Institute for Advanced Studies
in Basic Sciences
Gava Zang, Zanjan, Iran

Introduction to Machine Learning

By Dr. Parvin Razzaghi (p.razzaghi@iasbs.ac.ir)

Winter 2024

Slide credit

- Professor
  - Dr. Parvin Razzaghi

- Website
  - http://www.seas.upenn.edu/~cis519

- Original Slides:
  - Eric Eaton

- Acknowledgment
  - These slides have been prepared based on the materials created by Eric Eaton, with sincere appreciation for the many educators who have generously shared their course materials online. These slides may be reused or adapted for academic purposes, provided that appropriate attribution is given. For any comments or suggestions, please contact: p.razzaghi@iasbs.ac.ir

Course Information

- Slides and Assignments:
  - Accessible through the group:

- Quera:
  - https://quera.org/course/add_to_course/course/20341/
  - Password: ml1403

Communication

- **Contacting the Professor:**
  - p.razzaghi@iasbs.ac.ir

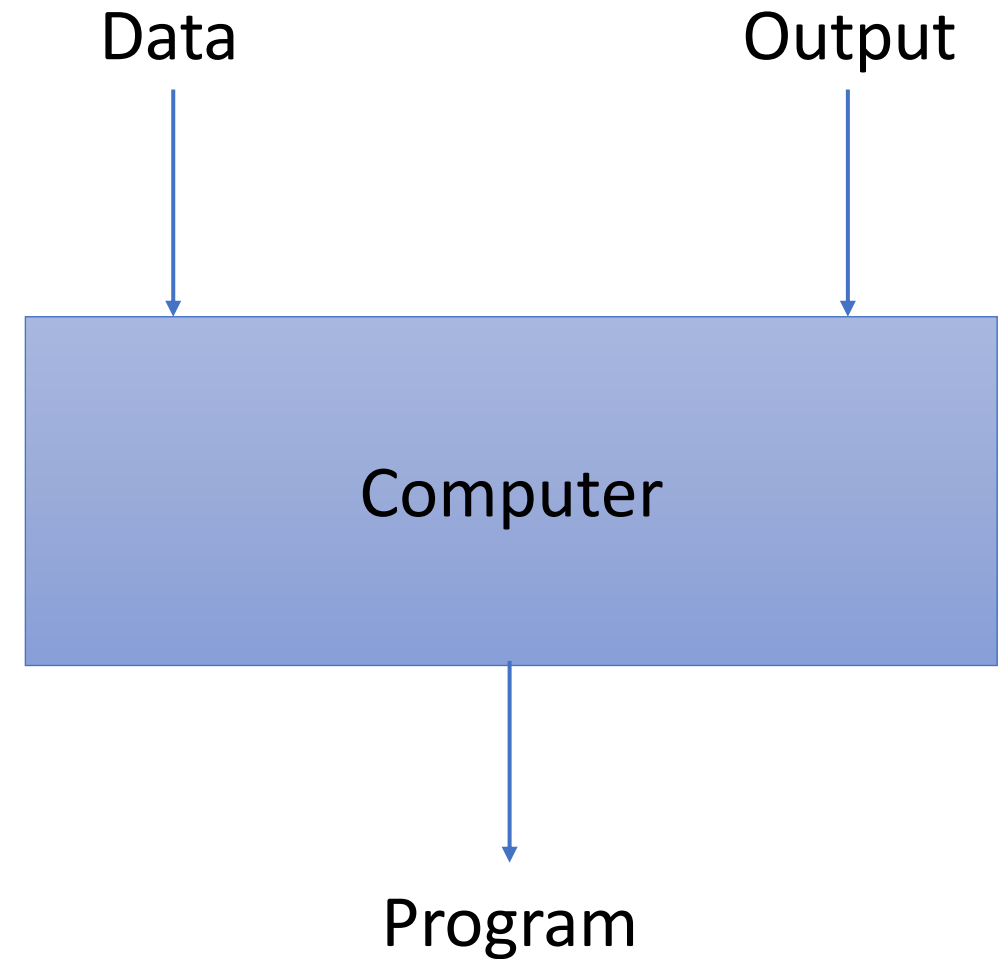- **Contacting the Teaching Assistants (TAs):**
  - TAs' Information:
    - Elaheh Sheikh: [elahehsheikhi@iasbs.ac.ir] | [@elisheikhi]
    - Maryam Lotfi: [maryam.lotfi@iasbs.ac.ir] | [@marymltf]
    - Reza Pishkar: [r.pishkar@iasbs.ac.ir] | [@rpishkr]
    - Mohammadreza Naderlou: [m.naderlo@iasbs.ac.ir] | [@mimre_non]

Traditional Programming

Machine Learning

Data   Program

Data   Output

Computer

Computer

Output

Program
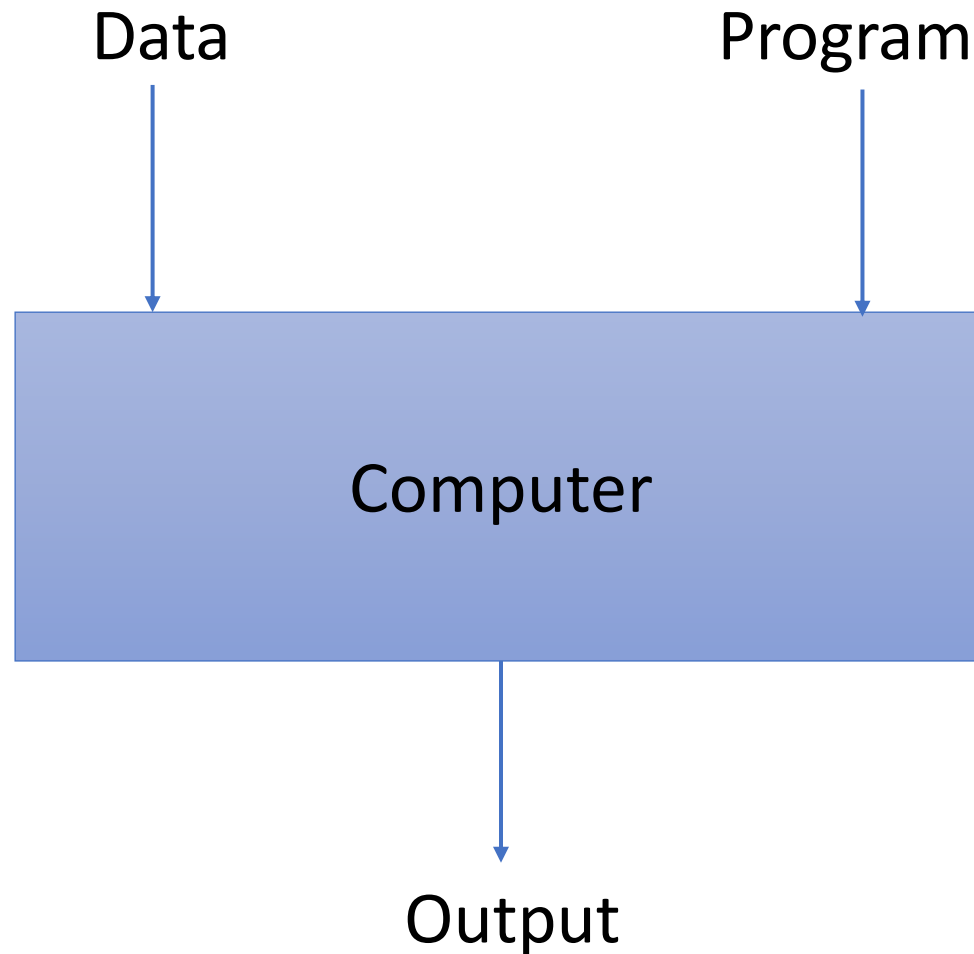
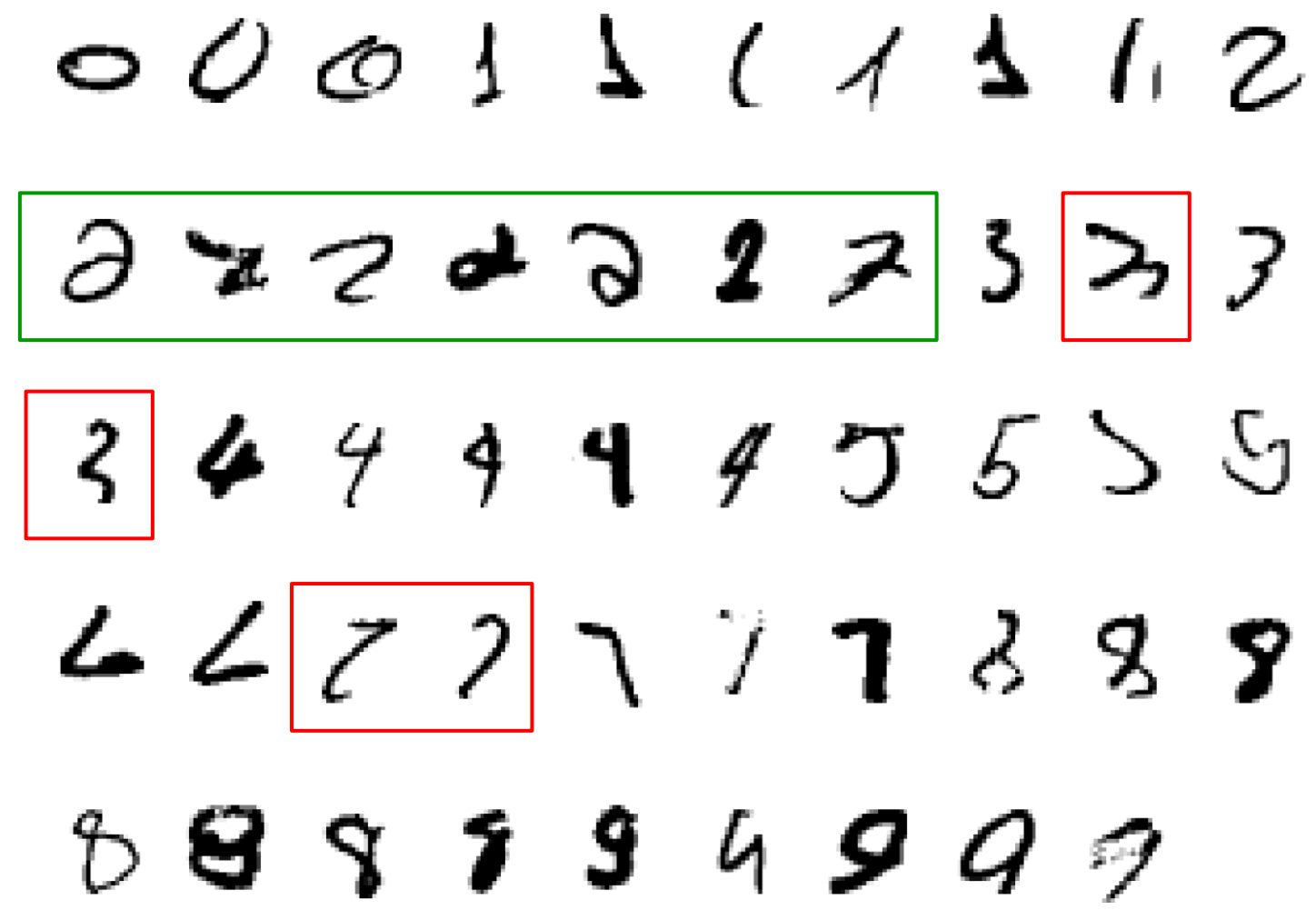# When Do We Use Machine Learning?

- ML is used when:

  - Human expertise does not exist (navigating on Mars)

  - Humans can't explain their expertise (speech recognition)

  - Models must be customized (personalized medicine)

  - Models are based on huge amounts of data (genomics)

A classic example of a task that requires machine learning:

It is very hard to say what makes a 2

Some more examples of tasks that are best solved by using a learning algorithm

- Recognizing patterns:
  - Facial identities or facial expressions
  - Handwritten or spoken words
  - Medical images

- Generating patterns:
  - Generating images or motion sequences

- Recognizing anomalies:
  - Unusual credit card transactions
  - Unusual patterns of sensor readings in a nuclear power plant

- Prediction:
  - Future stock prices or currency exchange rates

Slide credit: Geoffrey Hinton

Sample Applications

- Web search
- Computational biology

- Finance
- E-commerce
- Space exploration

- Robotics
- Information extraction

- Social networks
- Debugging software
- ….

Samuel's Checkers-Player

"Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed."

Arthur Samuel (1959)

What is Machine Learning?

"Learning is any process by which a system improves performance from experience."

Herbert Simon

- Definition by Tom Mitchell (1998):

  - Machine Learning is the study of algorithms that

    - improve their performance P

    - at some task T

    - with experience E.

  - A well-defined learning task is given by $< P, T, E >$.

Defining the Learning Task

Improve on task T, with respect to performance metric P, based on experience E

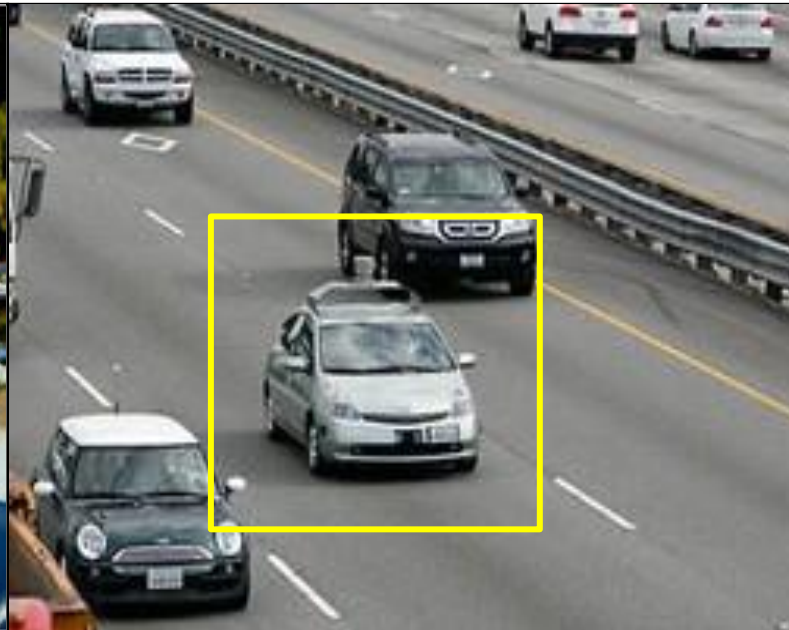|	

• T: Playing checkers
 	
• P: Percentage of games won against an
 	
  arbitrary opponent
 	
• E: Playing practice games against itself
	• T: Recognizing hand-written words
	
	• P: Percentage of words correctly classified
	
	• E: Database of human-labeled images of
	
	  handwritten words

• T: Driving on four-lane highways using vision sensors • P: Average distance traveled before a human-judged error • E: A sequence of images and steering commands recorded while observing a human driver.	• T: Categorize email messages as spam or legitimate. • P: Percentage of email messages correctly classified. • E: Database of emails, some with human-given labels

State of the Art Applications of Machine Learning

Autonomous Cars

- Nevada made it legal for autonomous cars to drive on roads in June 2011
- As of 2013, four states (Nevada, Florida, California, and Michigan) have legalized autonomous cars.
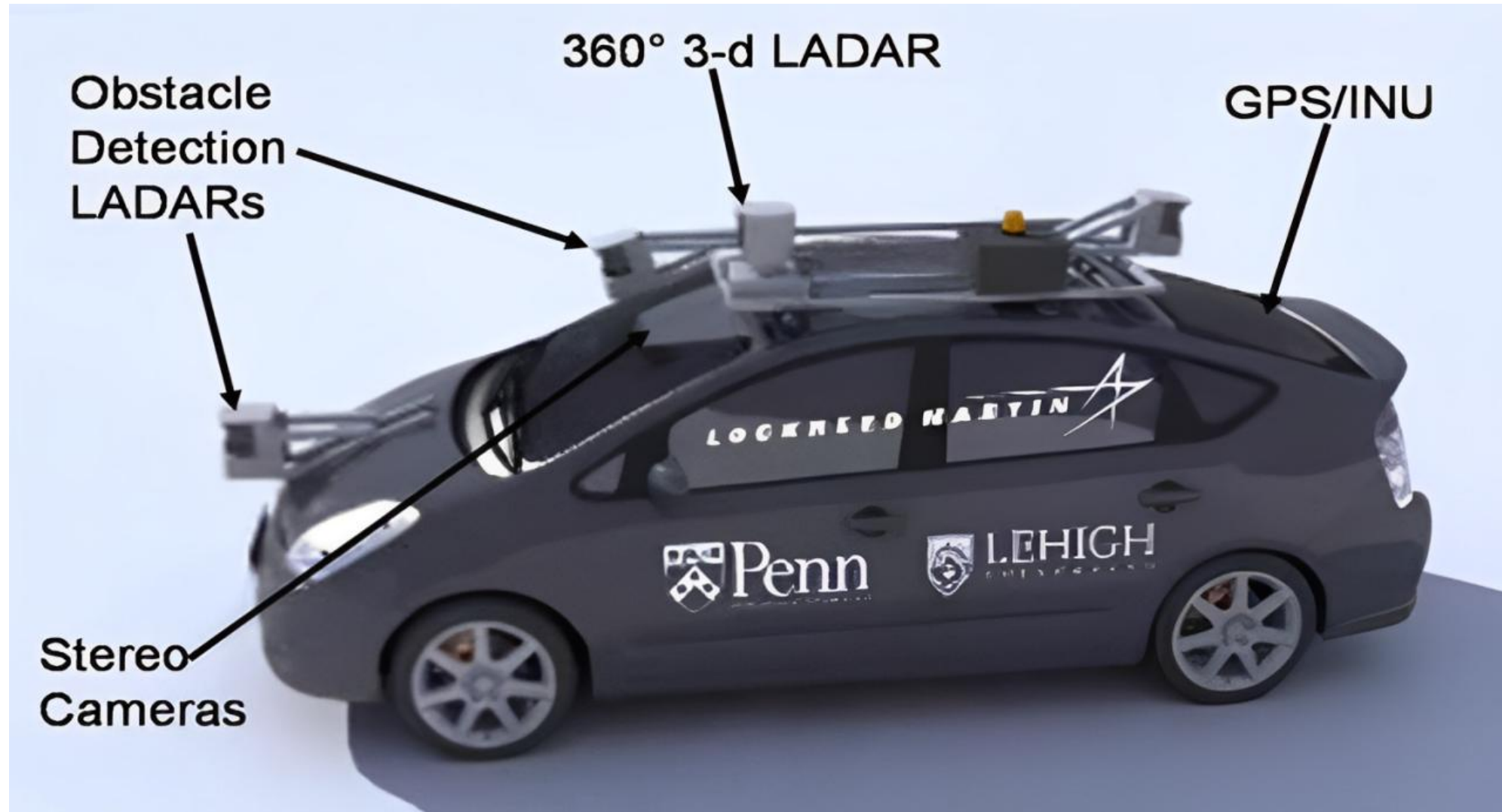
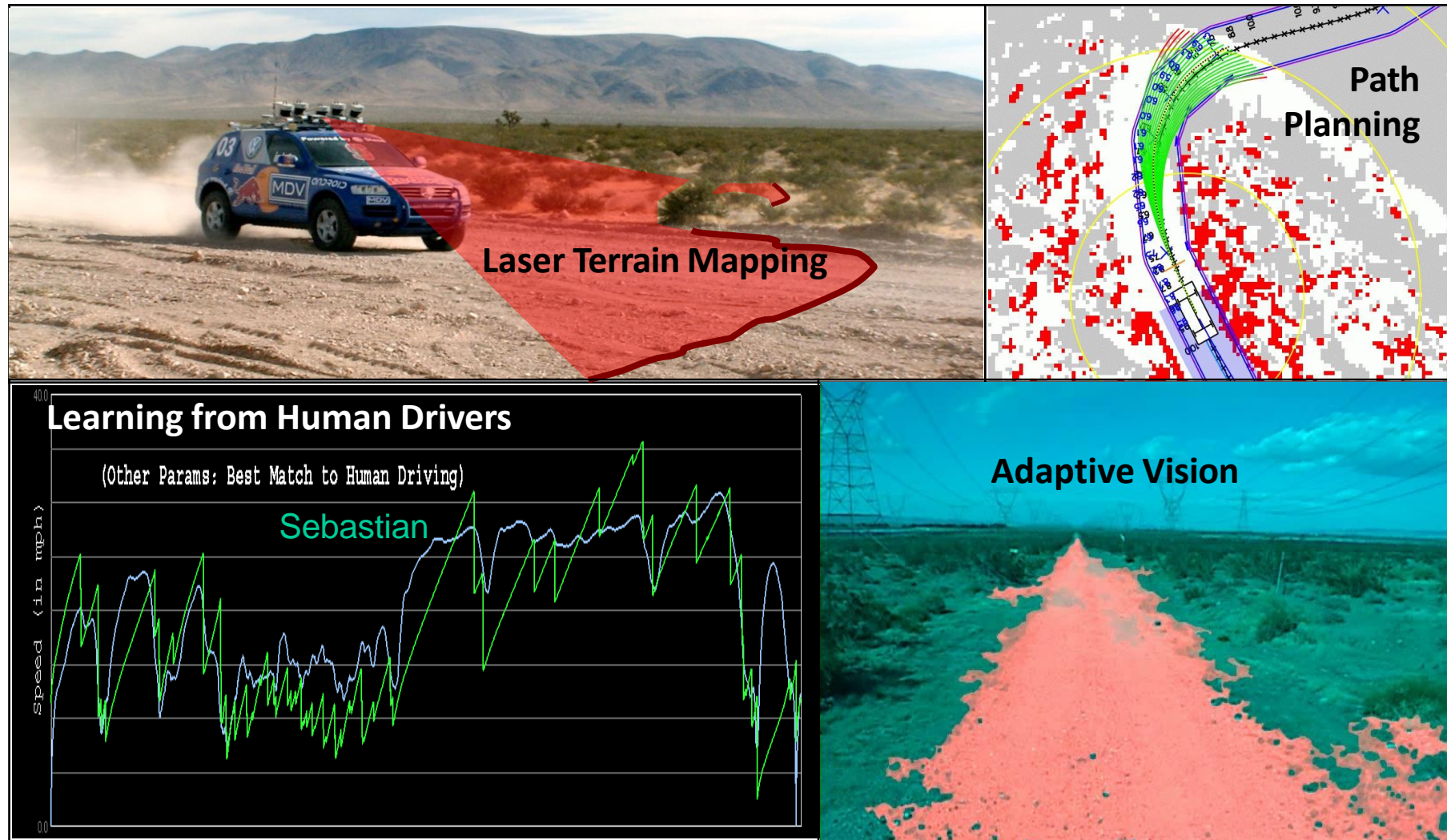Penn's Autonomous Car (Ben Franklin Racing Team)

Autonomous Car Sensors

Autonomous Car Technology

Laser Terrain Mapping

Path Planning

Learning from Human Drivers

Adaptive Vision

Deep Learning in the Headlines

Deep Belief Net on Face Images

object models

object parts (combination of edges)

edges

pixels
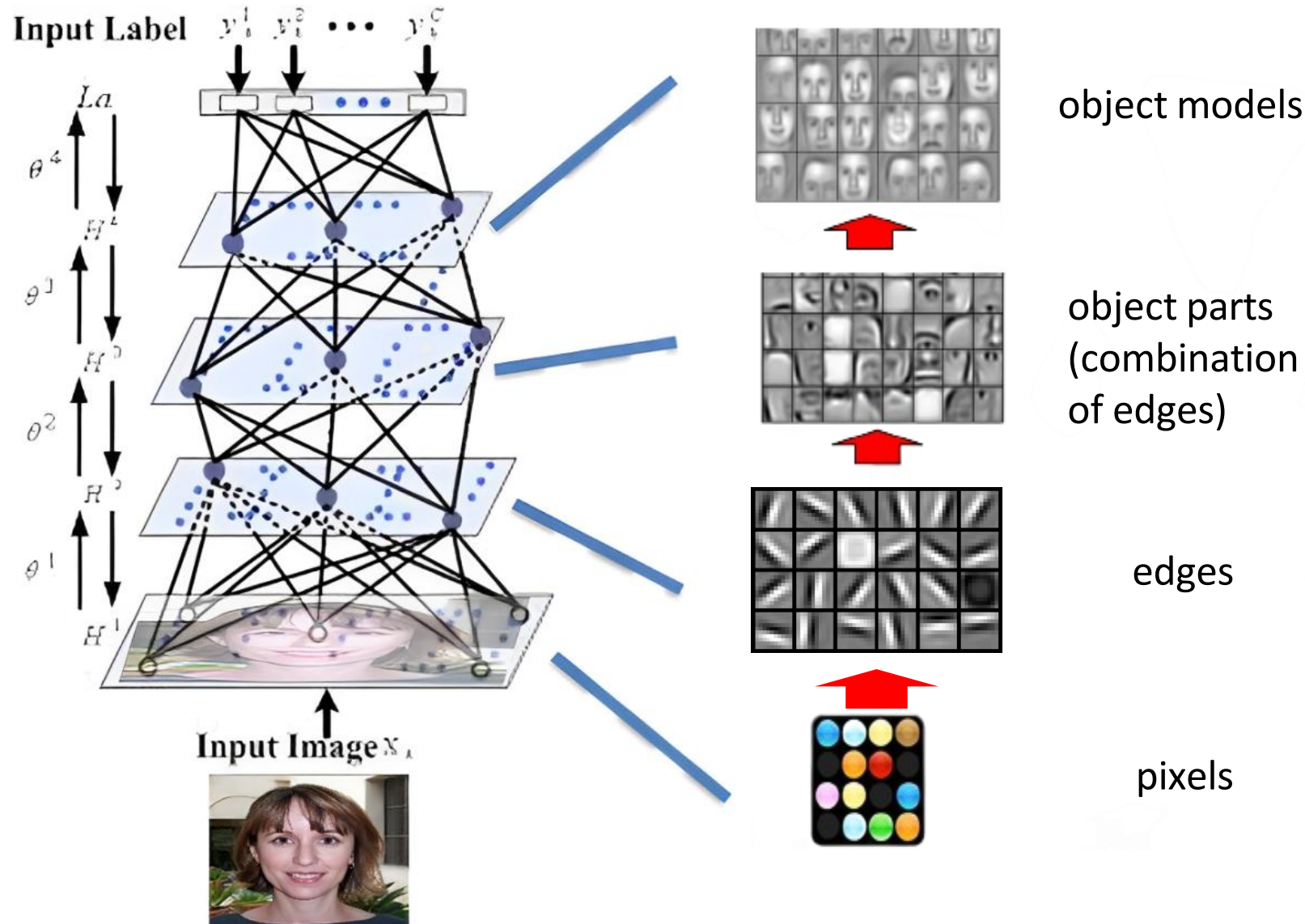
Training on Multiple Objects

- Trained on 4 classes (cars, faces, motorbikes, airplanes).

- Second layer: Shared-features and object-specific features.

- Third layer: More specific features.
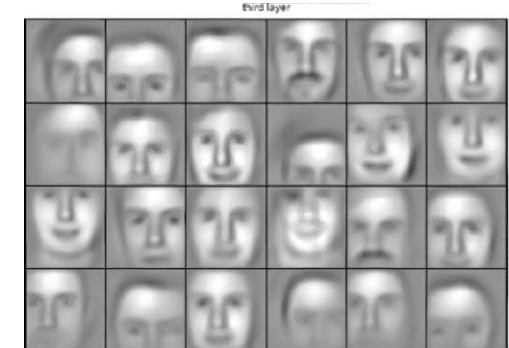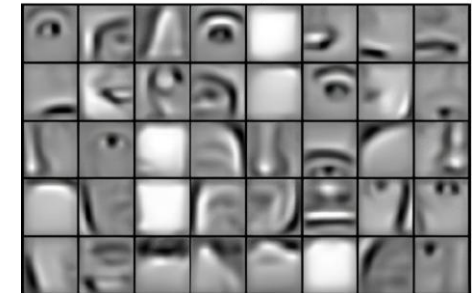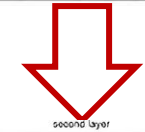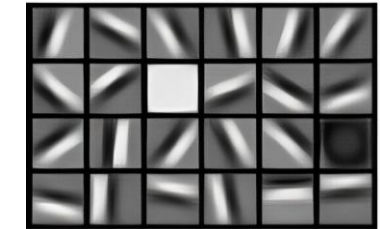
Scene Labeling via Deep Learning

Machine Learning in Automatic Speech Recognition

A Typical Speech Recognition System

Feature Extraction → Neural Network → Decoder → Transducer & Language Model → *Hi, how are you?*
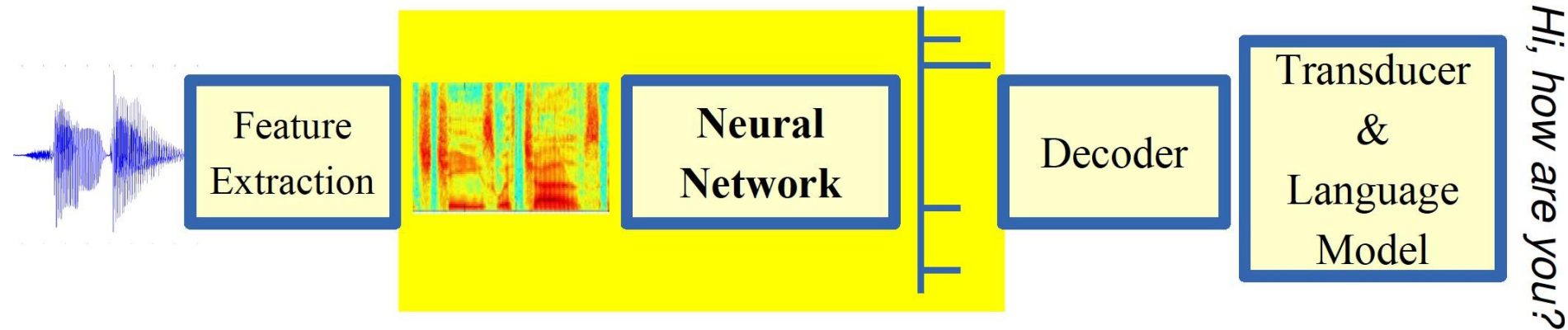
ML used to predict of phone states from the sound spectrogram
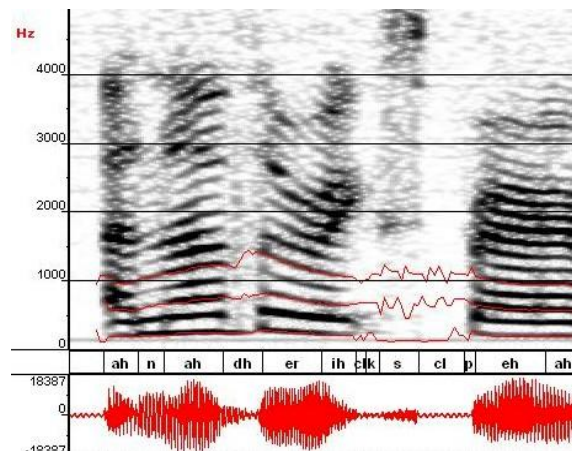
Deep learning has state-of-the-art results

# Hidden Layers	1	2	4	8	10	12
Word Error Rate %	16.0	12.8	11.4	10.9	11.0	11.1

Baseline GMM performance = 15.4%

[Zeiler et al. "On rectified linear units for speech recognition" ICASSP 2013]

Impact of Deep Learning in Speech Technology
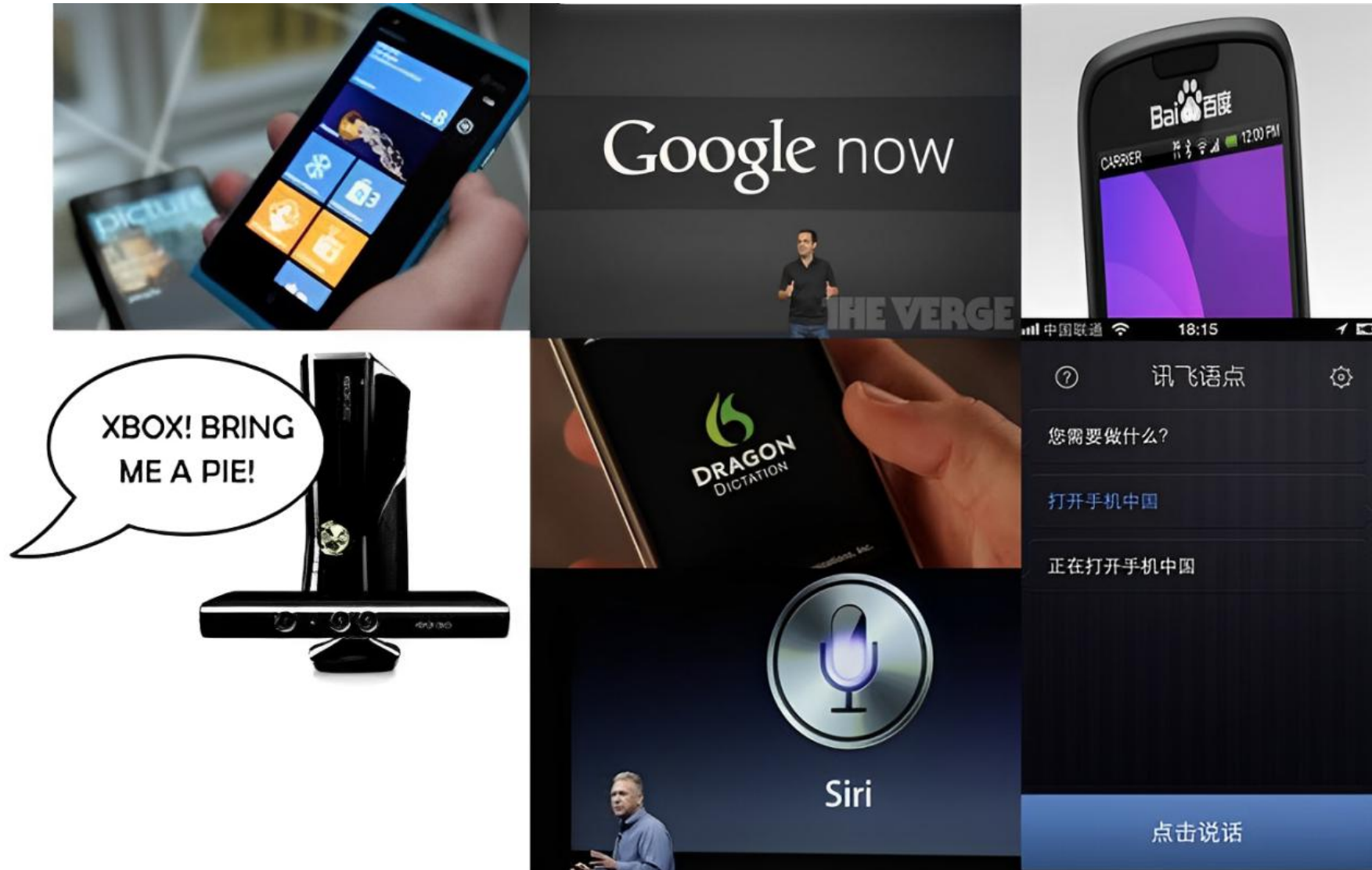
Types of Learning

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

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
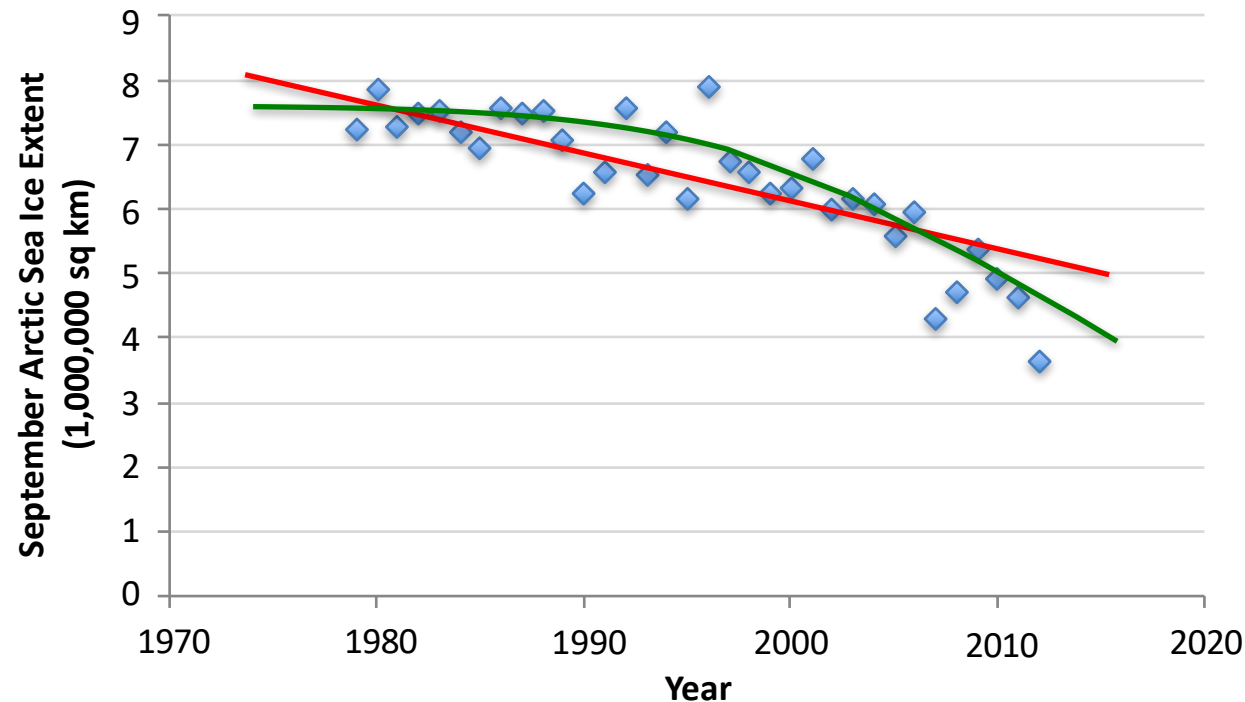
Supervised Learning: Classification

- 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 categorical == classification

Breast Cancer (Malignant / Benign)

Tumor Size

Supervised Learning: Classification

- 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 categorical == classification

Breast Cancer (Malignant / Benign)
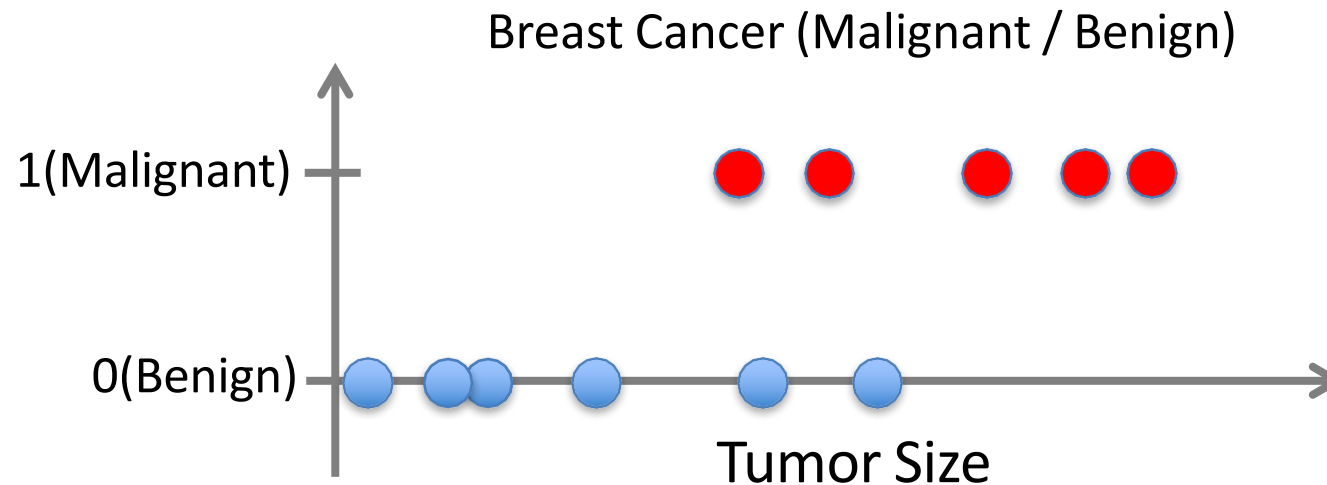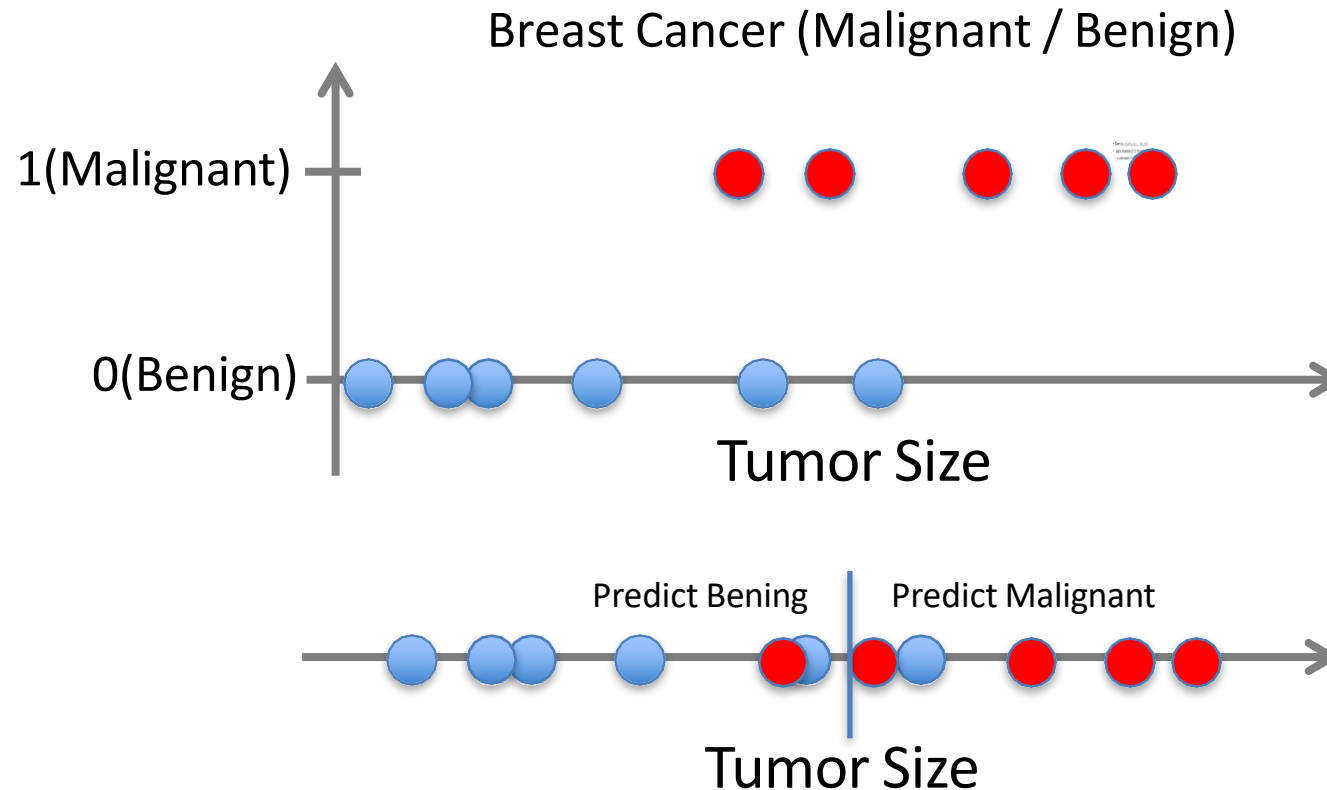
Supervised Learning

- x can be multi-dimensional
  - Each dimension corresponds to an attribute
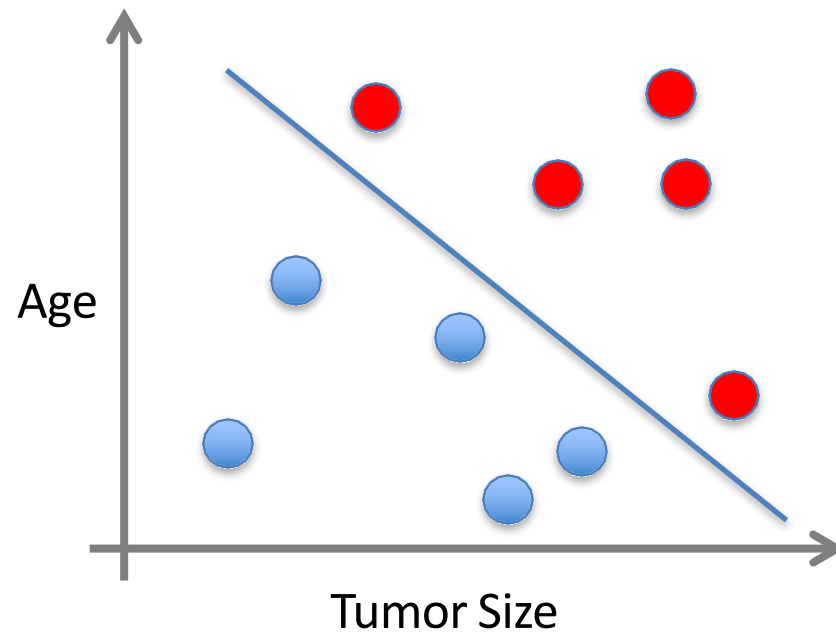
- Clump Thickness
- Uniformity of Cell Size
- Uniformity of Cell Shape

…

Unsupervised Learning

- Given $x_1, x_2, \ldots, x_n$ (without labels)

- Output hidden structure behind the x's

  - E.g., clustering
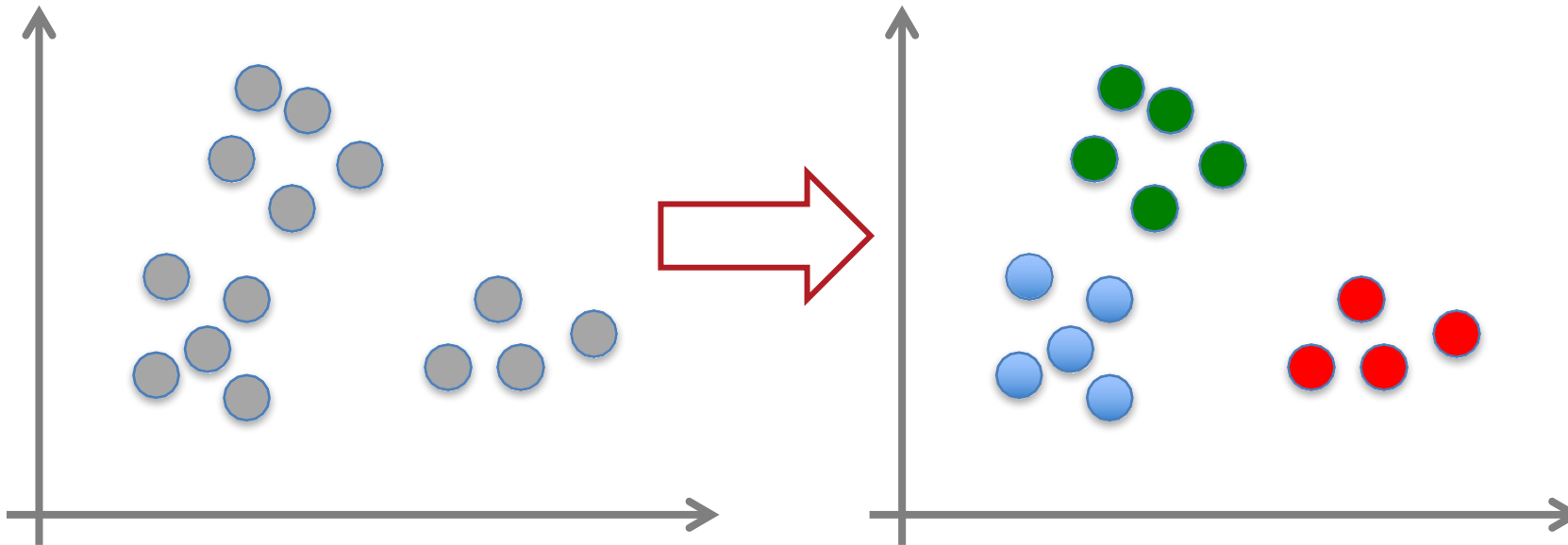
Unsupervised Learning

Genomics application: group individuals by genetic similarity

[Source: Daphne Koller]
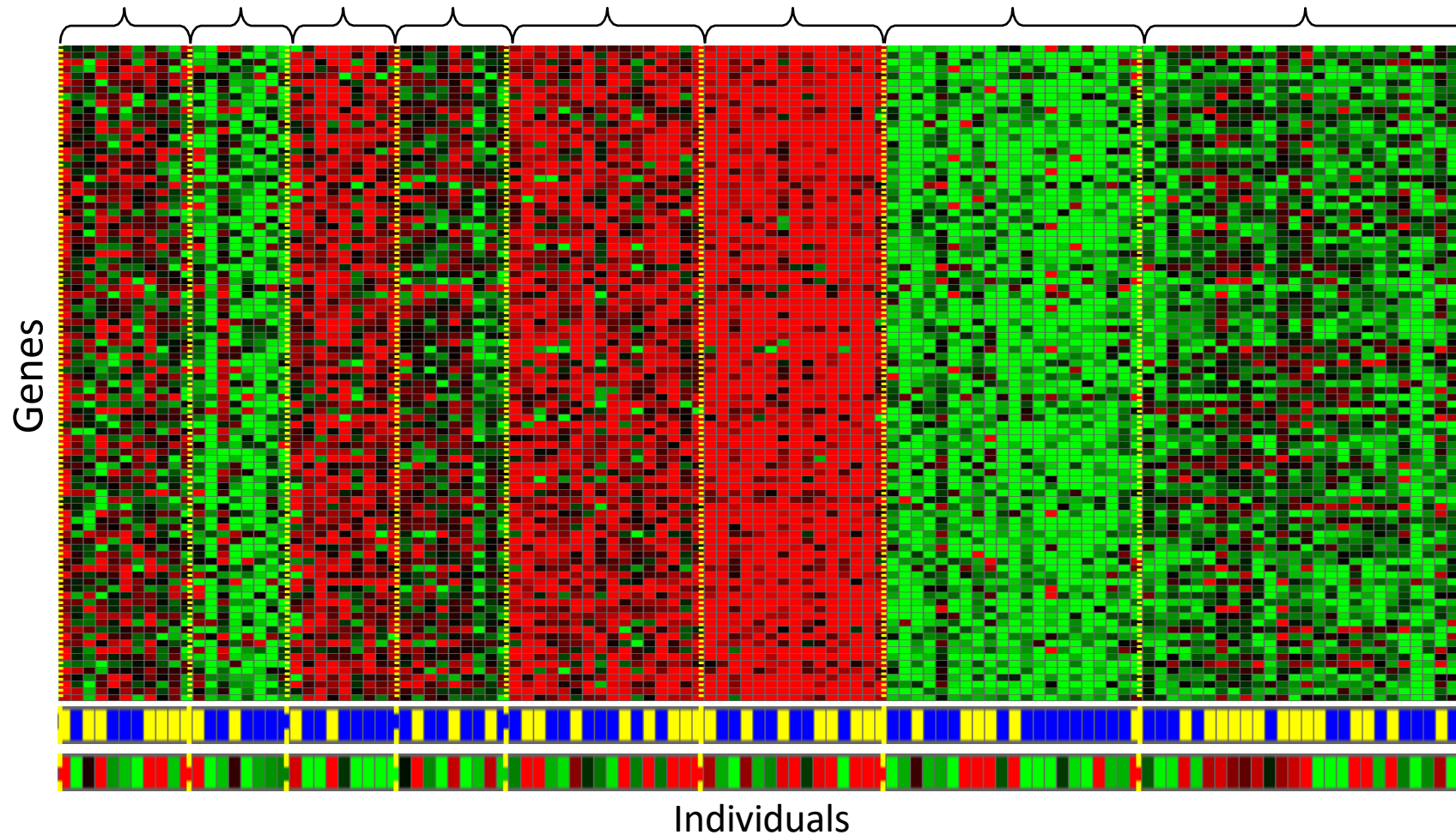
Unsupervised Learning

Organize computing clusters
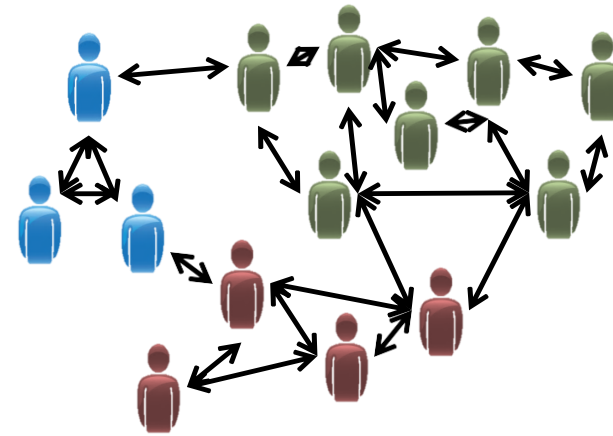
Social network analysis

Market segmentation

Astronomical data analysis

Slide credit: Andrew Ng

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

The Agent-Environment Interface
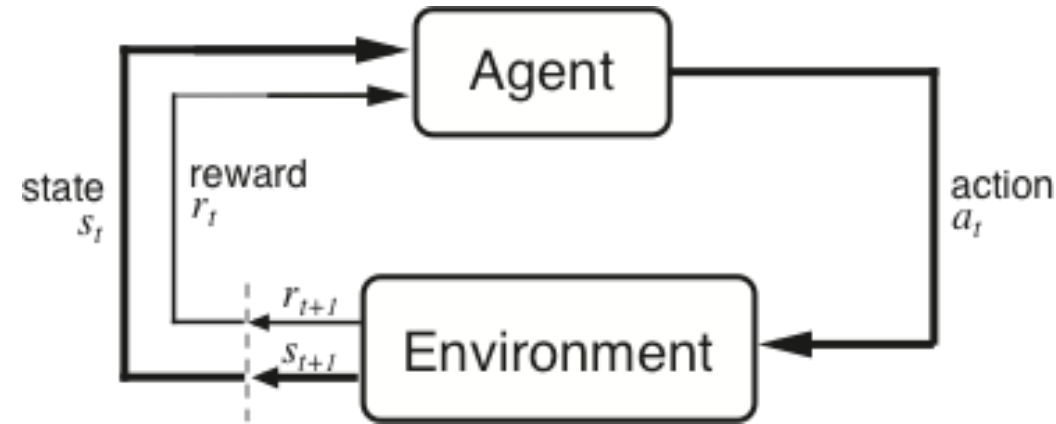
Agent and environment interact at discrete time steps : $t = 0, 1, 2, \mathrm{K}$

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}

Reinforcement Learning

https://www.youtube.com/watch?v=4cgWya-wjgY
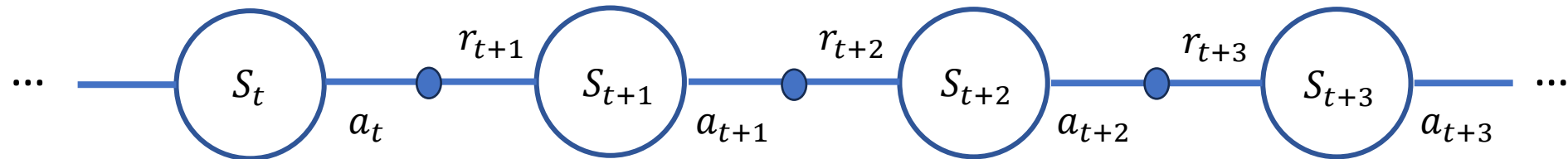
Inverse Reinforcement Learning

- Learn policy from user demonstrations

Stanford Autonomous Helicopter

http://heli.stanford.edu/

https://www.youtube.com/watch?v=VCdxqn0fcnE

Framing a Learning Problem

Designing a Learning System

- Choose the training experience
- Choose exactly what is to be learned
  - i.e. the **target function**

- Choose how to represent the target function
- Choose a learning algorithm to infer the target function from the experience
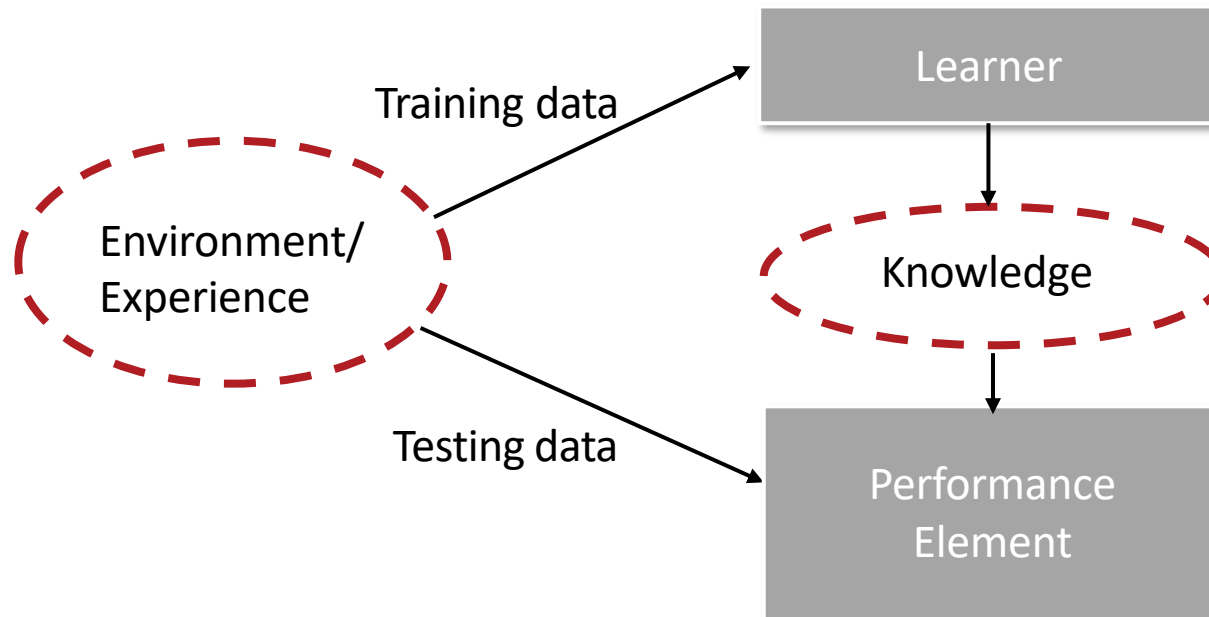
Training vs. Test Distribution

- We generally assume that the training and test examples are independently drawn from the same overall distribution of data
  - We call this "i.i.d" which stands for "independent and identically distributed"
- If examples are not independent, requires
  collective classification
- If test distribution is different, requires
  transfer learning

ML in a Nutshell

- Tens of thousands of machine learning algorithms

  - Hundreds new every year

- Every ML algorithm has three components:

  - **Representation**

  - **Optimization**

  - **Evaluation**

Various Function Representations

- Numerical functions
  - Linear regression
  - Neural networks
  - Support vector machines

- Symbolic functions
  - Decision trees
  - Rules in propositional logic
  - Rules in first-order predicate logic

- Instance-based functions
  - Nearest-neighbor
  - Case-based

- Probabilistic Graphical Models
  - Naïve Bayes
  - Bayesian networks
  - Hidden-Markov Models (HMMs)
  - Probabilistic Context Free Grammars (PCFGs)
  - Markov networks

Various Search/Optimization Algorithms

- Gradient descent
  - Perceptron
  - Backpropagation
- Dynamic Programming
  - HMM Learning
  - PCFG Learning
- Divide and Conquer
  - Decision tree induction
  - Rule learning
- Evolutionary Computation
  - Genetic Algorithms (GAs)
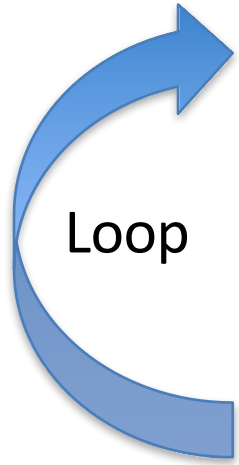 - Genetic Programming (GP)
  - Neuro-evolution

Evaluation

- Accuracy
- Precision and recall

- Squared error
- Likelihood
- Posterior probability

- Cost / Utility
- Margin

- Entropy
- K-L divergence
- etc.

ML in Practice

Loop

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

Lessons Learned about Learning

- Learning can be viewed as using direct or indirect experience to approximate a chosen target function.

- Function approximation can be viewed as a search through a space of hypotheses (representations of functions) for one that best fits a set of training data.

- Different learning methods assume different hypothesis spaces (representation languages) and/or employ different search techniques.

Any questions?