Introduction to Artificial Intelligence and Machine Learning

Patrick J. Martin, Ph.D.

Cognitive Science and Artificial Intelligence Department

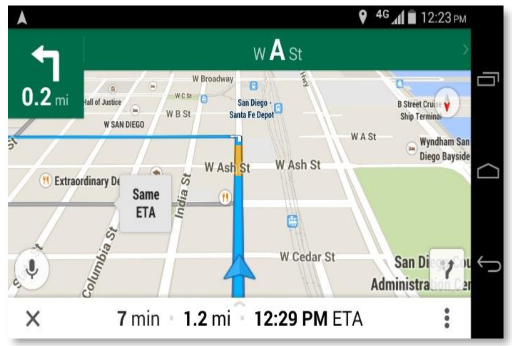
MITRE Corporation

GENERATION℠ AI NEXUS

AI in the Real World

AI technology is incorporated into current consumer products…

Personal Assistants



Mapping Software

Web Platforms/Content Delivery

…and is foundational to developing industries

Autonomous Vehicles

Internet of Things

Your Turn!

Write down your thoughts on AI! What do you know about it?

Artificial Intelligence

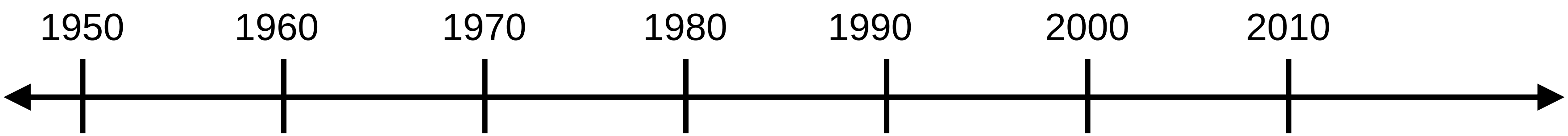
Artificial Intelligence

Algorithms and software that enable machines to:

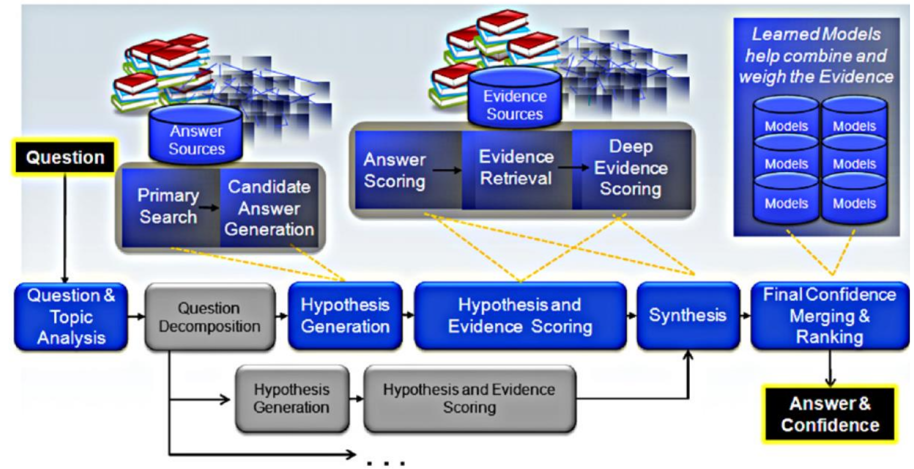
Brief History
Foundations

The field of AI builds upon many disciplines

Brief History
Timeline



Watson Jeopardy

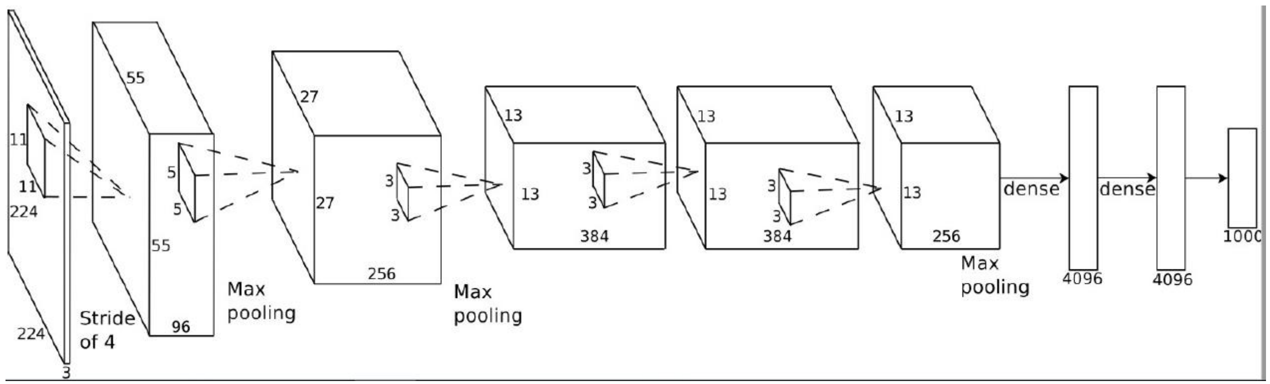


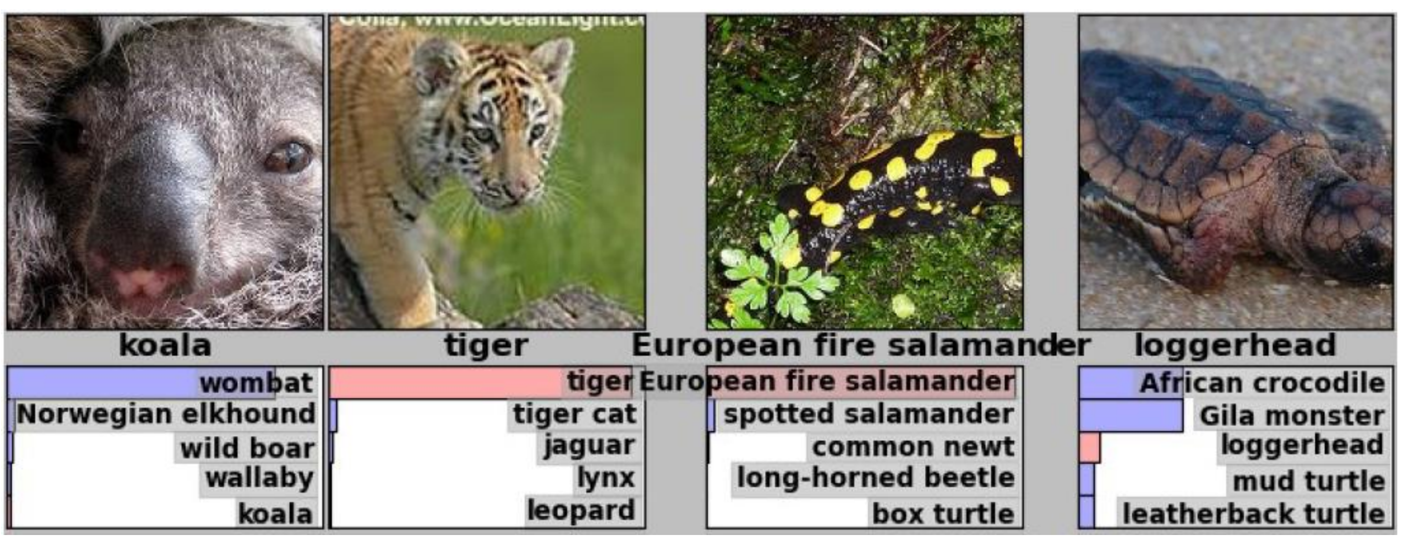
- **AI technologies used**
  - Graph search
  - Probabilistic reasoning
  - Machine learning
  - Natural language processing
- **Lessons learned**
  - Building real AI systems requires considerable engineering
  - Pushed the boundaries of Question Answering technology

D. Ferucci, et. al., Building Watson: An Overview of the DeepQA Project, *AI Magazine*, 2010

The Rise of Deep Learning

AlexNet (2012)

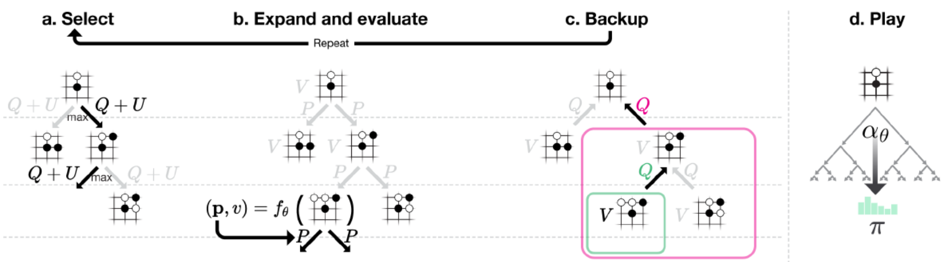




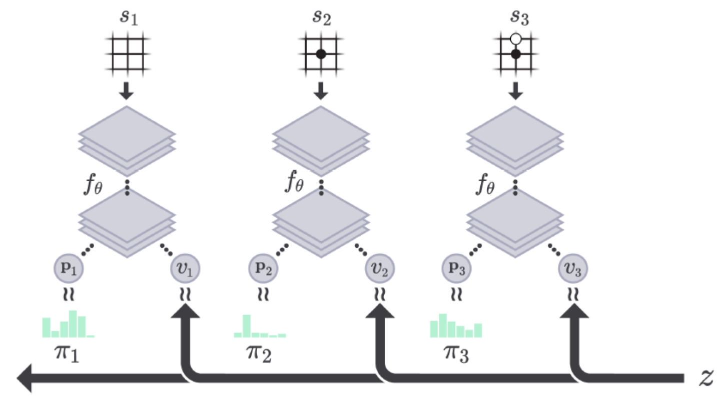
A. Krizhevsky et. al., ImageNet Classification with Deep Convolutional Neural Networks, NIPS, 2012

Alpha Go Zero

Tree search algorithm produces moves



Neural network trained with chosen moves

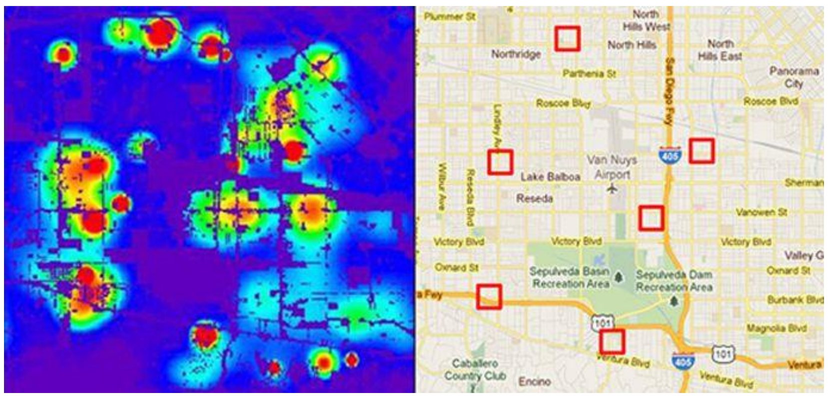


- **AI technologies used**
  - Deep reinforcement learning
  - Tree search
- **Lessons learned**
  - RL can be scaled to complex, but known, interactions with humans
  - Not very useful in unknown, dynamic environments

https://deepmind.com/blog/alphago-zero-learning-scratch/

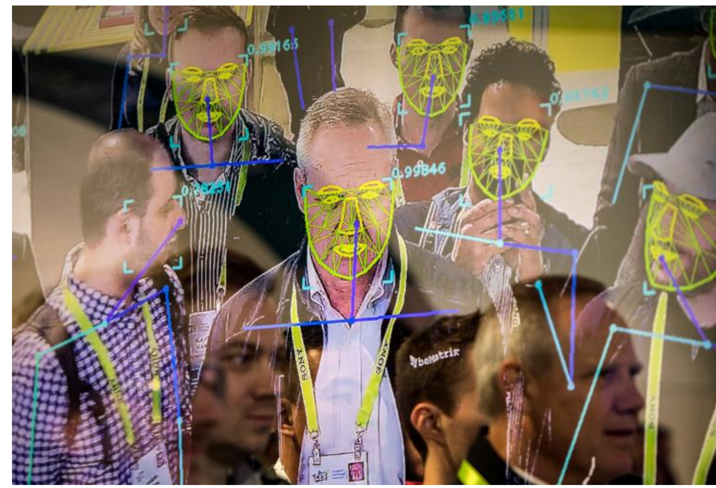
Challenges

- **Bias: At the mercy of our data**



Predictive Policing (Smithsonian)

- **Privacy: Our data is at the mercy of organizations**



Public Facial Precognition (NY Times)

GAIN MITRE

Computational Rationality

- **What do we mean by rational?**

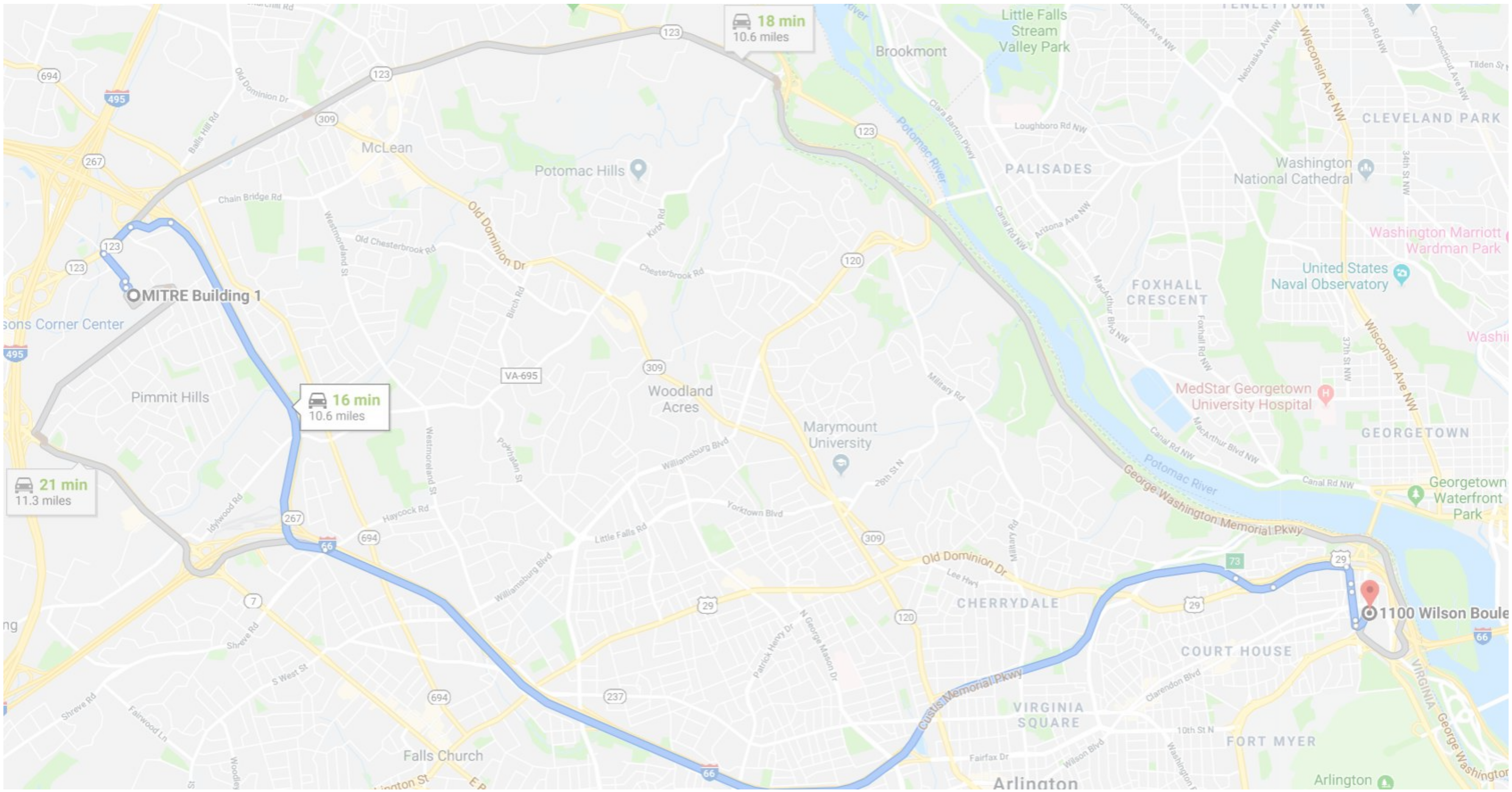
- **What is a goal?**

Knowledge, Reasoning, and Planning

- **Structures that maintain facts**

- **Algorithms that compute actions that achieve goals**

Knowledge, Reasoning, and Planning
(Un)Informed Search



Knowledge, Reasoning, and Planning
(Un)Informed Search

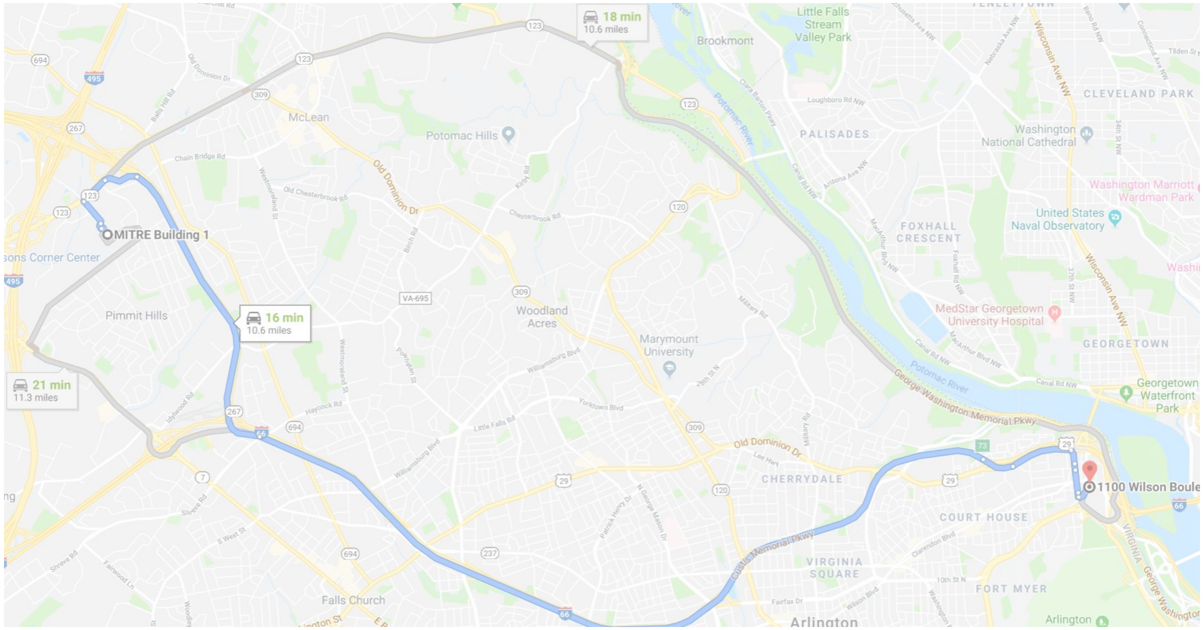
Search problems have the following pieces:

Search problems are solved with an algorithm

GAIN MITRE

Probabilistic Reasoning

Consider the map search problem…



Probabilistic Reasoning
Bayesian Methods

Making decisions when we are uncertain about…

Probabilistic Reasoning
Example: Bayesian Network

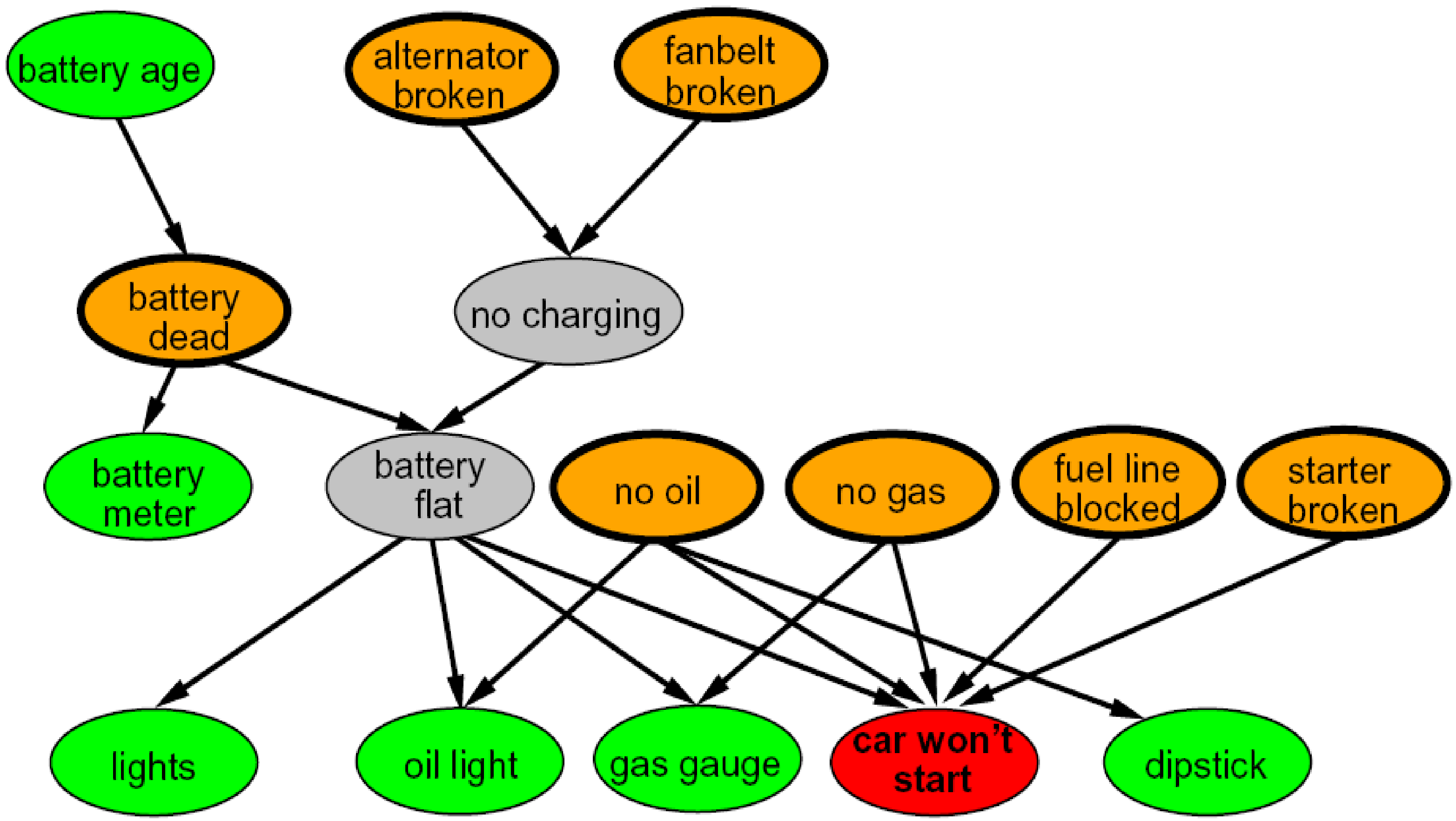
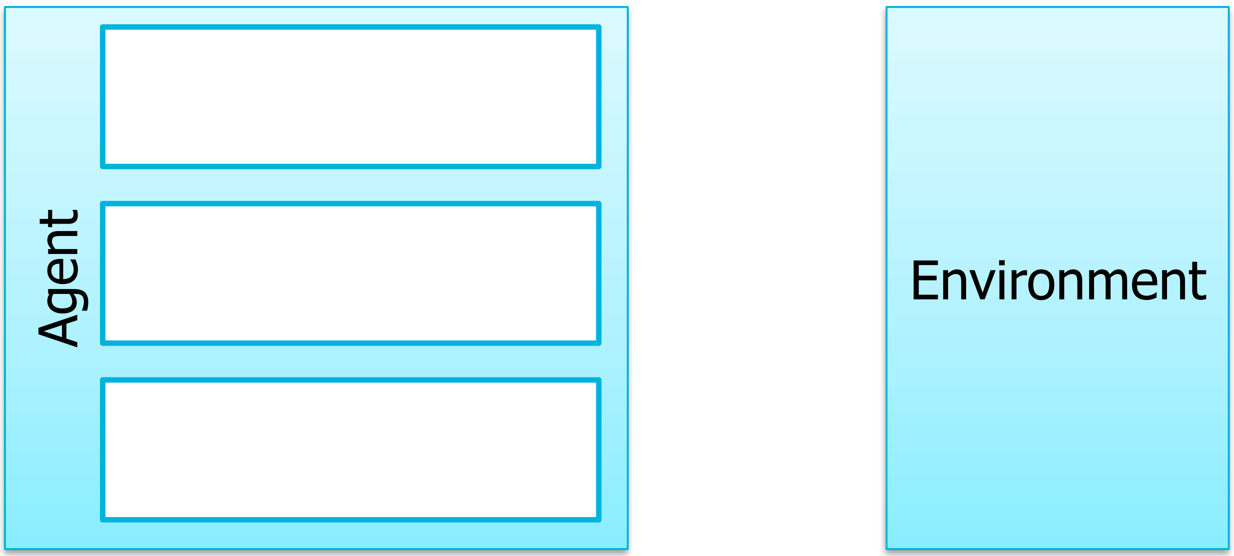


Image from http://ai.berkeley.edu

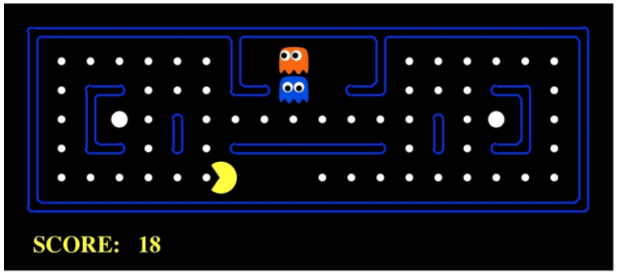
Communication, Perceiving, and Action

AI algorithms enable intelligent, rational agents



Communication, Perceiving, and Action
Applications

- **Pacman**



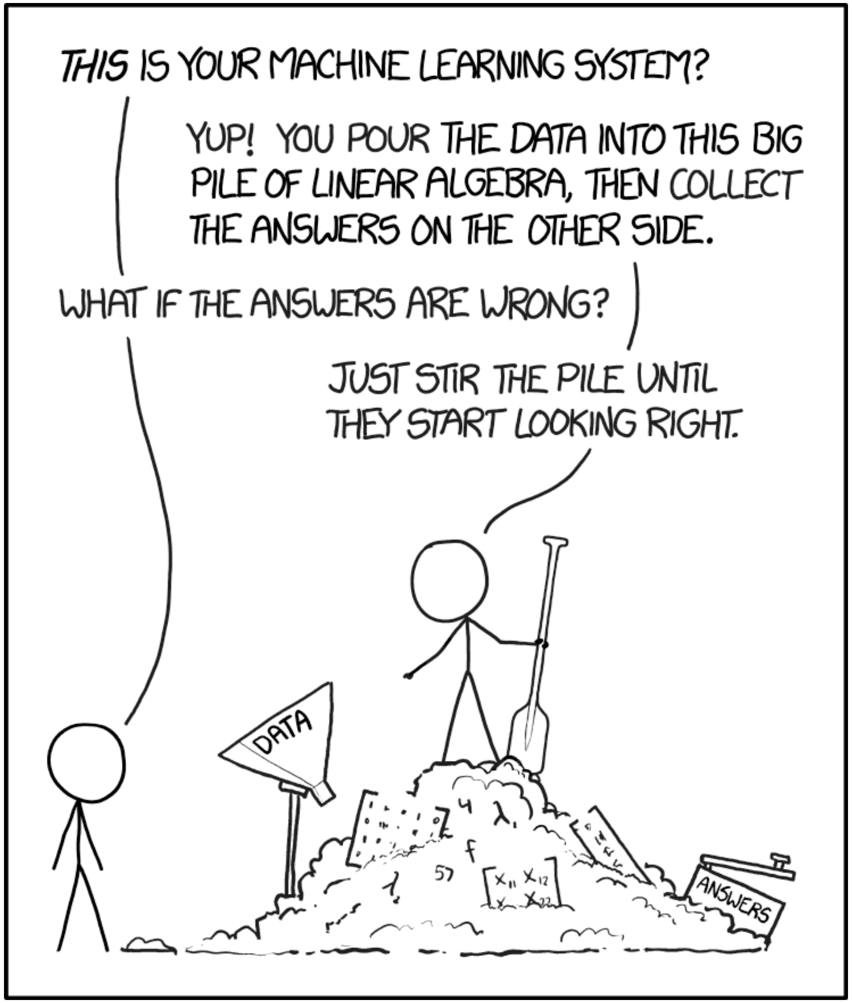
Pac-Man is a registered trademark of Namco-Bandai Games, used here for educational purposes

- **Autonomous Car**

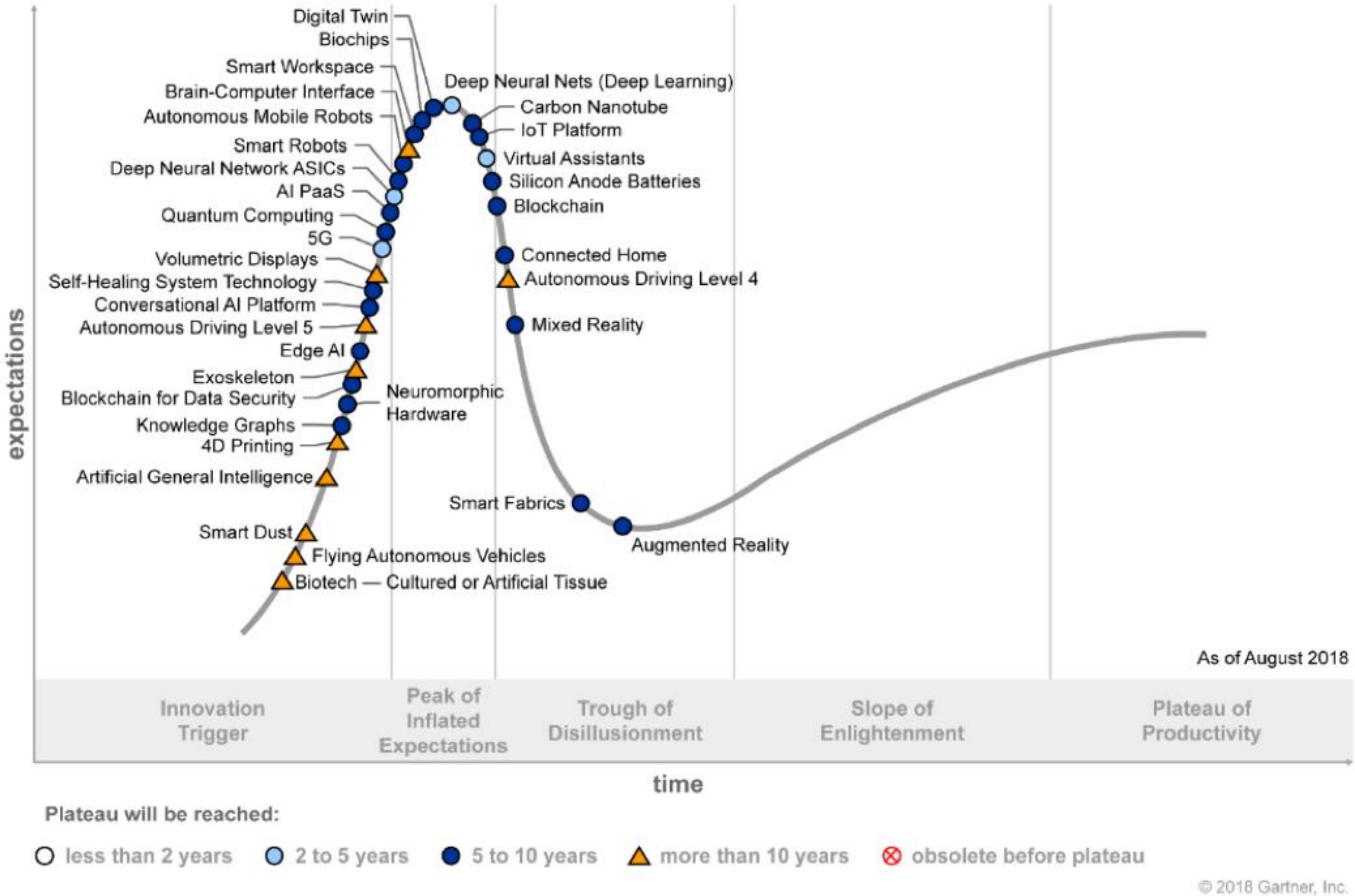
Image courtesy of Toyota and used for educational purposes

GAIN MITRE

Machine Learning

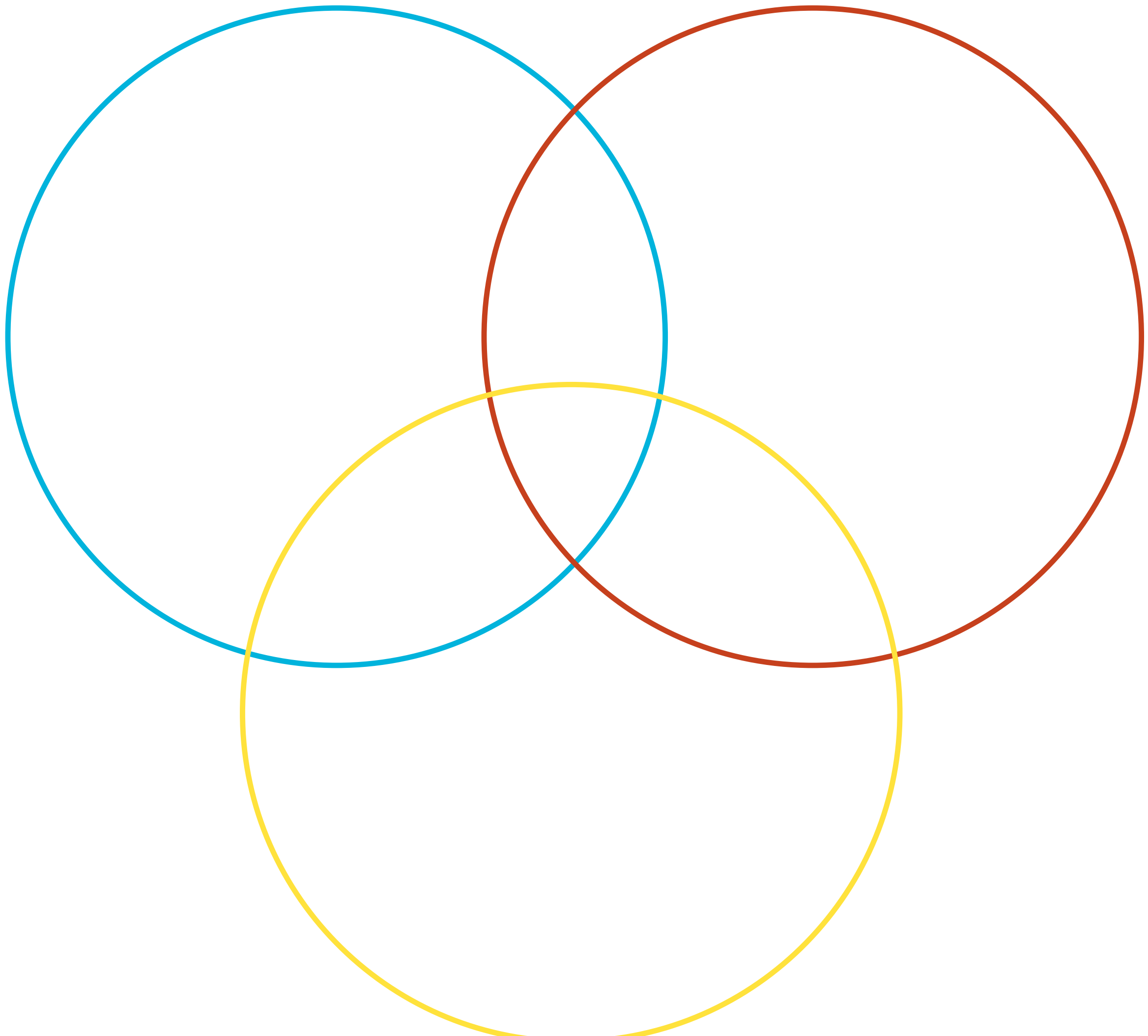


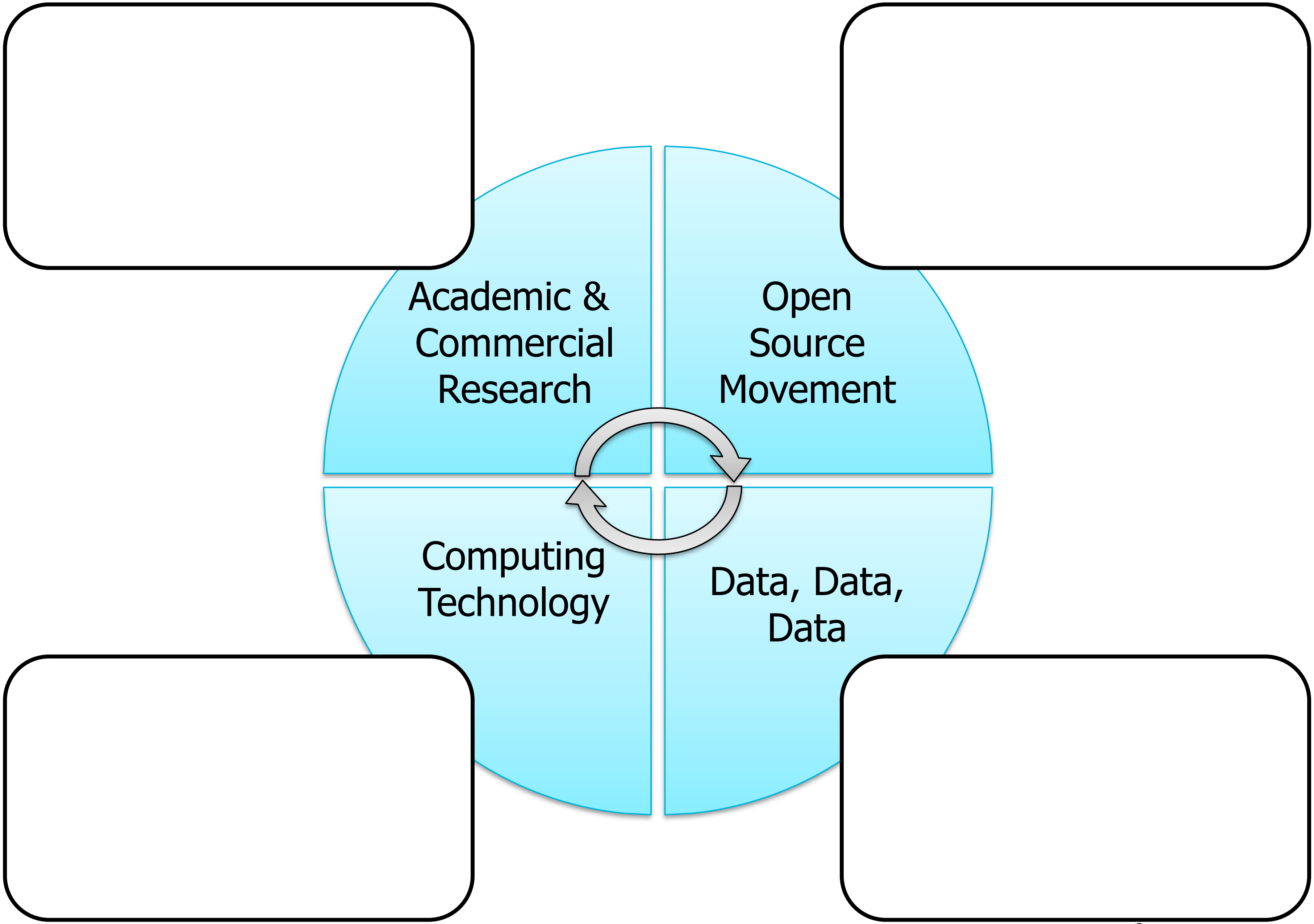
https://xkcd.com/1838/



[From Gartner Press Release, August 2018]

Theoretical Underpinnings




Academic & Commercial Research
Open Source Movement
Computing Technology
Data, Data, Data

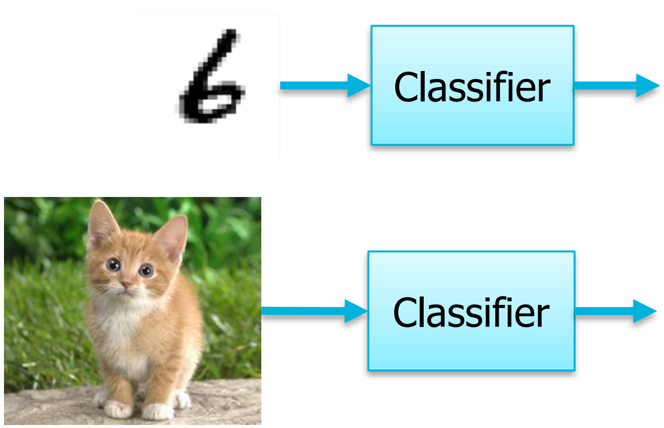
Supervised Learning

Definition

The process of learning a model from a set of **input vectors** and their **target vectors**

Classification

Regression



Unsupervised Learning

Definition

The process of learning a model from input data vectors with **no** target values/classes

imgflip.com

Applications

Frequently Used Algorithms

Reinforcement Learning

Definition

The process of learning an optimal output via trial and error in an environment; goal directed learning*

Applications

Frequently Used Algorithms

[*] R.S. Sutton and A.G. Barto. *Reinforcement Learning: An Introduction*, MIT Press, 1998

Exercise

Which type of machine learning is applicable to the following scenarios? Be sure to include why.

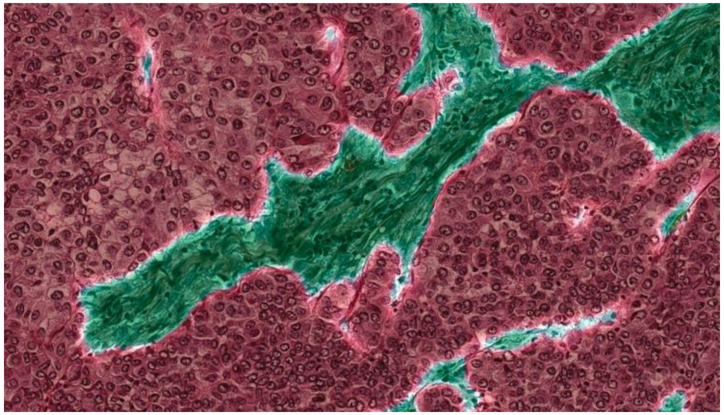
1. Amazon segmenting customers in a database.

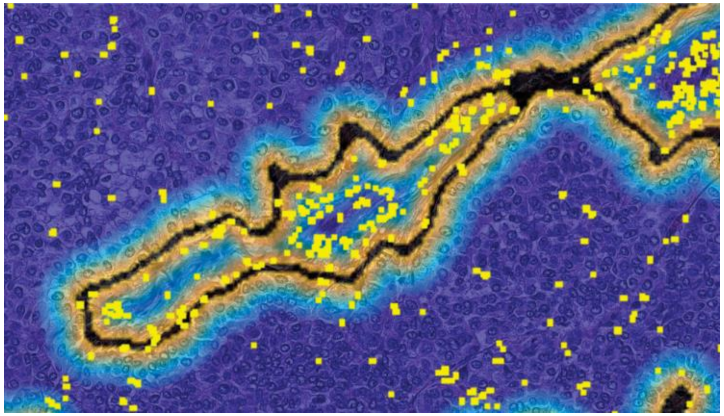
2. Enabling a smart taxi to correctly make a right-on-red

3. Detecting a person of interest in a surveillance video

Example Application
Sepsis Watch

- **Data Requirements**
- **Benefits**
- **Challenges**





Images from https://spectrum.ieee.org/biomedical/diagnostics/the-first-frontier-for-medical-ai-is-the-pathology-lab

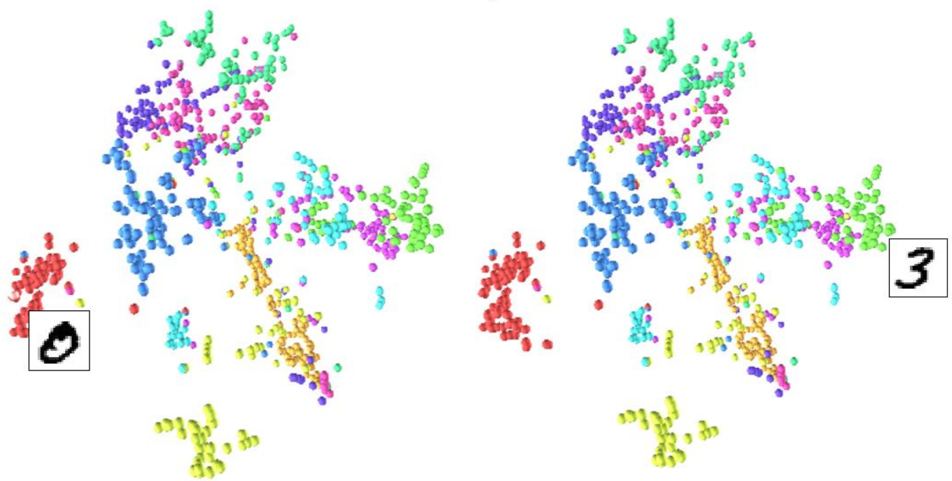
Development Process
Data Partitioning and Analysis

Chunk O' Data

- **Data Partitioning**

$28px \times 28px$

- **Analysis**



http://colah.github.io

GAIN MITRE

Development Process

ML Algorithm Assessment

- **Training algorithms fit a model to the training data**
- **Things to think about…**

Regression
Decision Trees

Random Forests

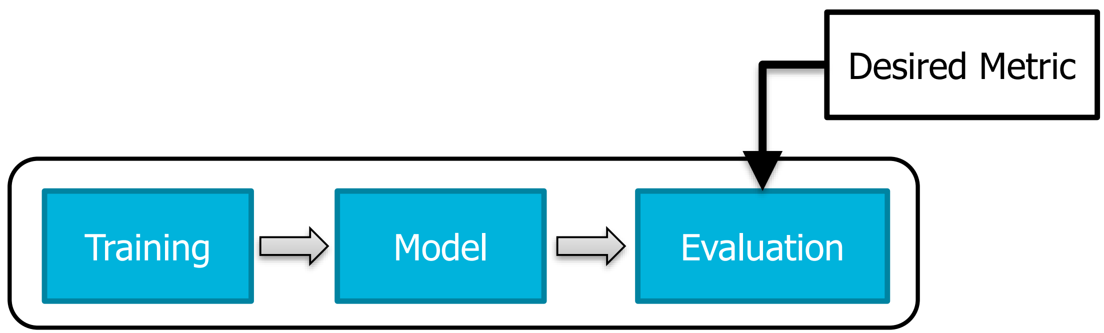
Deep
Neural Networks

Development Process

Model Training and Assessment

Goal: Learn a model that generalizes to data it has not seen before

Training	Test



*sklearn.model_selection has multiple CV tools to evaluate your models

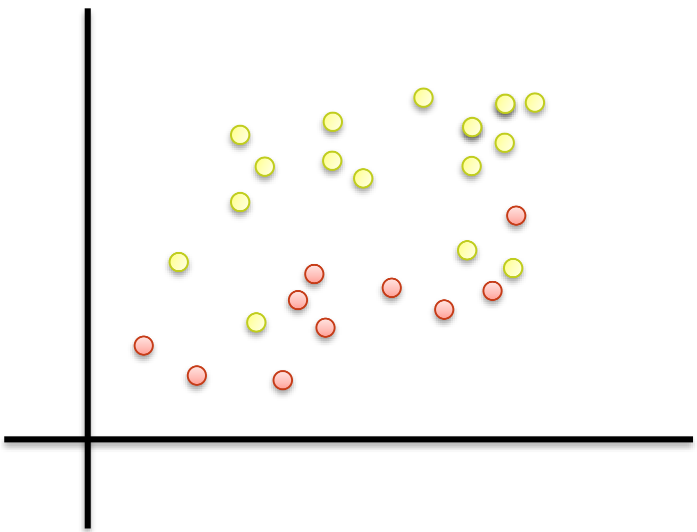
Performance Metrics

Metrics allow you to compare candidate models

Regression Metrics

Classification Metrics

Classification Metrics (1)



Confusion Matrix

Actual

Predicted

Classification Metrics (2)

Precision: How many classifications are true positives

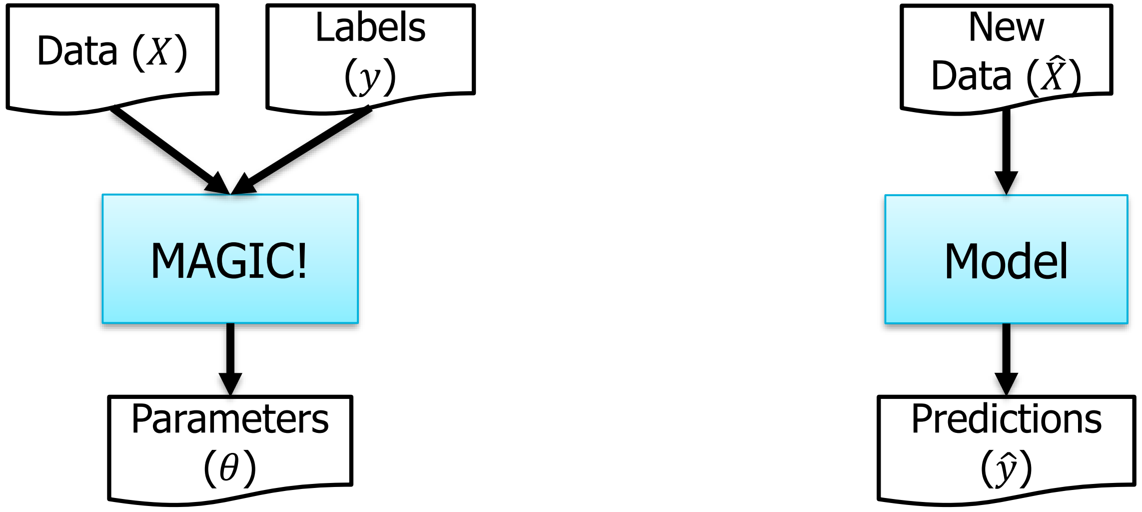
Recall: How many true positives were found

Accuracy: Number of correctly classified items over all items

Recap: Supervised Learning

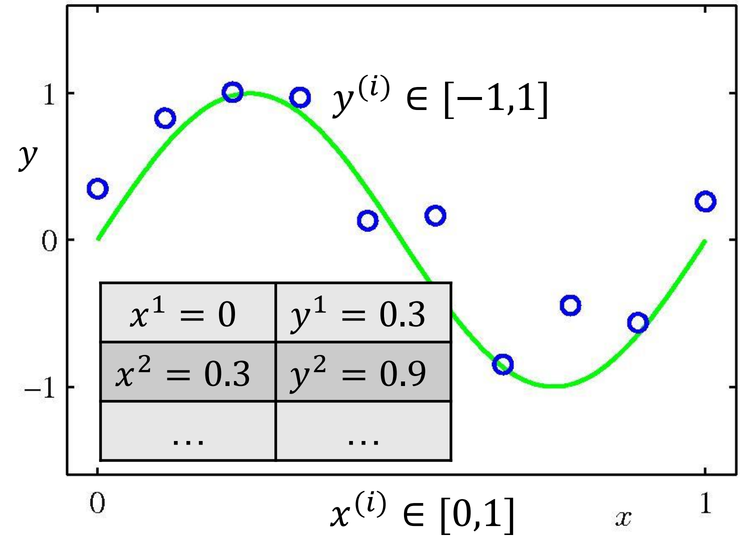
Definition

- **The process of learning a model from a set of input vectors and their target vectors**



Sounds Like Curve Fitting!

- **Start with a collection of data**
  - Each $x^{(i)} \in \mathbb{R}$ is an instance
- **Each input is associated with a target value/class**
  - $y^{(i)} \in \mathbb{R}$ (sometimes denoted $t^{(i)}$)
- **Learning Goal**
  - Predict the value of $\hat{y}$ for any new $\hat{x}$.



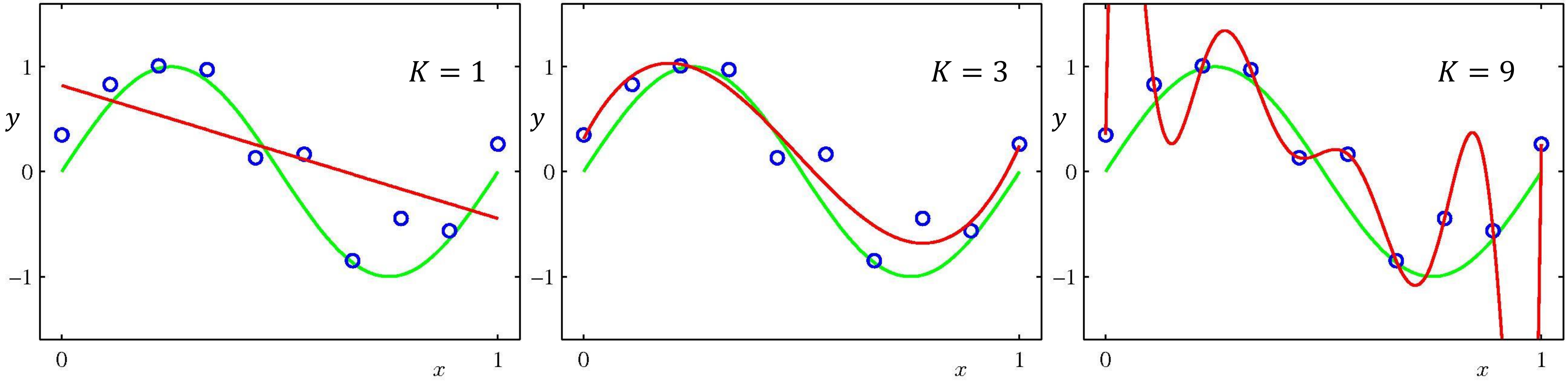
GAIN MITRE

Sounds Like Curve Fitting!

Example: polynomial model

$$h(x, \boldsymbol{\theta}) = \theta_0 + \theta_1 x + \theta_2 x^2 + \cdots = \sum_{k=0}^{K} \theta_k x^k$$

Some example models based on polynomial order, K



Linear Regression

Problem Statement

- **Input:** m **instances of** $x \in \mathbb{R}^n$**, targets,** $y \in \mathbb{R}$
- **Goal: Compute parameters** $\boldsymbol{\theta}$ **of a linear model** $h(x, \boldsymbol{\theta})$ **that predicts** $\hat{y}$

Linear Model

$$\hat{y} = \theta_0 + \theta_1 x_1 + \cdots + \theta_n x_n$$

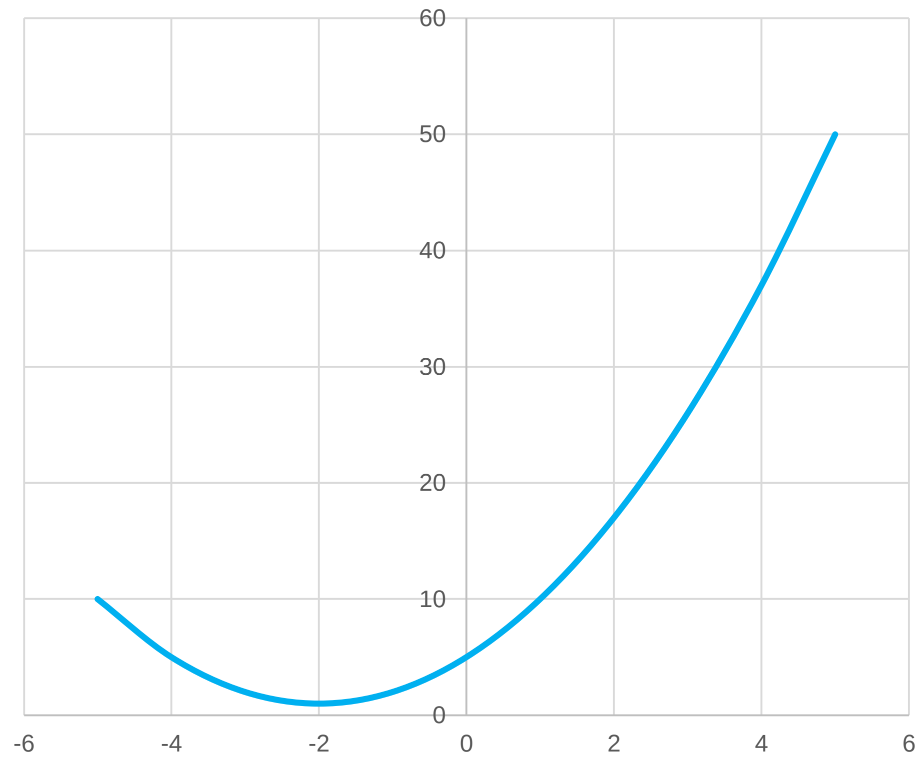
Loss Function

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(\left(\theta_1 x_1^i + \cdots + \theta_n x_n^i \right) - y^i \right)^2$$

Fitting the Model
Gradient Descent in a Nutshell

A critical algorithm for ML!

Basic Steps:



Regularization

- **Overfitting results in crazy model parameters, $\boldsymbol{\theta}$.**
  - Solution:
- **l_2 Regularization**

$$\frac{1}{2m}\sum_{i=1}^{m}\left(\left(\theta_1 x_1^i + \cdots + \theta_n x_n^i\right) - y^i\right)^2 + \boxed{\gamma\frac{1}{2}\sum_{i=1}^{n}\theta^2}$$

Benefit

- **l_1 Regularization**

$$\frac{1}{2m}\sum_{i=1}^{m}\left(\left(\theta_1 x_1^i + \cdots + \theta_n x_n^i\right) - y^i\right)^2 + \boxed{\gamma\sum_{i=1}^{n}|\theta|}$$

Benefit

Linear Regression Properties

Pros

- Versatile technique with lots of library support
- Scales well thanks to gradient descent

Cons

- In general, not globally optimal
- Nonlinear models require tricks…

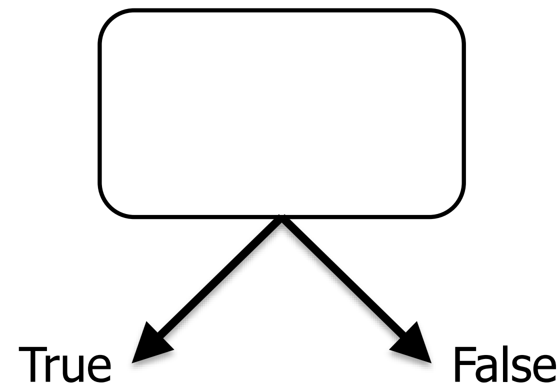
$$\hat{y} = \theta_0 + \theta_1 x_1 + \cdots + \theta_n x_n$$

Decision Trees

Where should we go out to lunch?

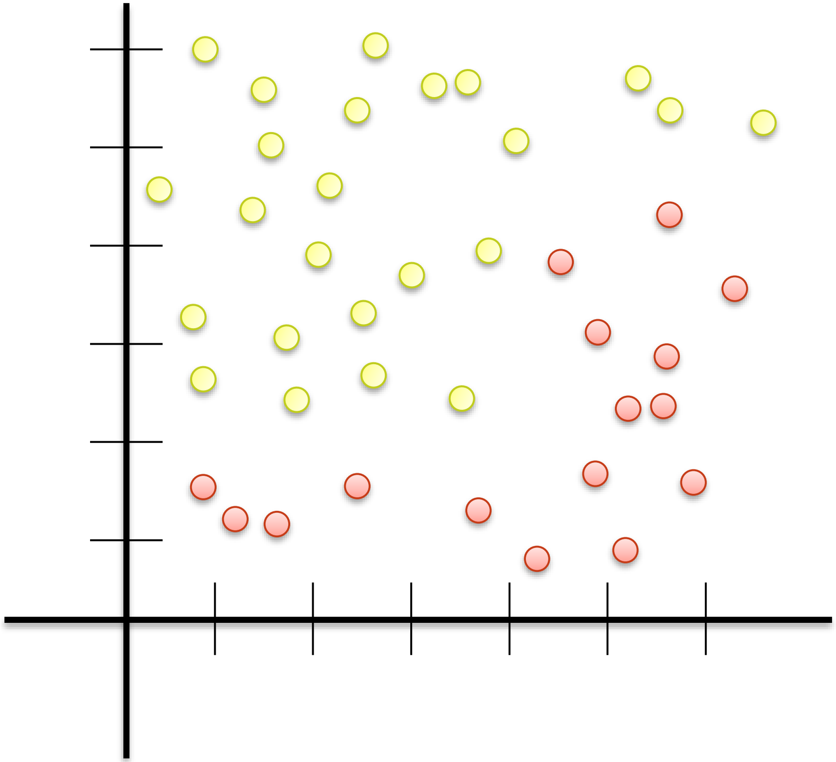
Indian   Pizza   Hamburgers

What features do we care about?



Decision Tree Example

What is the model for this data?



Decision Tree Properties

Pros

- **Minimal data preparation required**
- **Fast inference**
  - $O(\log_2(m))$
- **Models are easier to understand**

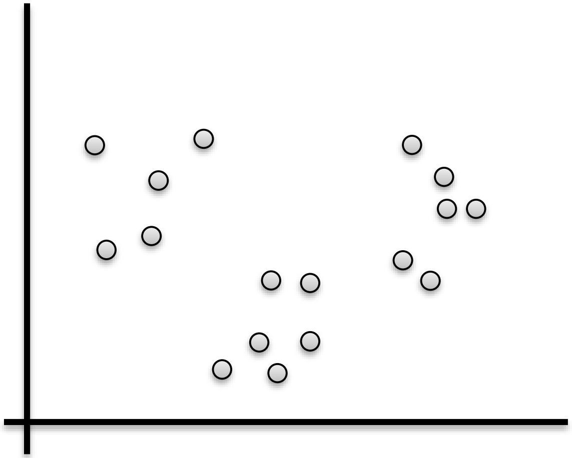
Cons

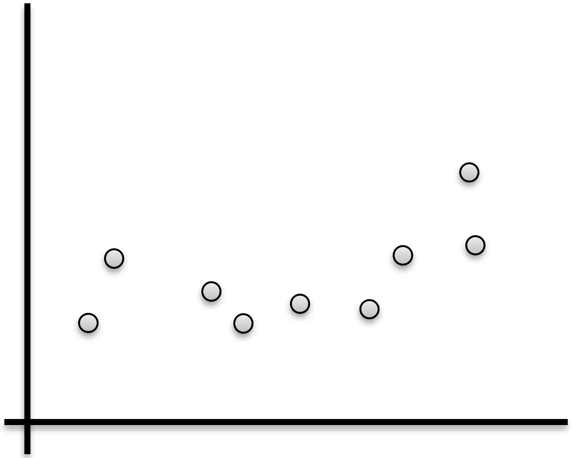
- **Sensitive to small variations in training data**
- **Models not globally optimal**
- **Training does not scale well with features**
  - $O(n \cdot m\ log(m))$

Recap: Unsupervised Learning

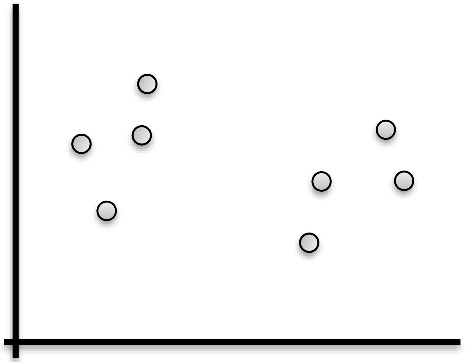
Definition

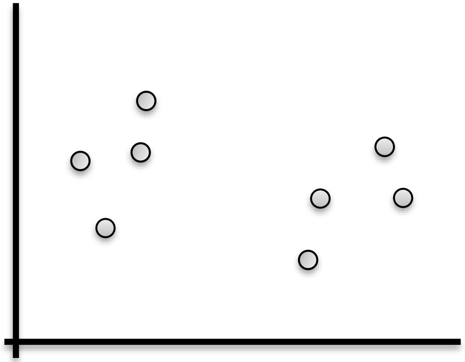
- **The process of learning a model from input data vectors with no target values/classes**





Clustering via K-Means

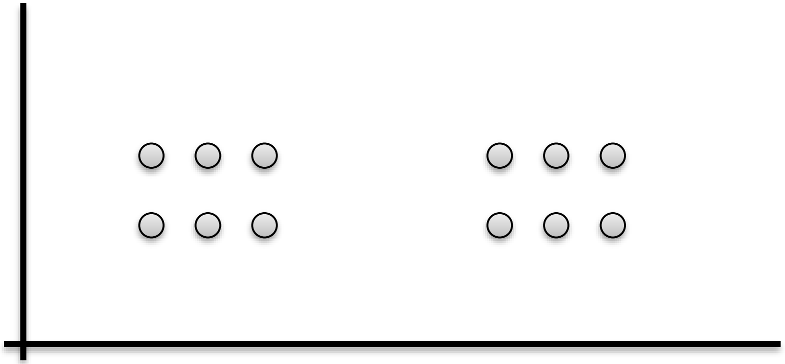




- **Minimize quadratic error between data and their cluster centers**
- **Gradient descent can result in local minima**

K-Means Summary

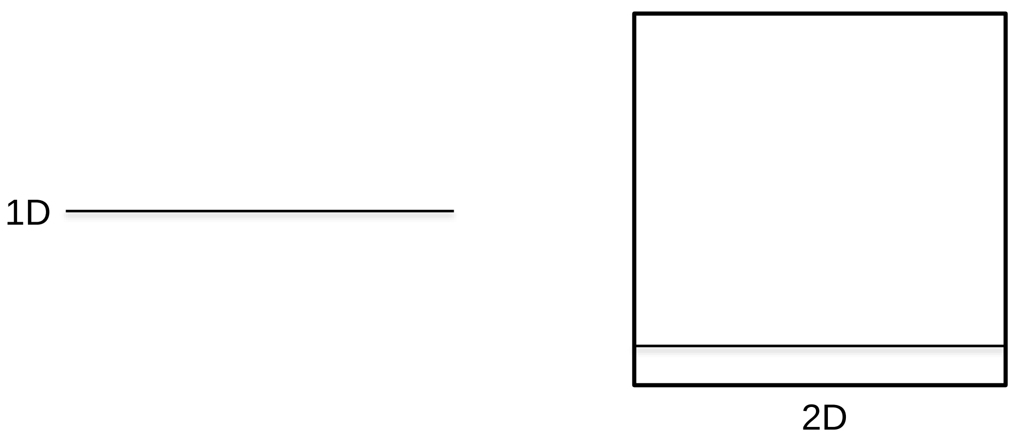
- Efficient on large data sets
- Benefit:
- Things to keep in mind:



Motivation for Dimensionality Reduction

- **Curse of Dimensionality!**
  - As the number of dimensions grow, the amount of data we need grows exponentially
- **Let's sample some spaces…**

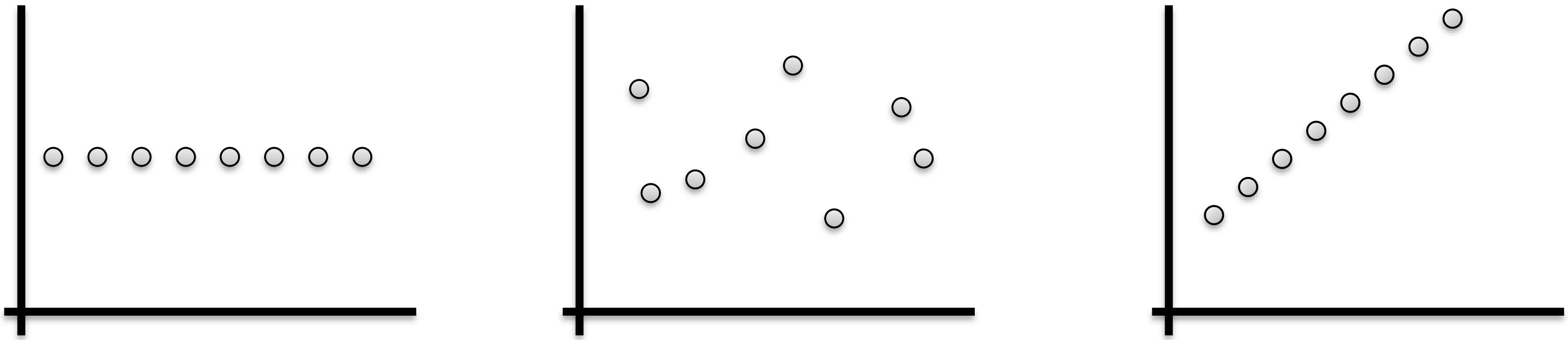
http://muppet.wikia.com/wiki/Count_von_Count



GAIN MITRE

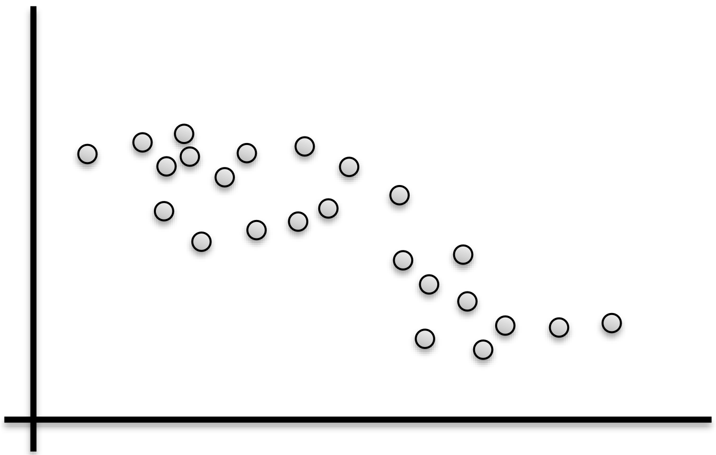
Principal Component Analysis (PCA)

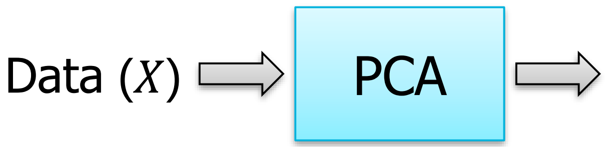
- **PCA is a projection technique**
  - Hypothesis:
- **In what dimension does the data reside?**



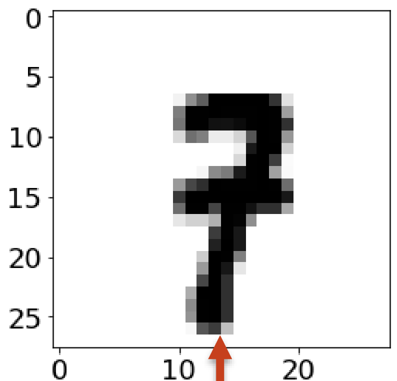
PCA Technical Approach

- **PCA transforms data of any shape into a new coordinate system**
- **Project with "minimal loss of information"**

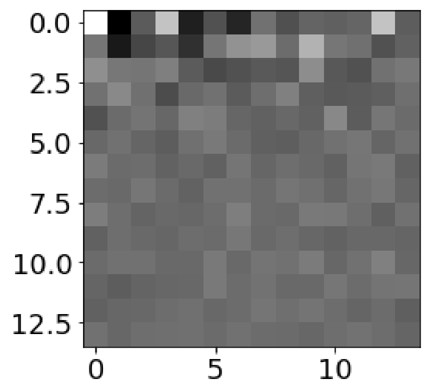




PCA Example

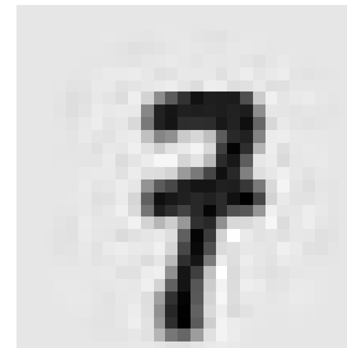


PCA Transform!



¯_(ツ)_/¯

Inverse Transform!



Each feature is a number between 0 and 255

0 255

PCA Summary

- **Great for managing large, sparse data sets**

- **Benefit:**

- **Things to keep in mind:**

Back to some algorithmic supervision!

What do we know now?

Supervised Methods

Linear Regression

Decision Trees

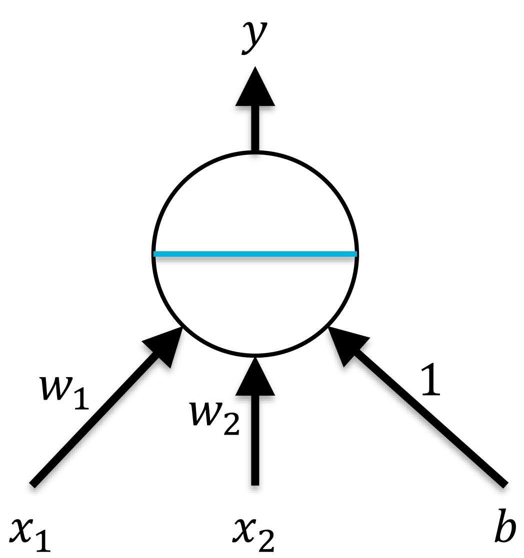
Unsupervised Methods

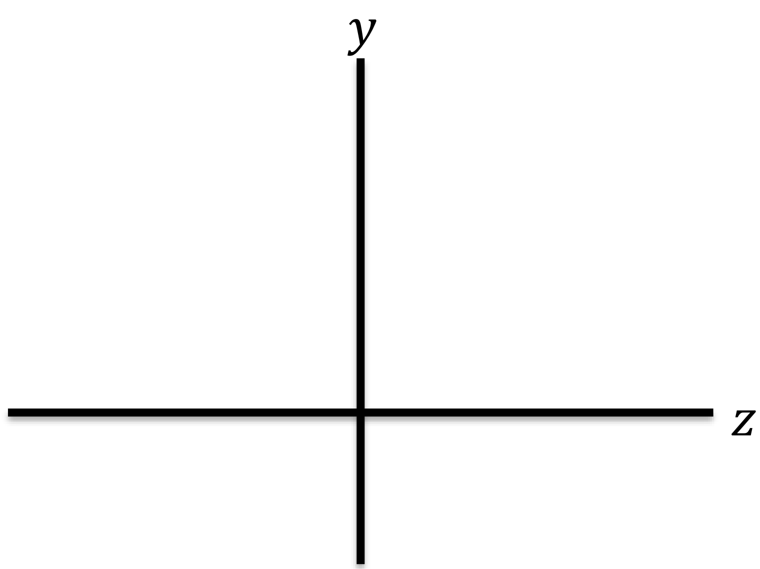
Principal Component Analysis

Neural Networks
To the (sort of) rescue!

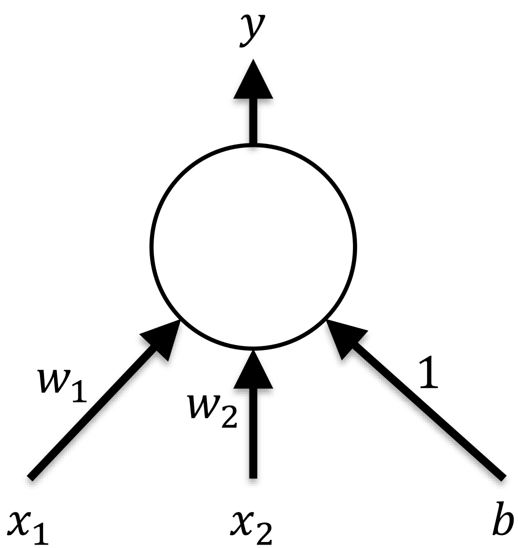
Neural networks go back to the 1950s!

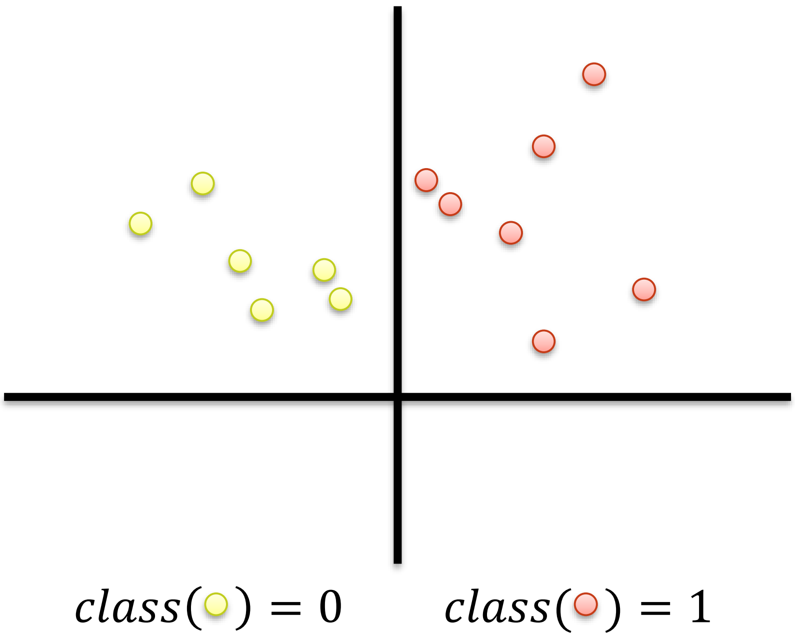
Linear threshold unit (LTU)



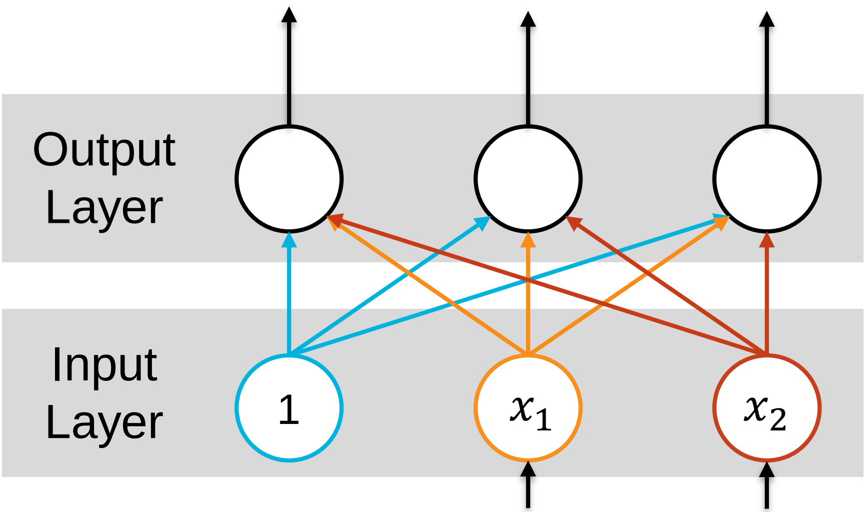


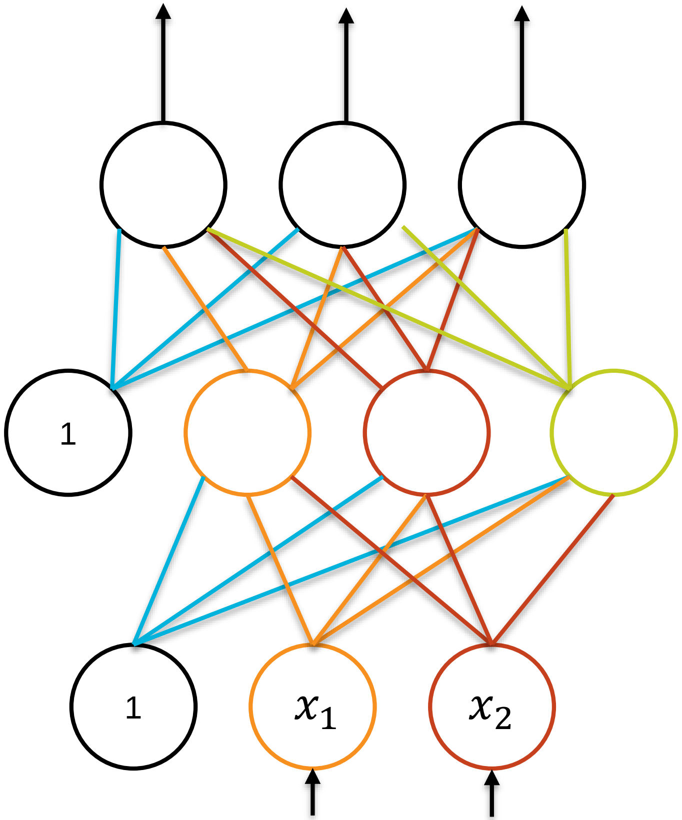
A Closer Look





Neural Network Architectures





Why Neural Networks?

- NNs with one hidden layer universal function approximators

- NNs have efficient architectures and training algorithms

Summary

- **Machine learning techniques provide powerful statistical tools**
  – Classification
  – Regression
  – Clustering
  – Dimensionality reduction

- **Remember:**

MITRE

MITRE's mission-driven teams are dedicated to solving problems for a safer world. Through our federally funded R&D centers and public-private partnerships, we work across government to tackle challenges to the safety, stability, and well-being of our nation.

Learn more www.mitre.org