learn more about our projects and tools.

AI Apprenticeship Programme - AI Singapore
https://www.aisingapore.org/aiap/
About the **AI** Apprenticeship Programme (AIAP). The 1st TechSkills Accelerator (TeSA) company-led training programme in Artificial Intelligence (**AI**) . TeSA is an initiative of SkillsFuture. AIAP is a 9 months of full-time structured training programme comprising: 3 months of **AI** coursework consisting of classrooms, online, ...

AI Singapore: Home
https://www.aisingapore.org/
AI SINGAPORE is a national initiative to anchor deep capabilities in Artificial Intelligence, thereby creating social and economic impact, growing local talent, building an **AI** ecosystem, and putting Singapore on the world map.

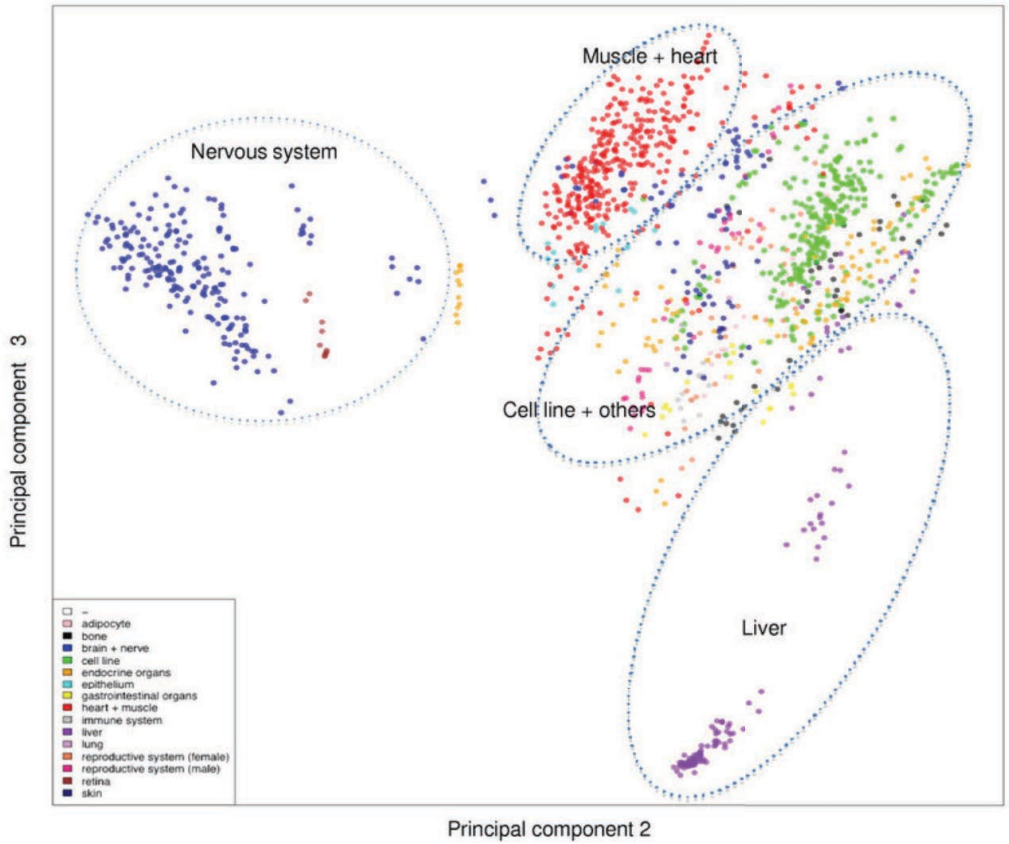
Artificial Intelligence (AI) – Hacker Noon
https://hackernoon.com/artificial-intelligence/home
Artificial Intelligence (**AI**) is intelligence exhibited by machines. Hacker Noon publishes **AI** trends & stories by tech professionals who build & invest in **AI**. how hackers start their afternoons.

Ai | TED.com
https://www.ted.com/topics/ai
A collection of TED Talks (and more) on the topic of **AI**.



Example 3: Data analysis

- Gene expression with PCA (unsupervised data representation):



Example 4: Data generation

- Generative Adversarial Networks (GANs):

NVIDIA'17

Quiz

- Of the following examples, which would you address using an unsupervised learning algorithm?
  - Given email labeled as spam/not spam, learn a spam filter.
  - Given a set of news articles found on the web, group them into set of articles about the same story.
  - Given a database of customer data, automatically discover market segments and group customers into different market segments.
  - Given a dataset of patients diagnosed as either having diabetes or not, learn to classify new patients as having diabetes or not.

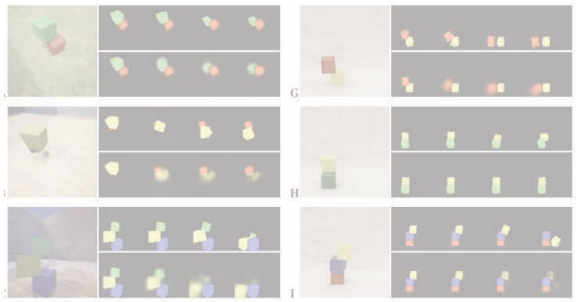
Learning techniques

- Supervised learning (recent breakthrough 2012-)

- Unsupervised learning (still incomplete)

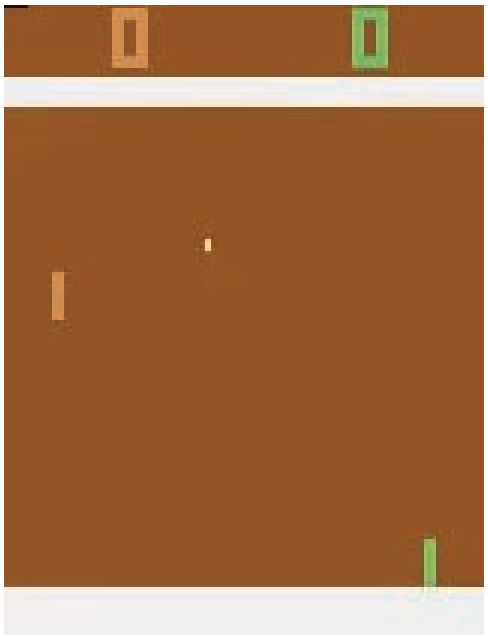
- **Reinforcement learning (recent breakthrough 2015-)**

- Predictive learning (main obstacle to general AI)



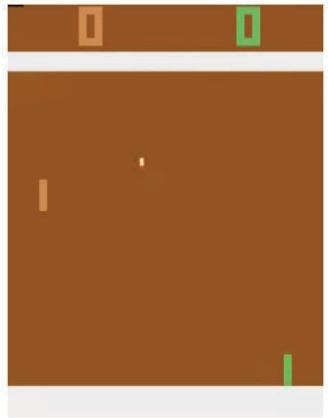
Atari Pong

- AI gamer:

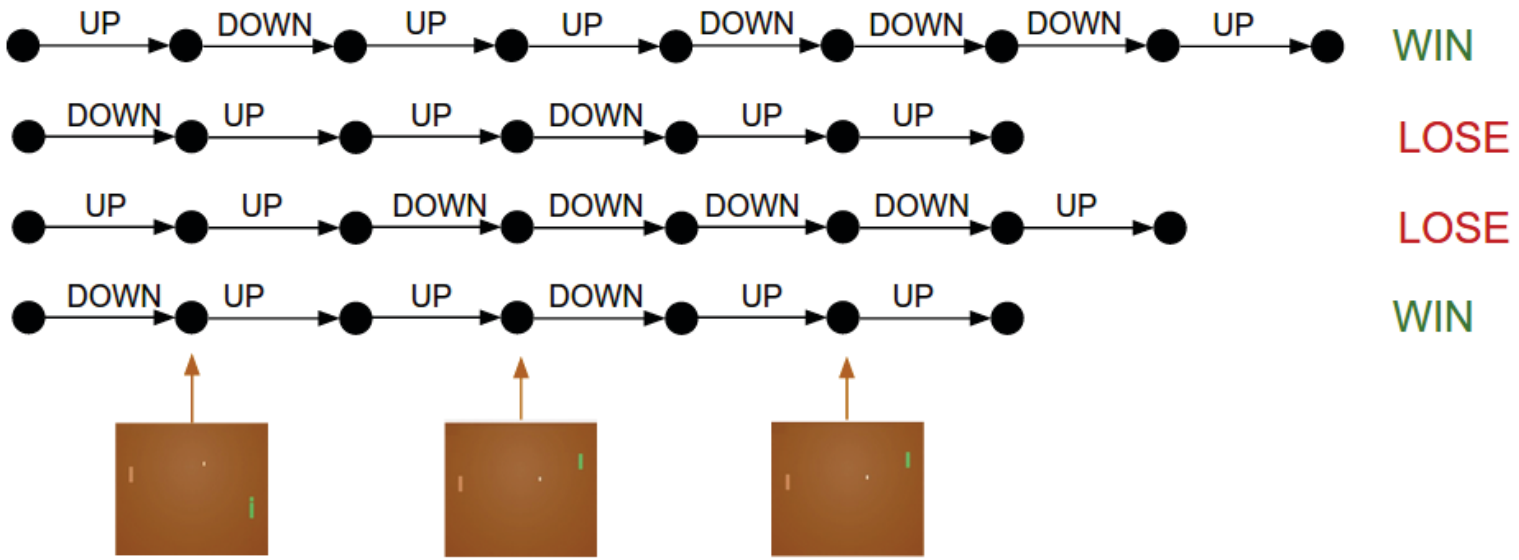


- Move Up or Down
- Reward:
  - +1 if the ball makes it past the opponent (winning game)
  - -1 if we missed the ball (lost game)
- Goal: Move paddle to maximize reward

Atari Pong



- Actions:
  - Move Up or Down
- Reward:
  - +1 if the ball makes it past the opponent (winning game)
  - -1 if we missed the ball (lost game)
- Goal: Move paddle to maximize reward



Examples

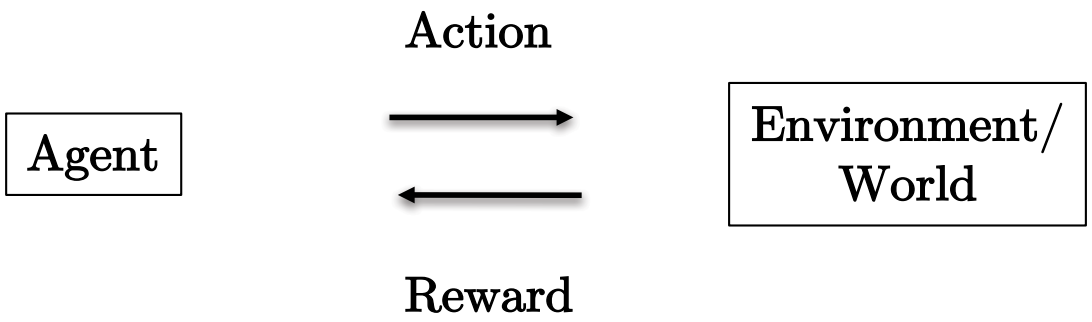
- AI Gaming (Doom), Go

Examples

- Robots

Reinforcement learning

- Definition: Learn a sequence of actions that maximizes a reward.



- RL agents learn to plan the future to win.

- Big issue: Lots of data required because the reward is sparse (more difficult than supervised learning because at each action, a positive reward/label is given).

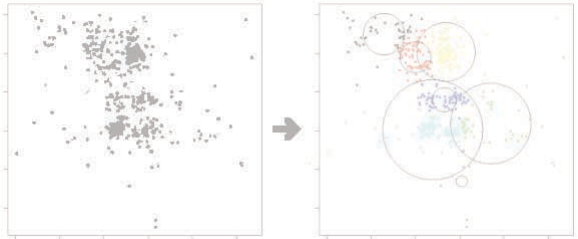
Quiz

- Which of these examples use supervised learning and reinforcement learning?
  - Given a robot that can move up, down, left, right in known mazes, develop a strategy to find the exit in a new unseen maze.
  - Given the chess game and a professional player who evaluates and provides all moves for several games, design a technique to win new games.
  - Given a self-driving car and millions of miles collected from human driving, develop a smart car system.

Learning techniques

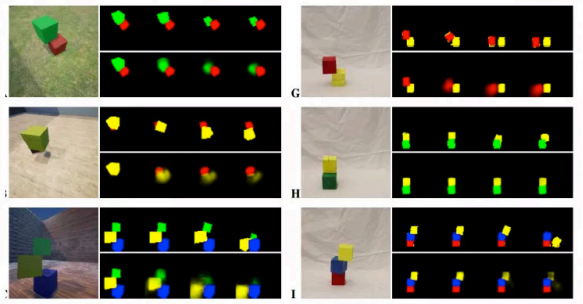
- Supervised learning (recent breakthrough 2012-)

- Unsupervised learning (still incomplete)



- Reinforcement learning (recent breakthrough 2015-)

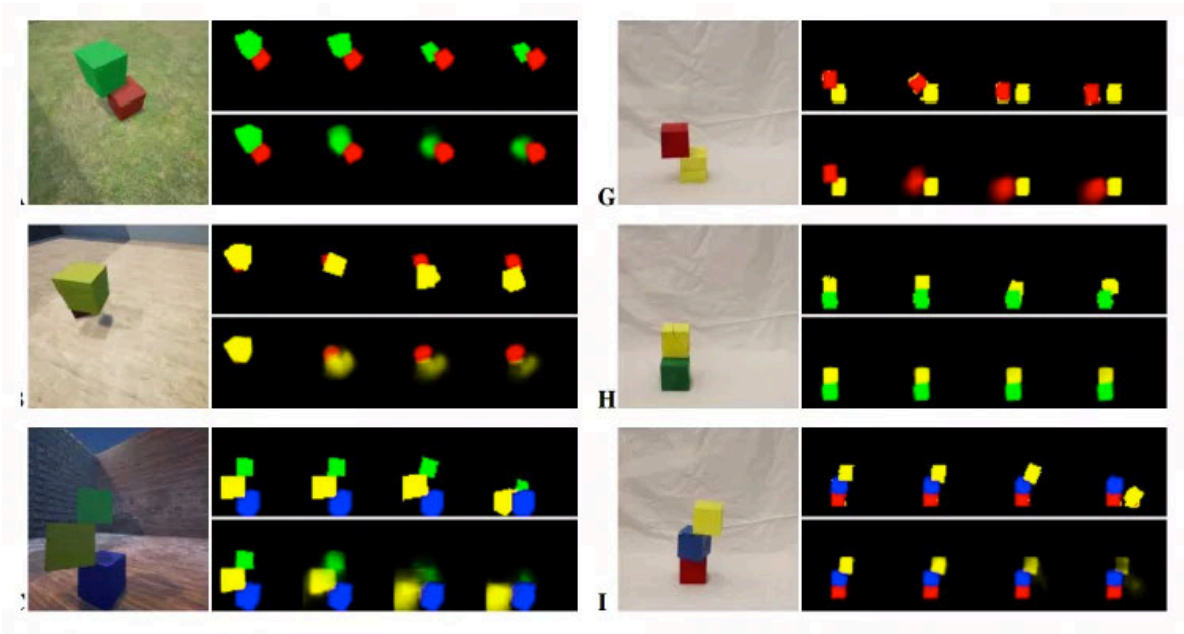
- **Predictive learning (main obstacle to general AI)**

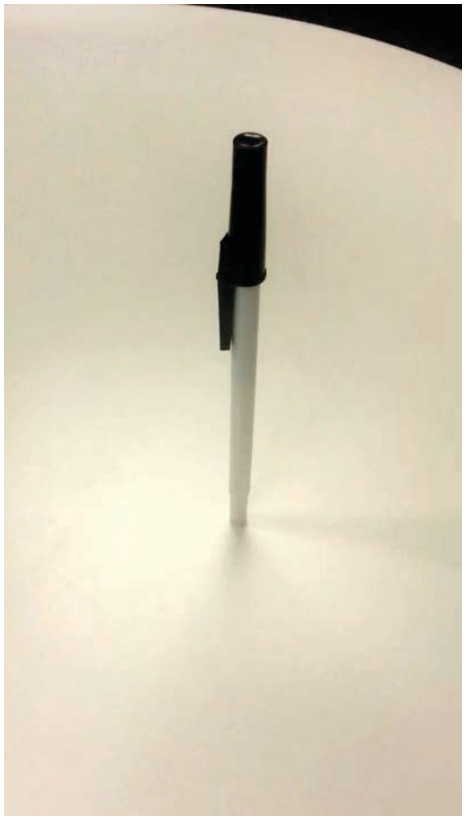


Predictive learning/common sense [LeCun]

- Definition: Unsupervised learning of the world and its laws by observing data, acting, and predicting the future, like Physics.

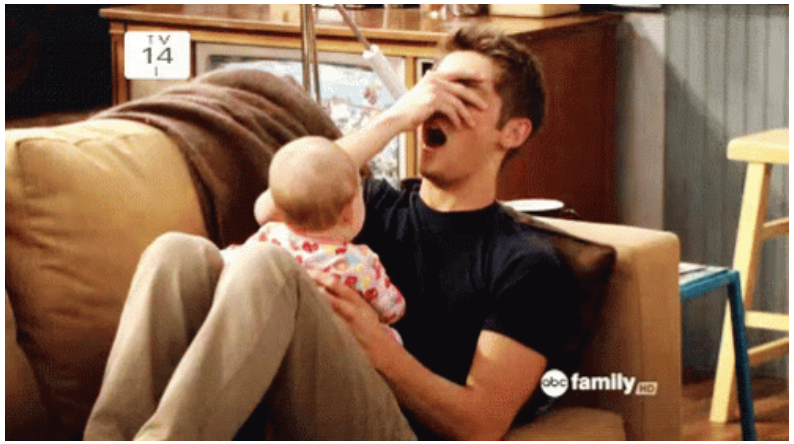
- Main obstacle to general AI today.



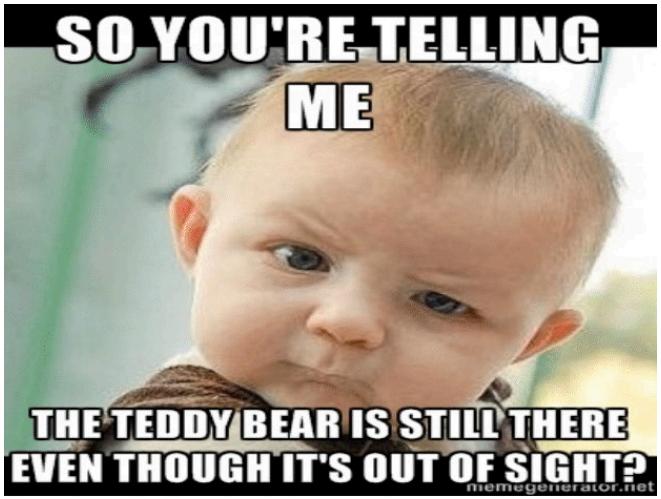


Common sense

- Human learning is unsupervised, and use observation-action on the world.
- Humans are able to learn in the first years physical perceptions like depth, 3-dimensionality, gravity, object permanence, etc.
- No need for advanced mathematical concepts like PDEs (Newton's laws).



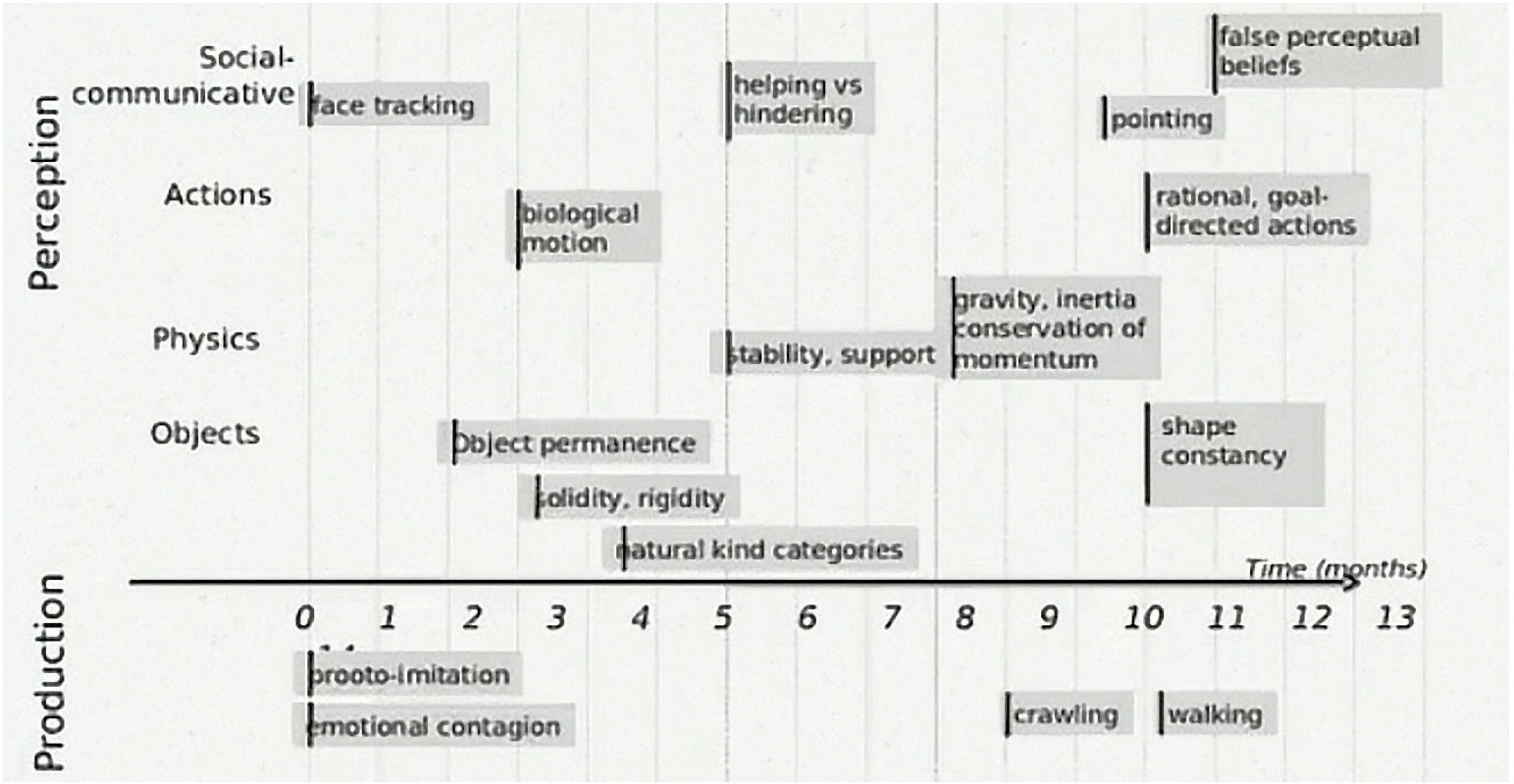
Peek-a-boo



Learn object permanence

Early concept acquisition

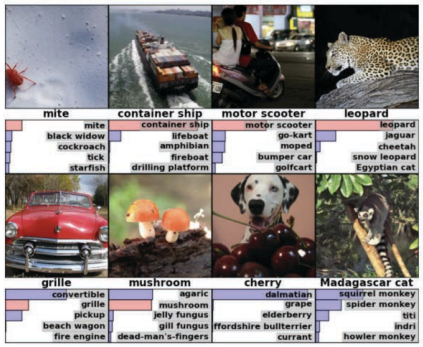
- Humans learn common sense through perception and production



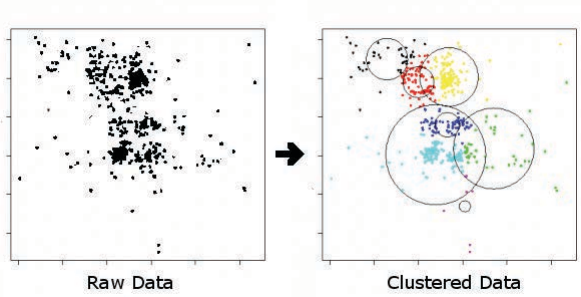
- How do we get the machines to learn common sense? Predicting past, present or future from available information?

Conclusion

- Supervised learning (recent breakthrough 2012-)

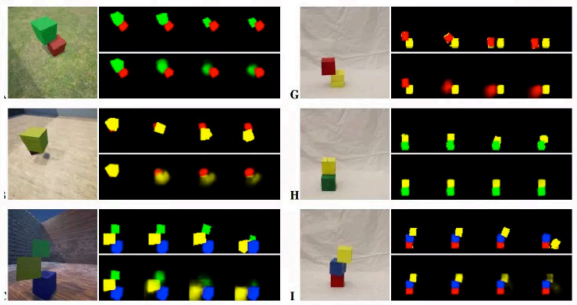


- Unsupervised learning (still incomplete)



- Reinforcement learning (recent breakthrough 2015-)

- Predictive learning (main obstacle to general AI)



Questions?