COMP/EECE 7/8745 Machine Learning

Topics:

Learning approaches

- Different Machine Learning(ML) approaches
- How and what does machine learn?
- Ecosystem for Machine Learning (DL)

Md Zahangir Alom
Department of Computer Science
University of Memphis, TN

Types of Learning

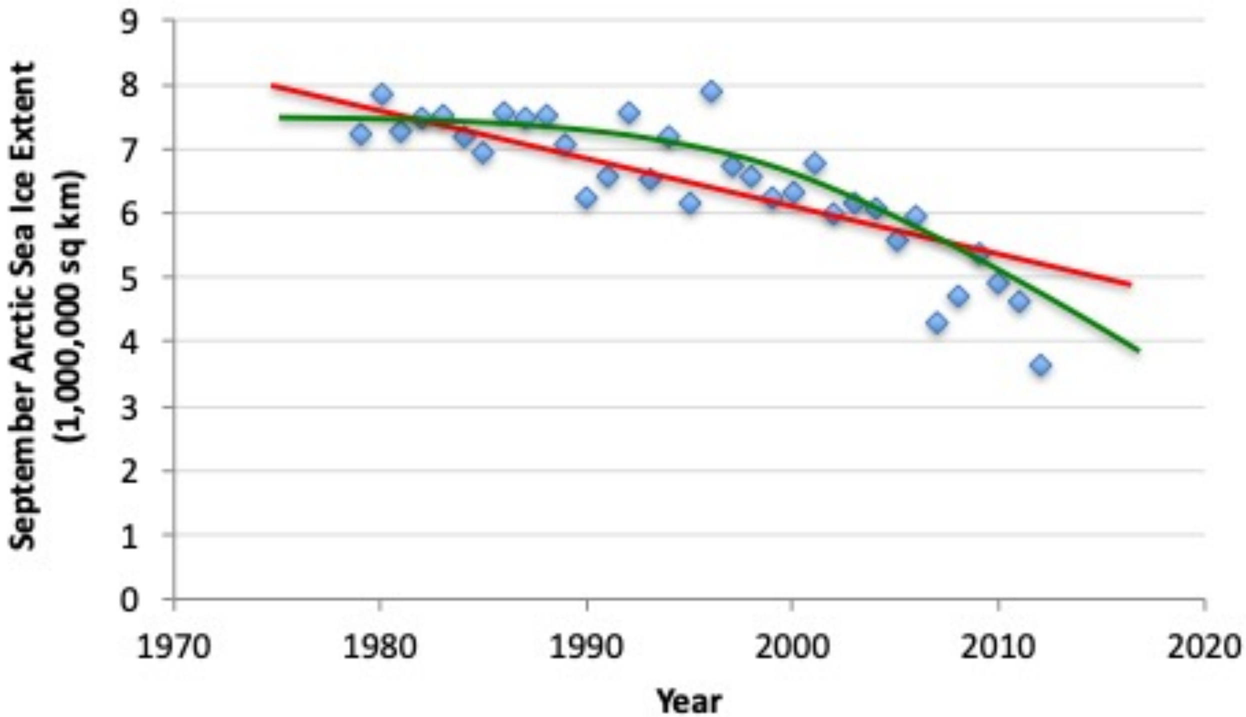
- **Supervised (inductive) learning**
  – Given: training data + desired outputs (labels)
- **Unsupervised learning**
  – Given: training data (without desired outputs)
- **Semi-supervised learning**
  – Given: training data + a few desired outputs
- **Reinforcement learning**
  – Rewards from sequence of actions

Based on slide by Pedro Domingos

24

Supervised Learning: Regression

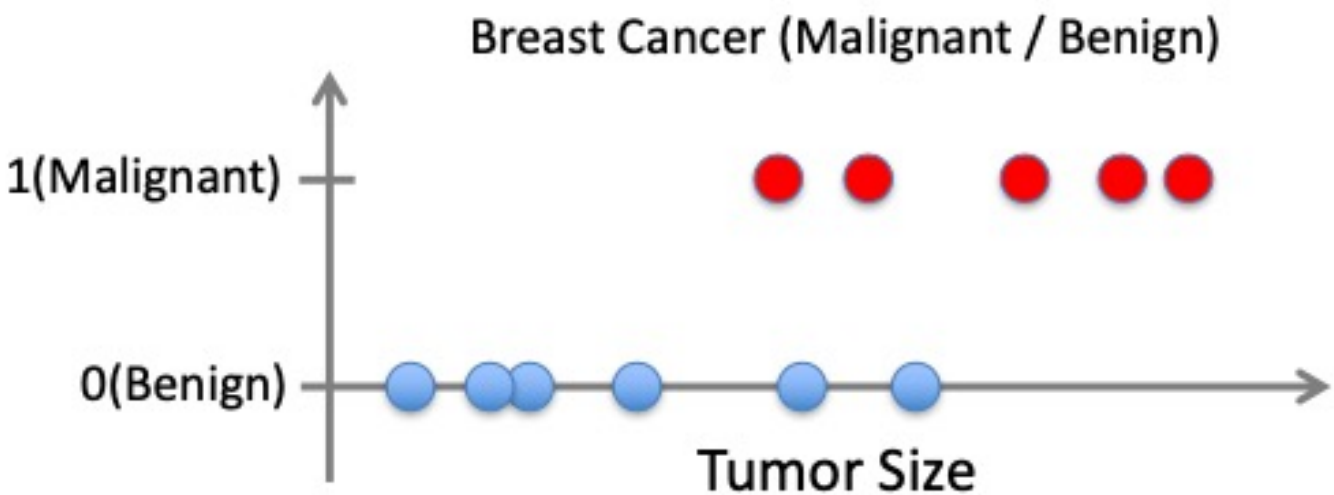
- Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
  - y is real-valued == regression



Data from G. Witt. Journal of Statistics Education, Volume 21, Number 1 (2013)

Supervised Learning: Classification

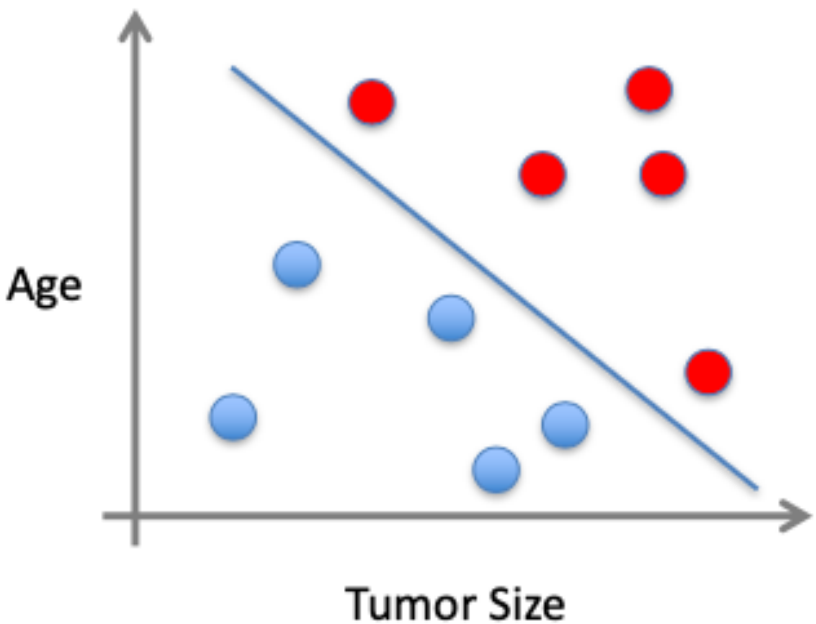
- Given $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$
- Learn a function $f(x)$ to predict y given x
  - y is categorical == classification



Based on example by Andrew Ng

Supervised Learning

- x can be multi-dimensional
  - Each dimension corresponds to an attribute



- Clump Thickness
- Uniformity of Cell Size
- Uniformity of Cell Shape

...

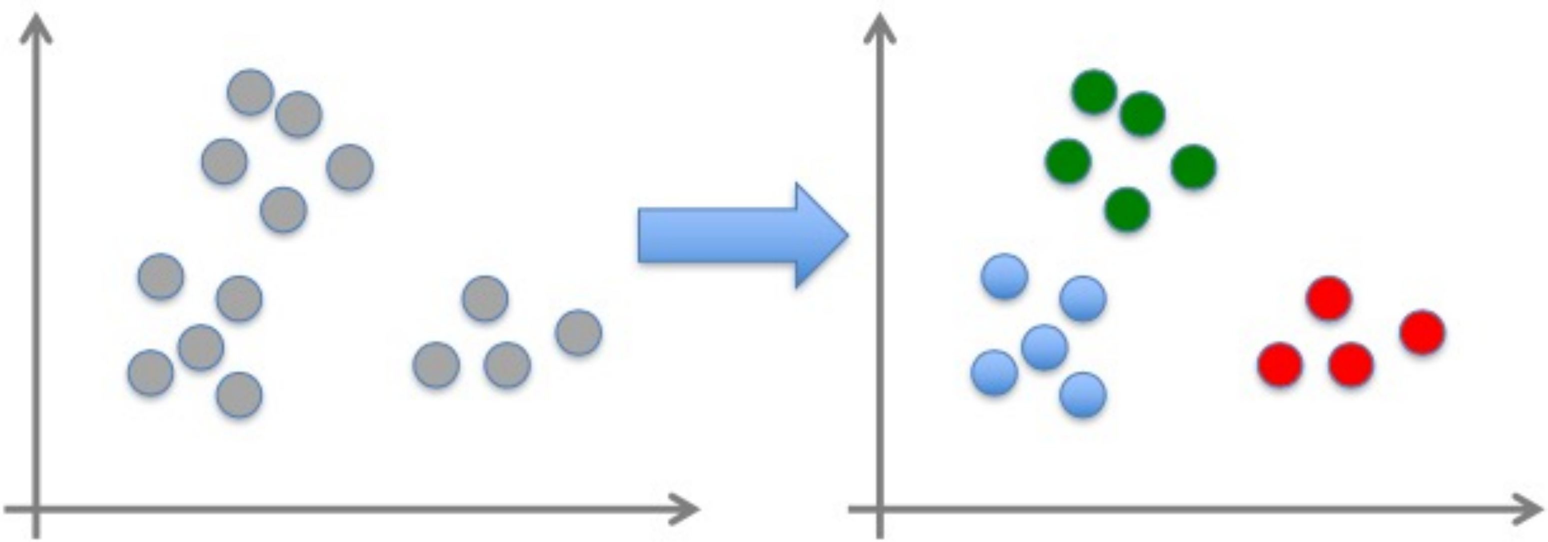
Based on example by Andrew Ng

Supervised Learning : Important Concepts

- Data: labeled instances <x_i, y>, e.g. emails marked spam/not spam
  - Training Set
  - Held-out Set
  - Test Set
- Features: attribute-value pairs which characterize each x
- Experimentation cycle
  - Learn parameters (e.g. model probabilities) on training set
  - (Tune hyper-parameters on held-out set)
  - Compute accuracy of test set
  - Very important: never "peek" at the test set!
- Evaluation
  - Accuracy: fraction of instances predicted correctly
- Overfitting and generalization
  - Want a classifier which does well on test data
  - Overfitting: fitting the training data very closely, but not generalizing well

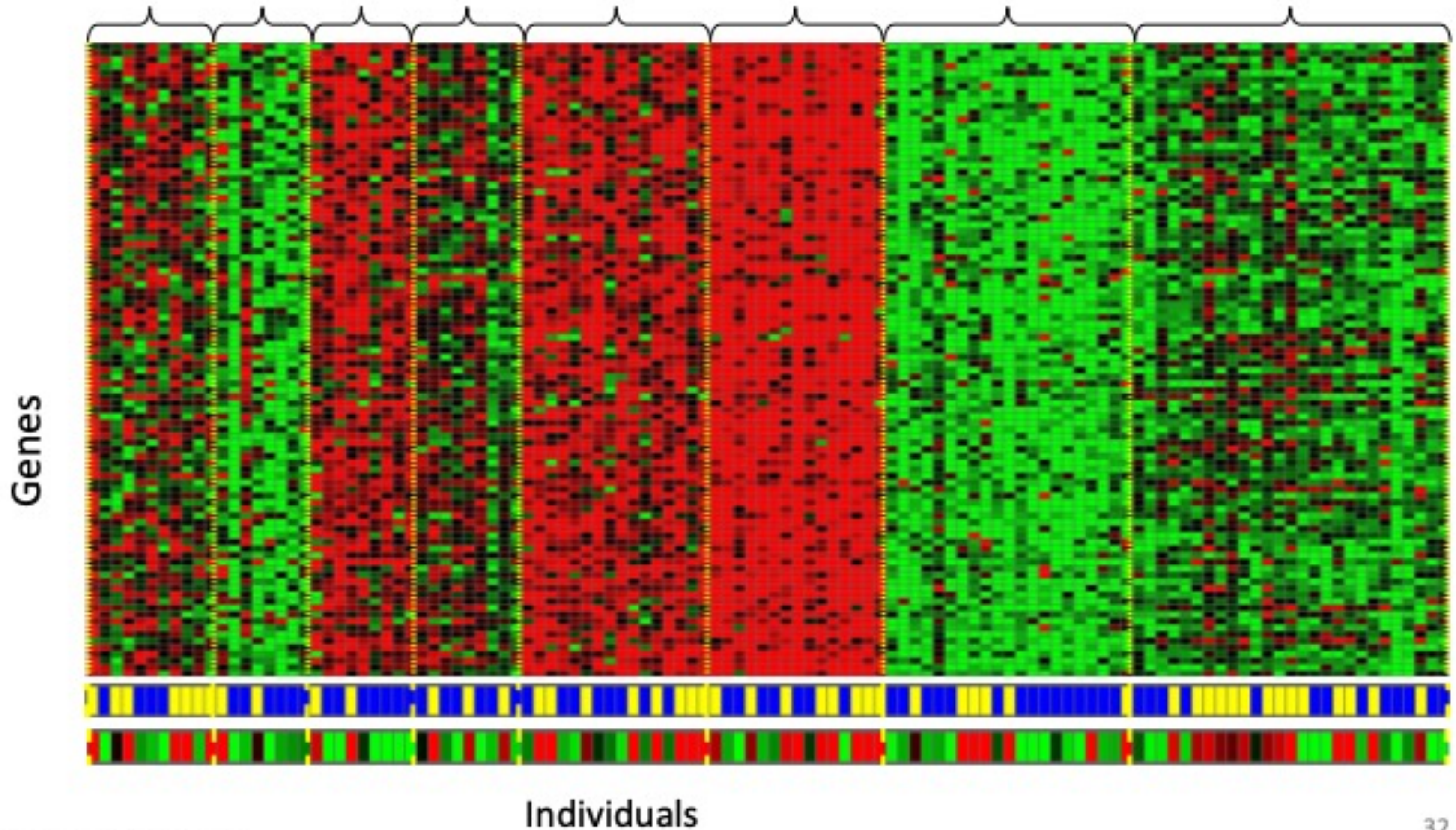
Unsupervised Learning

- Given $x_1, x_2, \ldots, x_n$ (without labels)
- Output hidden structure behind the x's
  - E.g., clustering



Unsupervised Learning

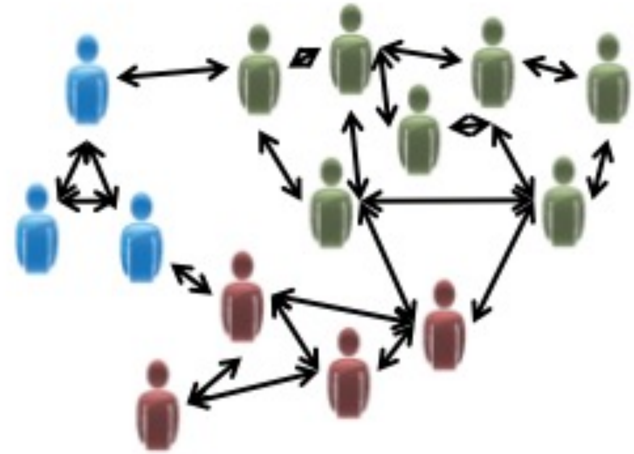
Genomics application: group individuals by genetic similarity



[Source: Daphne Koller]

Unsupervised Learning

Organize computing clusters



Social network analysis

Market segmentation

Astronomical data analysis

33

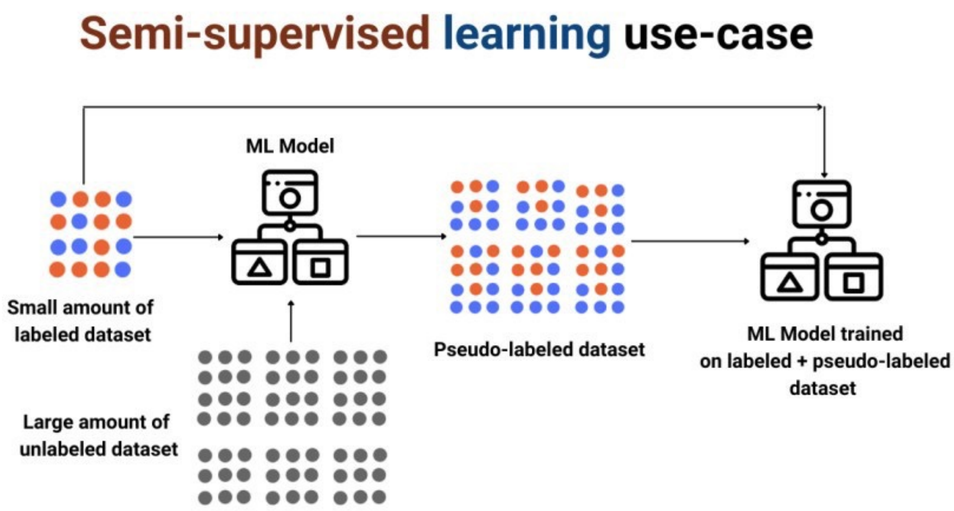
Slide credit: Andrew Ng

Semi-supervised Learning

- Semi-supervised learning is **an approach to machine learning that combines a small amount of labeled data with a large amount of unlabeled data during training**.
  - Semi-supervised learning falls between unsupervised learning (with no labeled training data) and supervised learning (with only labeled training data).
  - Training data very closely, **but not generalizing well**

Examples:

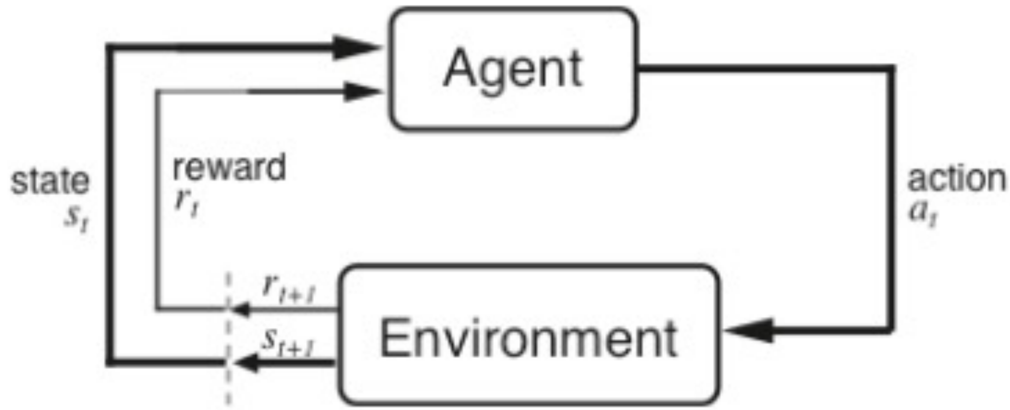
A common example of an application of semi-supervised learning is a **text document classifier**.



Reinforcement Learning

- Given a sequence of states and actions with (delayed) rewards, output a policy
  - Policy is a mapping from states → actions that tells you what to do in a given state
- Examples:
  - Credit assignment problem
  - Game playing
  - Robot in a maze
  - Balance a pole on your hand

Reinforcement Learning



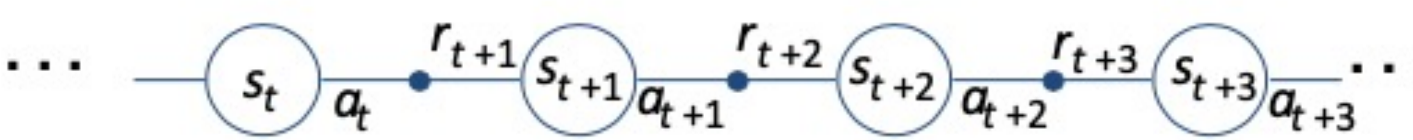
Agent and environment interact at discrete time steps : $t = 0, 1, 2, K$

Agent observes state at step t: $s_t \in S$

produces action at step t : $a_t \in A(s_t)$

gets resulting reward : $r_{t+1} \in \Re$

and resulting next state : s_{t+1}



Slide credit: Sutton & Barto

How and what does machine learn?

Input: X

Output: Y

Label"motorcycle"

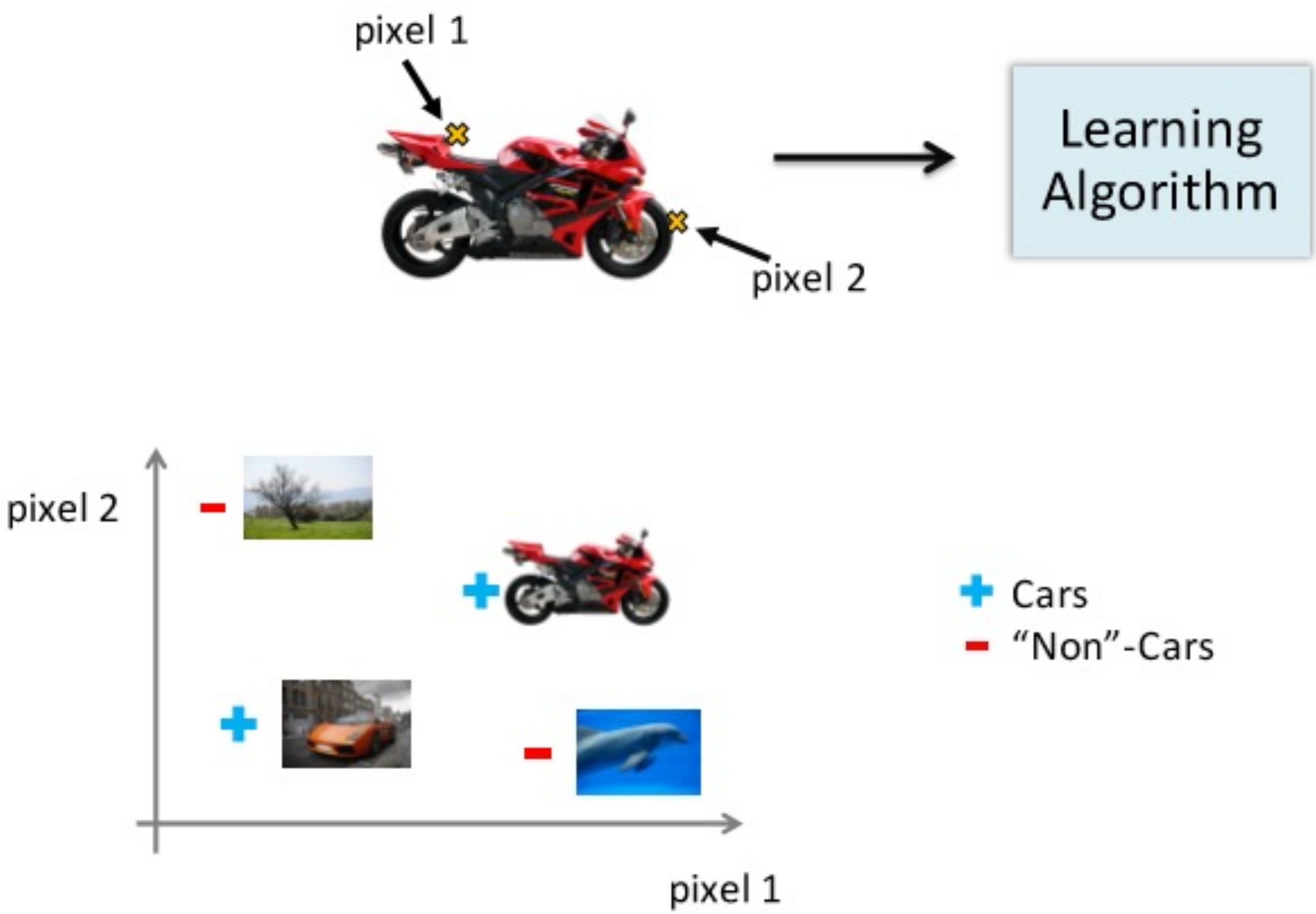
Why is it hard?

You see this

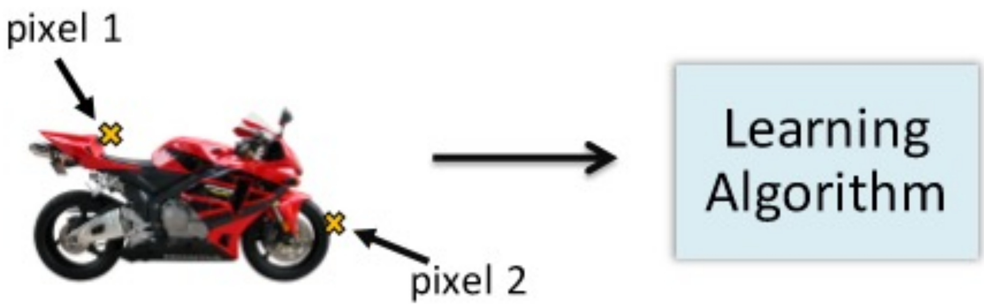
But the camera sees this:

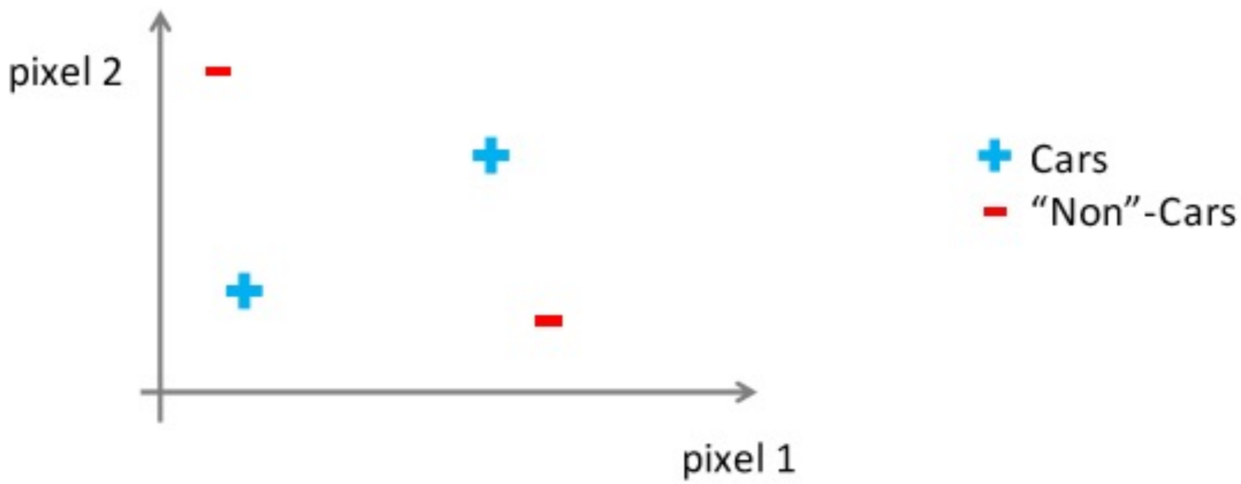
194	210	201	212	199	213	215	195	178	158	182	209
180	189	190	221	209	205	191	167	147	115	129	163
114	126	140	188	176	165	152	140	170	106	78	88
87	103	115	154	143	142	149	153	173	101	57	57
102	112	106	131	122	138	152	147	128	84	58	66
94	95	79	104	105	124	129	113	107	87	69	67
68	71	69	98	89	92	98	95	89	88	76	67
41	56	68	99	63	45	60	82	58	76	75	65
20	43	69	75	56	41	51	73	55	70	63	44
50	50	57	69	75	75	73	74	53	68	59	37
72	59	53	66	84	92	84	74	57	72	63	42
67	61	58	65	75	78	76	73	59	75	69	50

Raw Image Representation

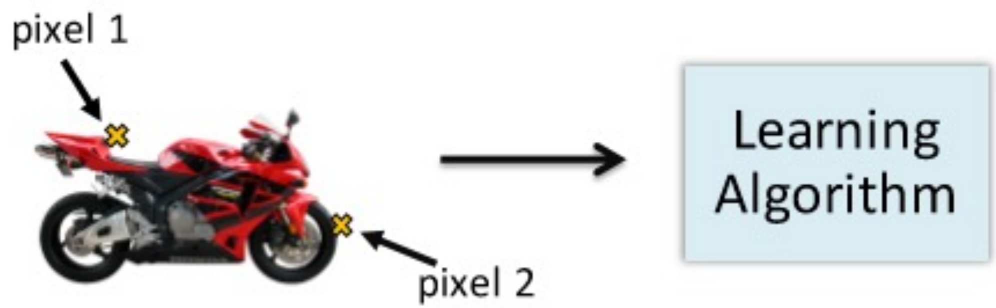


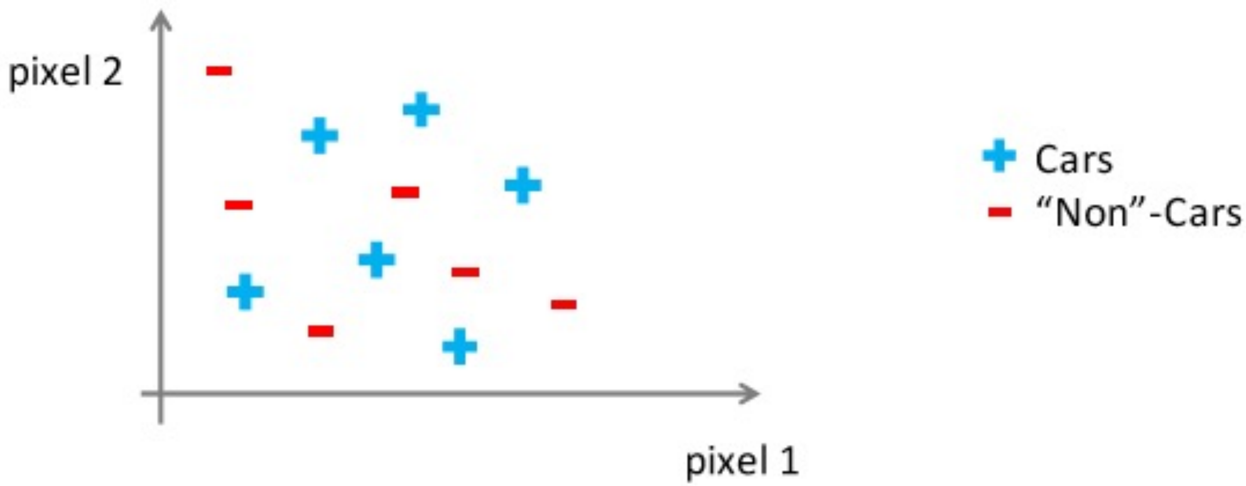
Raw Image Representation



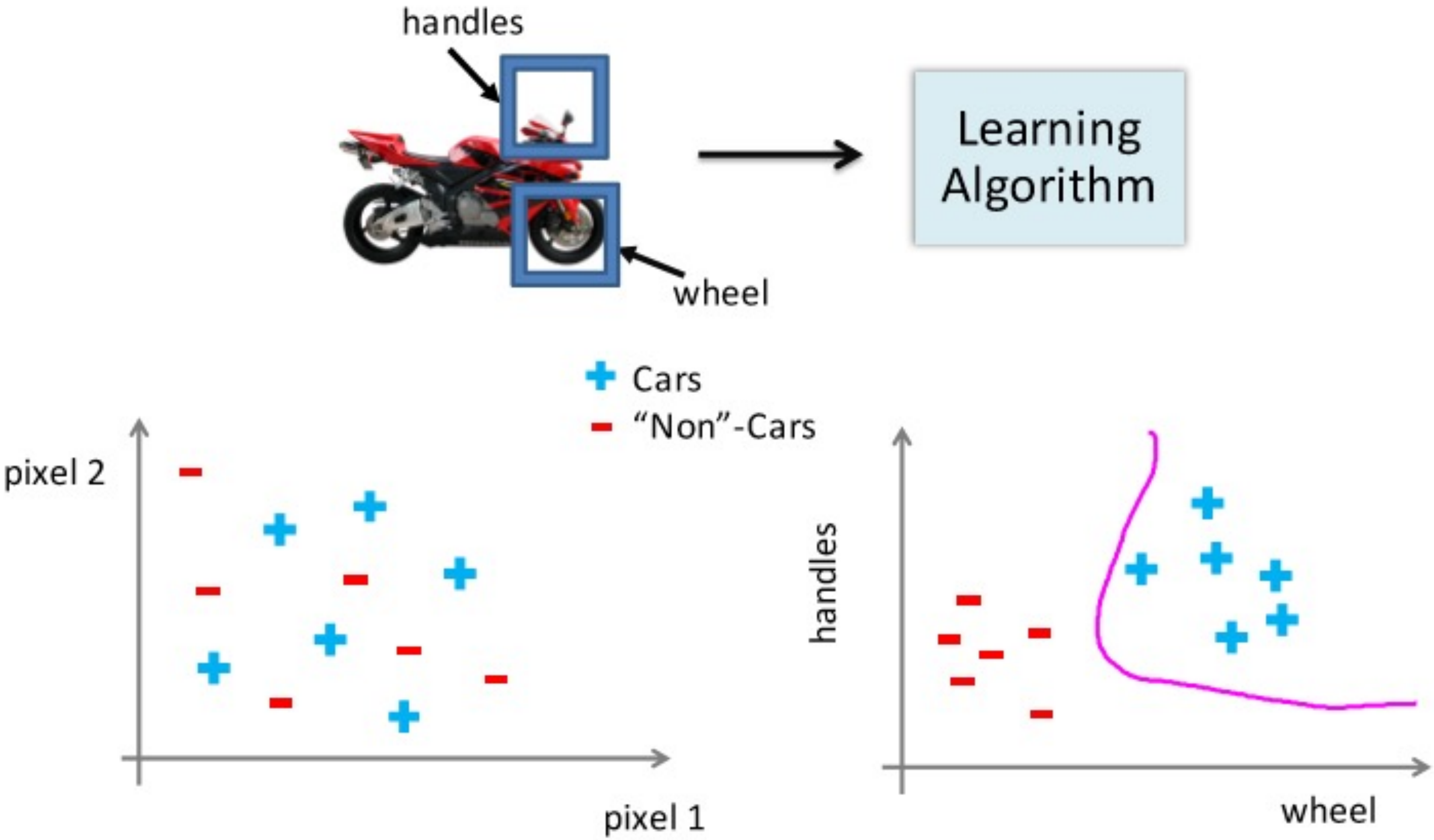


Raw image representation





Better feature representation



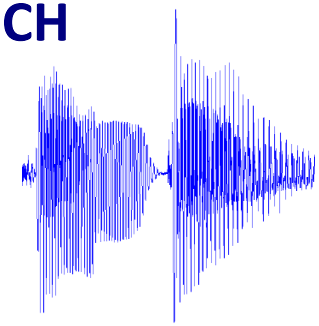
Traditional Machine Learning

VISION

→ **hand-crafted features** **SIFT/HOG** → **your favorite classifier** → "car"

fixed | learned

SPEECH



→ **hand-crafted features** **MFCC** → **your favorite classifier** → \ˈd ē p\

fixed | learned

NLP

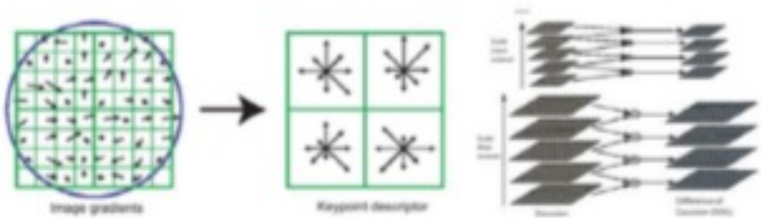
This burrito place is yummy and fun!

→ **hand-crafted features** **Bag-of-words** → **your favorite classifier** → "+"

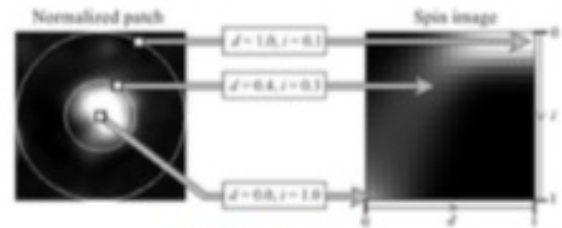
fixed | learned

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

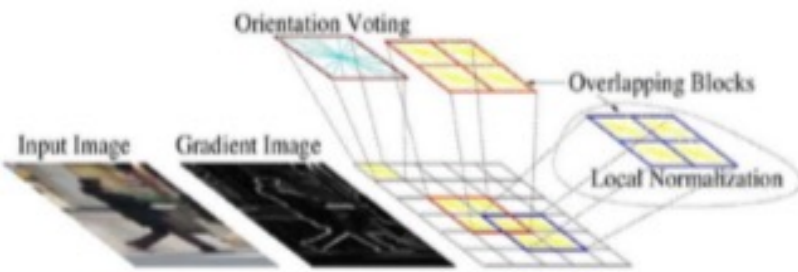
Feature representation methods



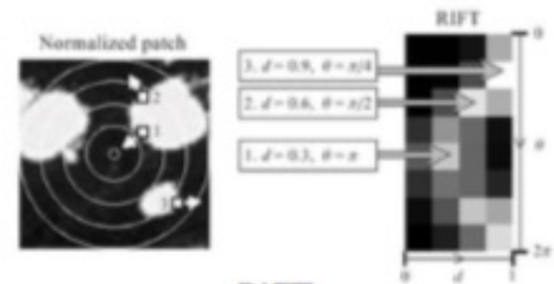
SIFT



Spin image

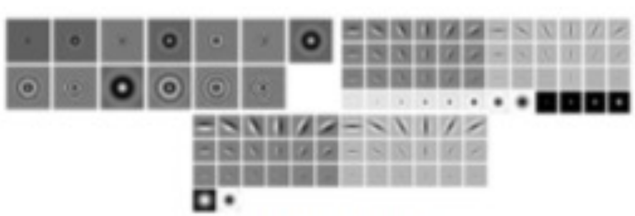


HoG

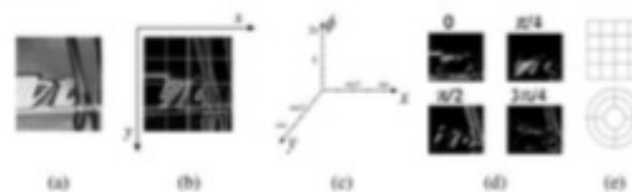


RIFT

Expert Knowledge!



Textons



GLOH

Source: feature representations in computer vision(Honglak lee)

Traditional Machine Learning (more accurately)

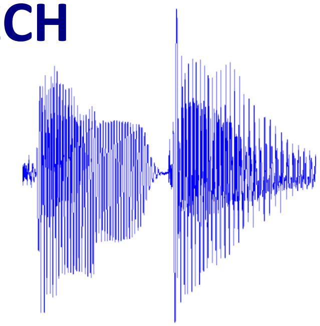
"Learned"

VISION

SIFT/HOG → K-Means/ pooling → classifier → "car"

fixed | unsupervised | supervised

SPEECH



MFCC → Mixture of Gaussians → classifier → \ˈd ē p\

fixed | unsupervised | supervised

NLP

This burrito place is yummy and fun!

Parse Tree Syntactic → n-grams → classifier → "+"

fixed | unsupervised | supervised

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

[Mel Frequency Cepstral Coefficient (MFCC)]

Properties of Machine Learning approaches

- (Hierarchical) Compositionality
  - Cascade of non-linear transformations
  - Multiple layers of representations

- End-to-End Learning
  - Learning (goal-driven) representations
  - Learning to feature extraction

- Distributed Representations , Scalability, and Genericity
  - No single neuron “encodes” everything
  - Groups of neurons work together

Hierarchical Compositionality

VISION

pixels → edge → texton → motif → part → object

SPEECH

sample → spectral band → formant → motif → phone → word

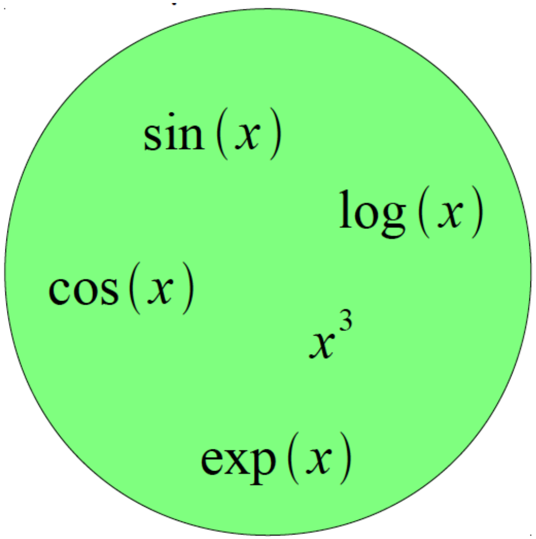
NLP

character → word → NP/VP/.. → clause → sentence → story

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

Building A Complicated Function

Given a library of simple functions

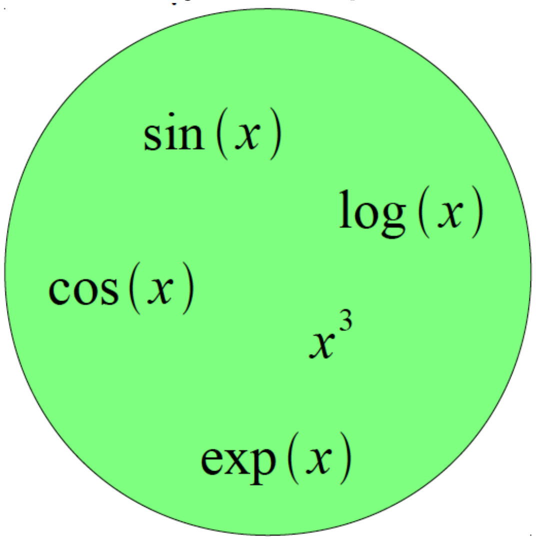


Compose into a

complicate function

Building A Complicated Function

Given a library of simple functions

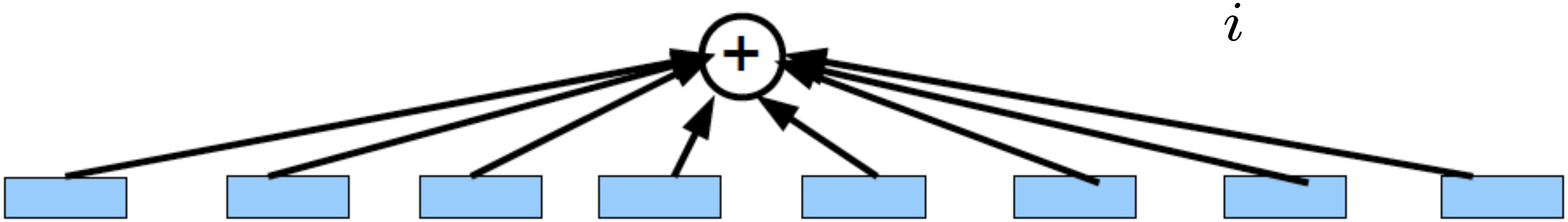


Compose into a complicate function

Idea 1: Linear Combinations

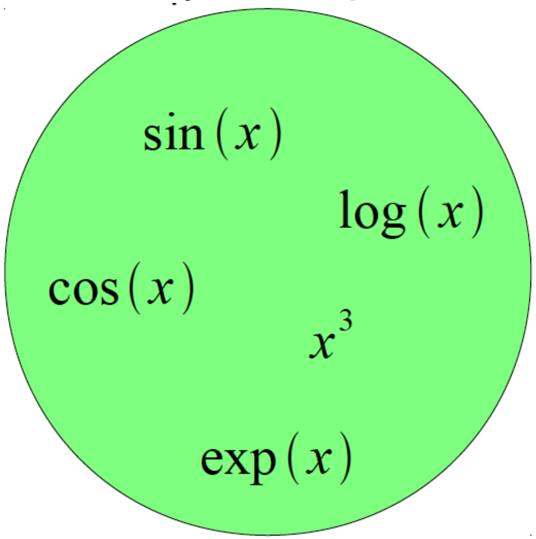
- Boosting
- Kernels
- …

$$f(x) = \sum_i \alpha_i g_i(x)$$



Building A Complicated Function

Given a library of simple functions

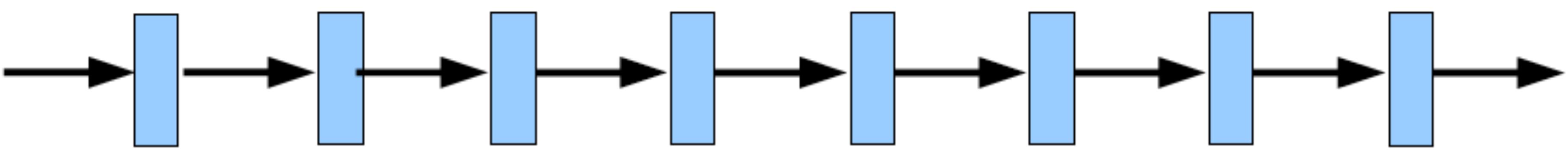


Compose into a complicate function

Idea 2: Compositions

- **Deep Learning**
- Grammar models
- Scattering transforms…

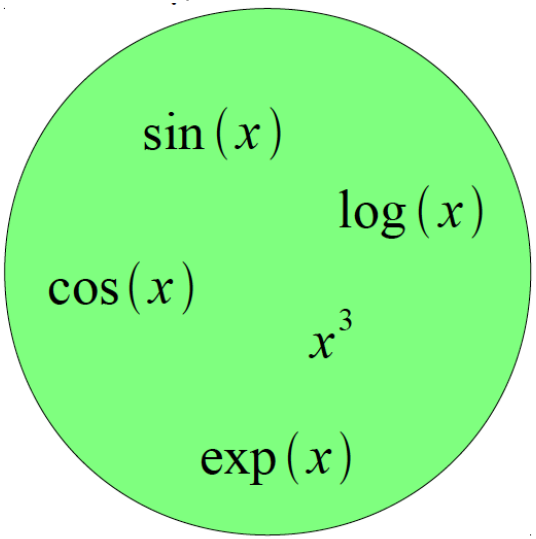
$$f(x) = g_1(g_2(\ldots(g_n(x)\ldots))$$



Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

Building A Complicated Function

Given a library of simple functions

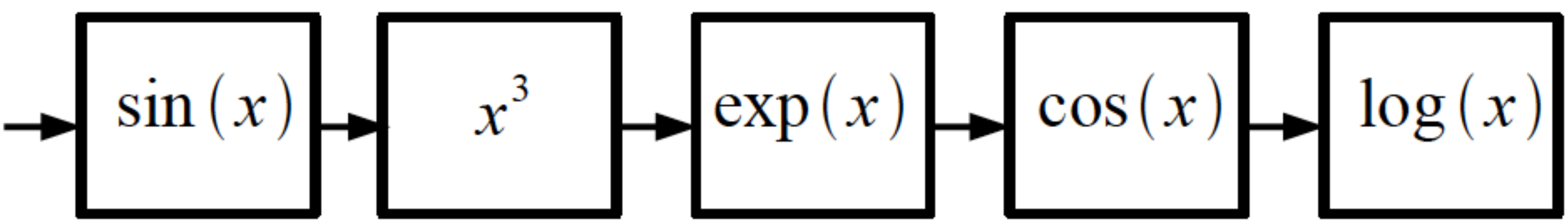


Compose into a complicate function

Idea 2: Compositions

- **Deep Learning**
- Grammar models
- Scattering transforms…

$$f(x) = \log(\cos(\exp(\sin^3(x))))$$



Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

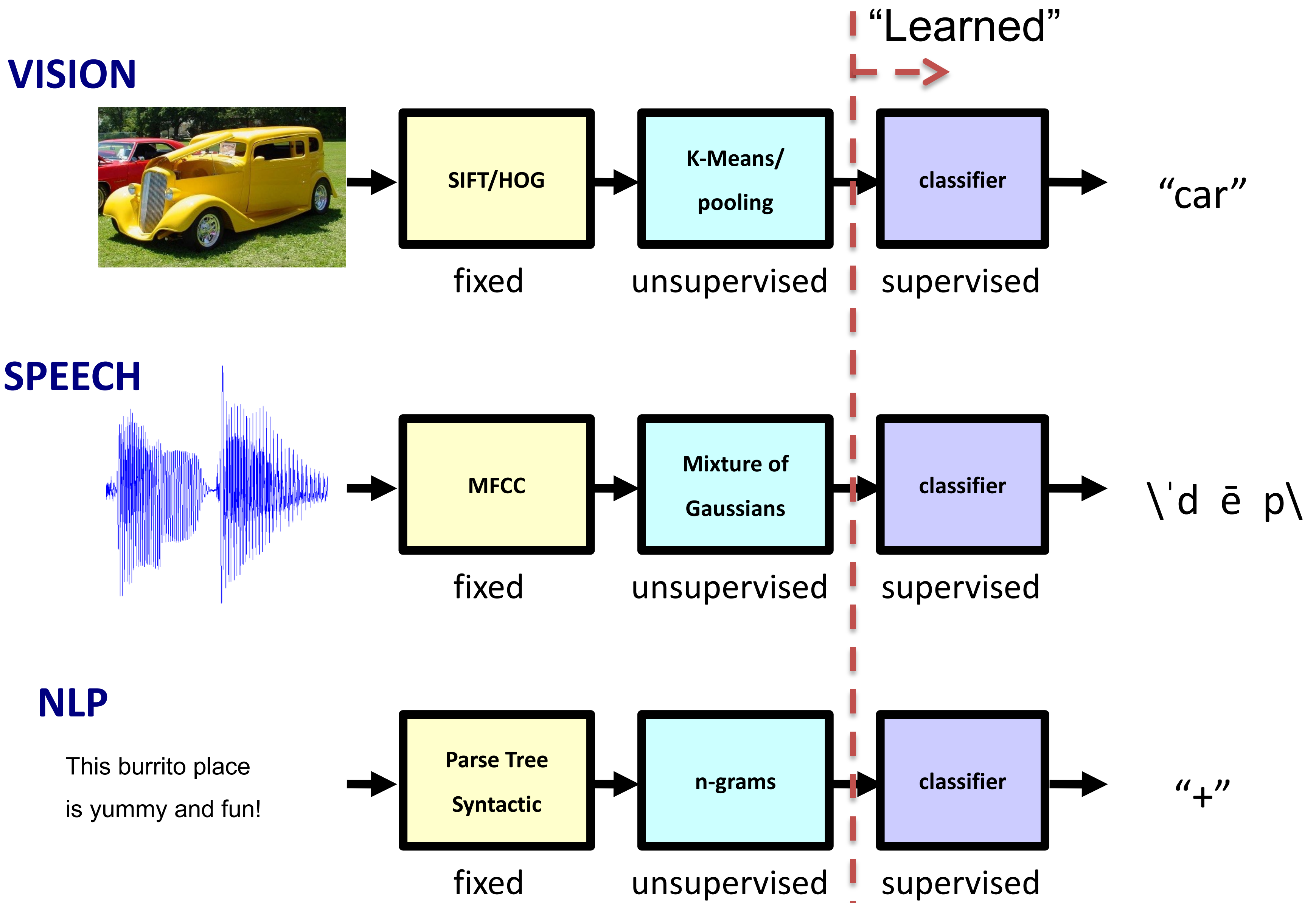
Properties of Deep (Machine) Learning approaches

- (Hierarchical) Compositionality
  - Cascade of non-linear transformations
  - Multiple layers of representations

- End-to-End Learning
  - Learning (goal-driven) representations
  - Learning to feature extraction

- Distributed Representations , Scalability, and Genericity
  - No single neuron “encodes” everything
  - Groups of neurons work together

Machine Learning = End-to-End Learning ?



“Shallow” vs Deep Learning

- “Shallow” models

hand-crafted Feature Extractor → “Simple” Trainable Classifier

fixed | learned

- Deep models

Trainable Feature-Transform / Classifier → … → Trainable Feature-Transform / Classifier → Trainable Feature-Transform / Classifier →

Learned Internal Representations

Slide Credit: Marc'Aurelio Ranzato, Yann LeCun

Properties of Machine Learning approaches

- (Hierarchical) Compositionality
  - Cascade of non-linear transformations
  - Multiple layers of representations

- End-to-End Learning
  - Learning (goal-driven) representations
  - Learning to feature extraction

- Distributed **Representations , Scalability, and Genericity**
  - No single neuron “encodes” everything
  - Groups of neurons work together

One Model To Learn Them All

Kaiser, Lukasz, Aidan N. Gomez, Noam Shazeer, Ashish Vaswani, Niki Parmar, Llion Jones, and Jakob Uszkoreit. "One model to learn them all." *arXiv preprint arXiv:1706.05137*(2017).

ML in a Nutshell

- Tens of thousands of machine learning algorithms
  - Hundreds new every year

- Every ML algorithm has three components:
  - **Representation**
  - **Optimization**
  - **Evaluation**

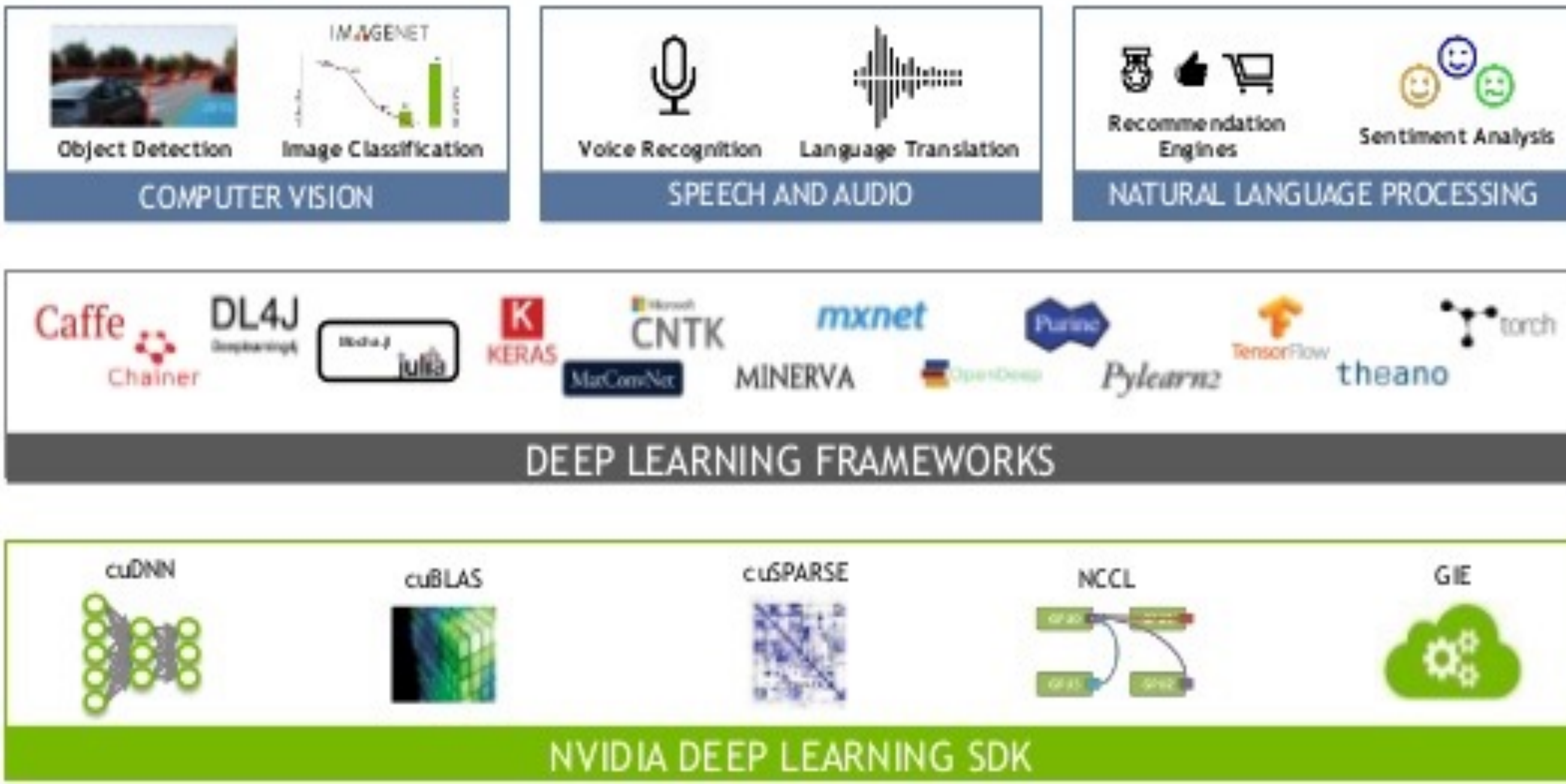
ML in practice

- Understand domain, prior knowledge, and goals
- Data integration, selection, cleaning, pre-processing, etc.
- Learn models
- Interpret results
- Consolidate and deploy discovered knowledge

Based on a slide by Pedro Domingos

Powering the Deep Learning Ecosystem

NVIDIA SDK Accelerates Every Major Framework



per.nvidia.com/deep-learning-software

Summary

- Machine learning system and types
- How and what does machine learn?
- How to design ML system?

- What’s next:
  - Data acquisition and labeling
  - Data preparation