



Introduction to Machine Learning

*Knowing is not enough; we must apply.
Willing is not enough; we must do.*

Johann Wolfgang von Goethe

Acknowledgement: Eric Eaton, www.seas.upenn.edu/~cis519

What is Machine Learning?

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

- Herbert Simon

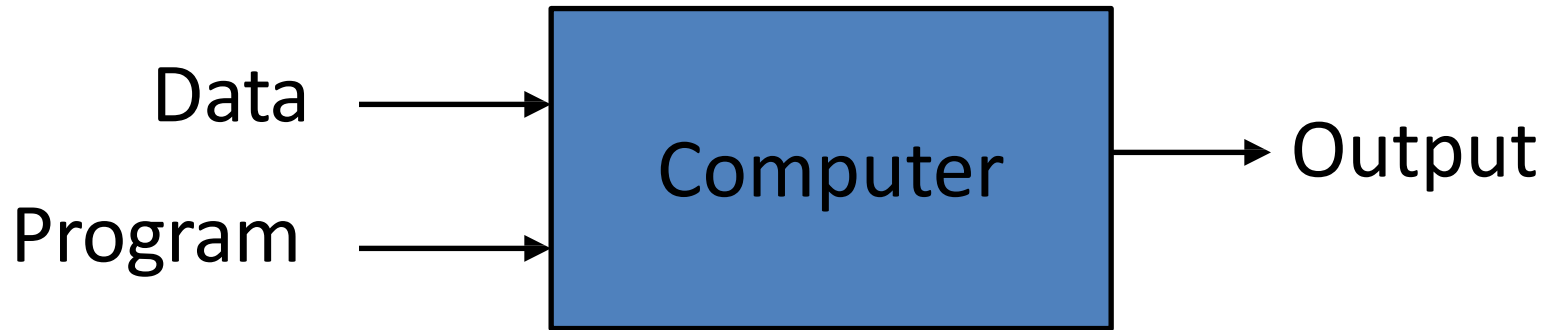
Definition by Tom Mitchell (1998):

Machine Learning is the study of algorithms that

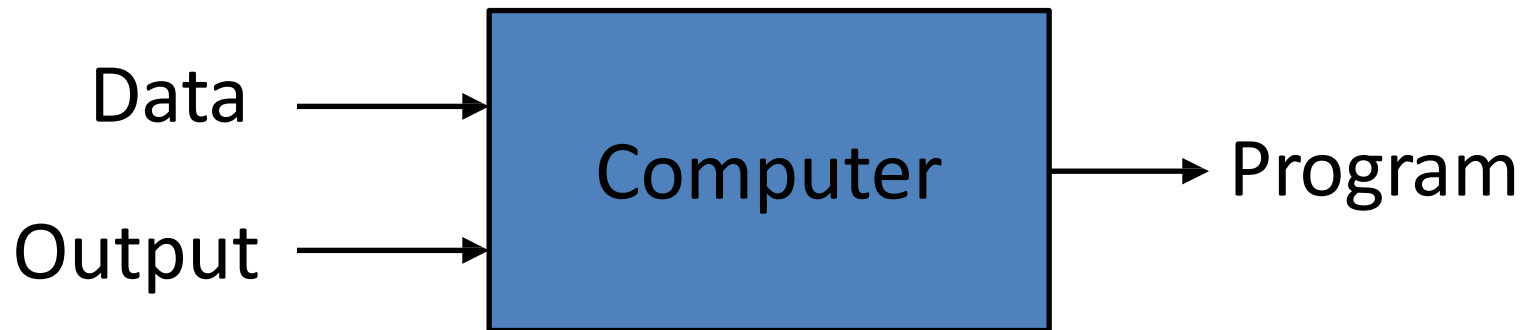
- improve their performance P
- at some task T
- with experience E .

A well-defined learning task is given by $\langle P, T, E \rangle$.

Traditional Programming



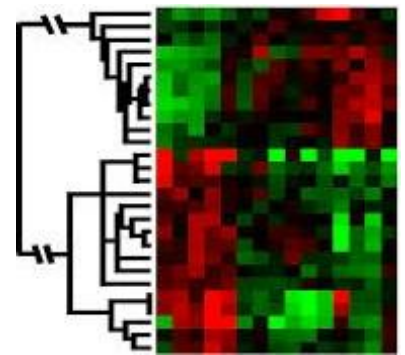
Machine Learning



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)



Learning isn't always useful:

- There is no need to “learn” to calculate payroll

A classic example of a task that requires machine learning:
It is very hard to say what makes a 2

0 0 0 1 1 1 1 1 1 2

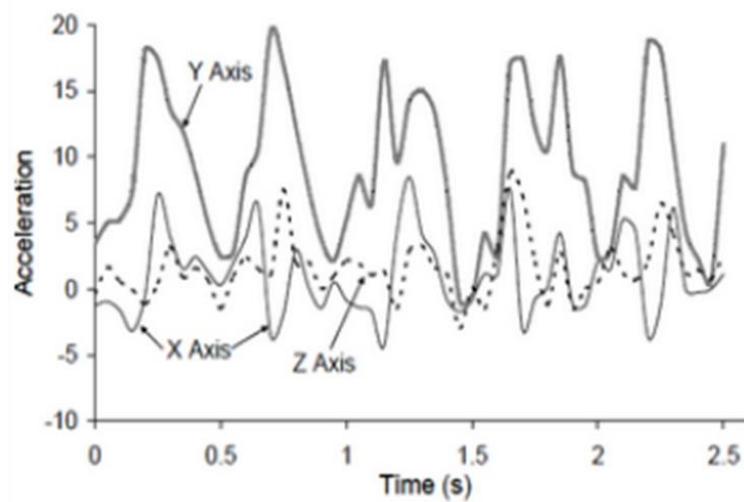
2 2 2 2 2 2 2 3 3 3

3 4 4 4 4 4 5 5 5 5

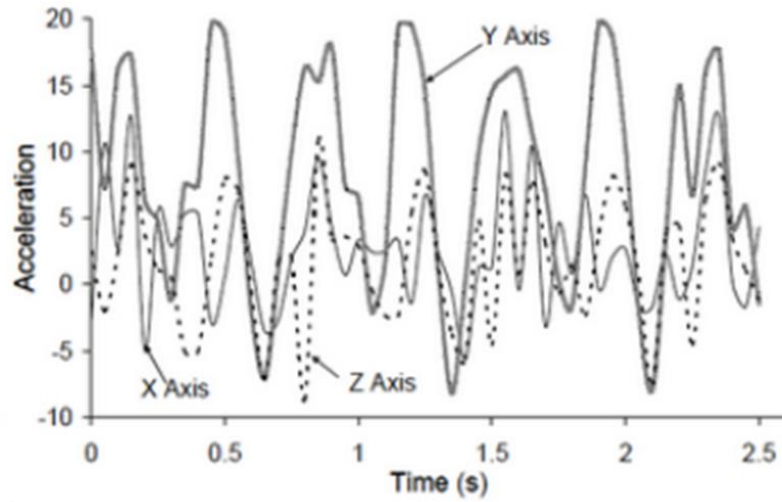
6 6 7 7 7 7 7 8 8 8

9 9 9 9 9 9 9 9 9

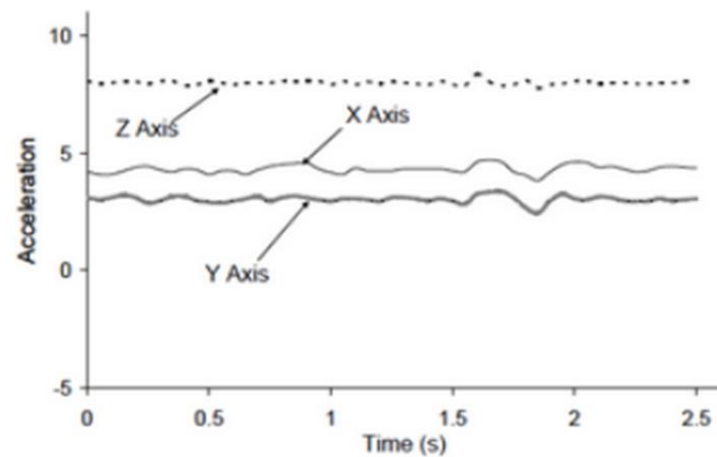
Another Example: Activity Recognition



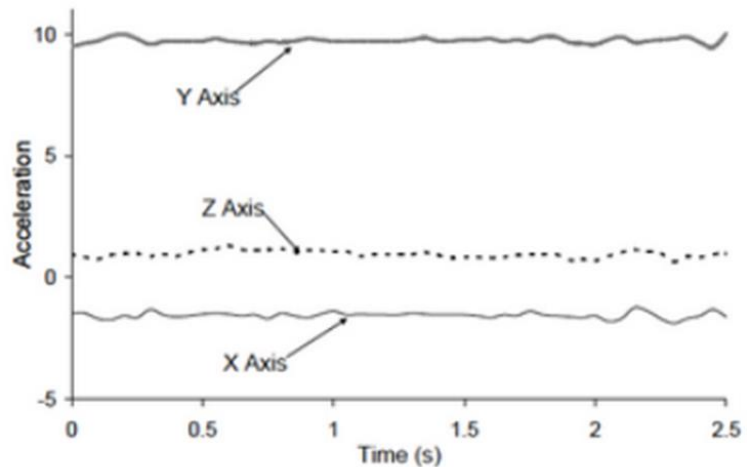
Waking



Jogging



Sitting



Standing

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

Sample Applications

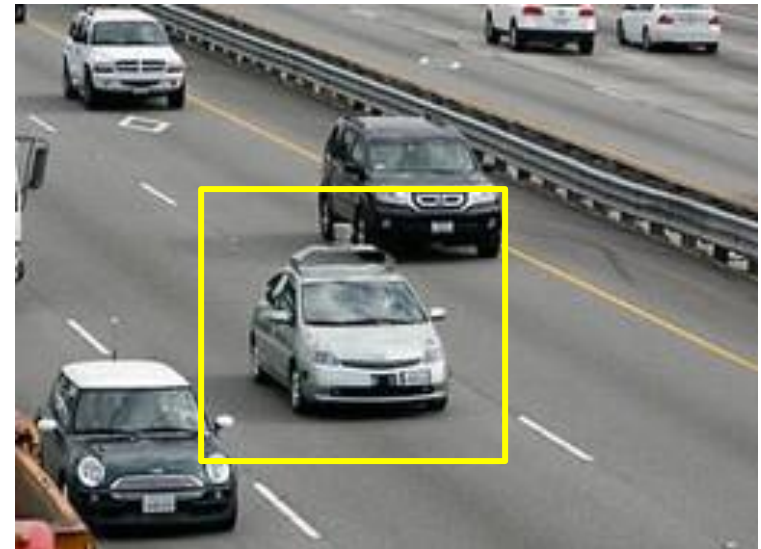
- Web search
- Computational biology
- Finance
- E-commerce
- Space exploration
- Robotics
- Information extraction
- Social networks
- Debugging software
- Many real-world problems that you are working on

Samuel's Checkers-Player

“Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.” -Arthur Samuel (1959)



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)



Speech Technology



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

A Specific Example: Activity Recognition

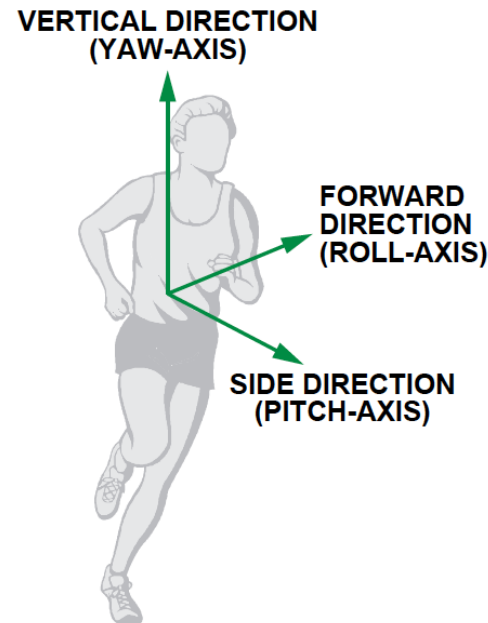
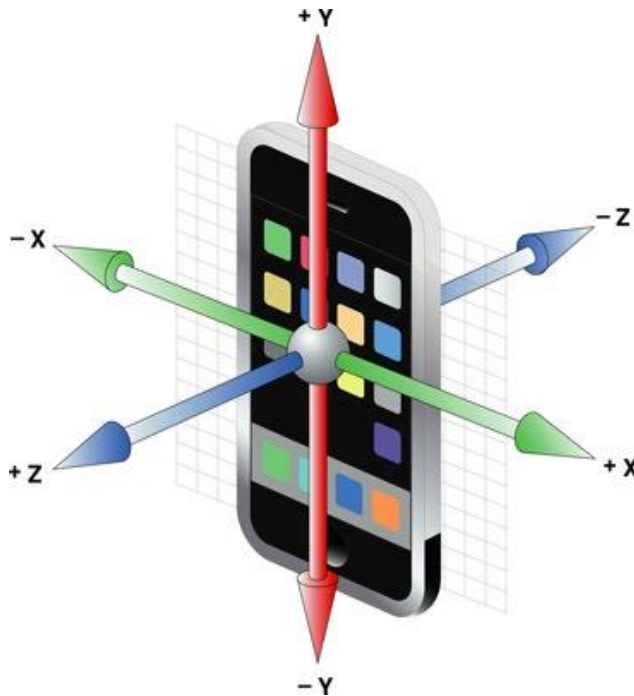
Example: Mobile Activity Tracker



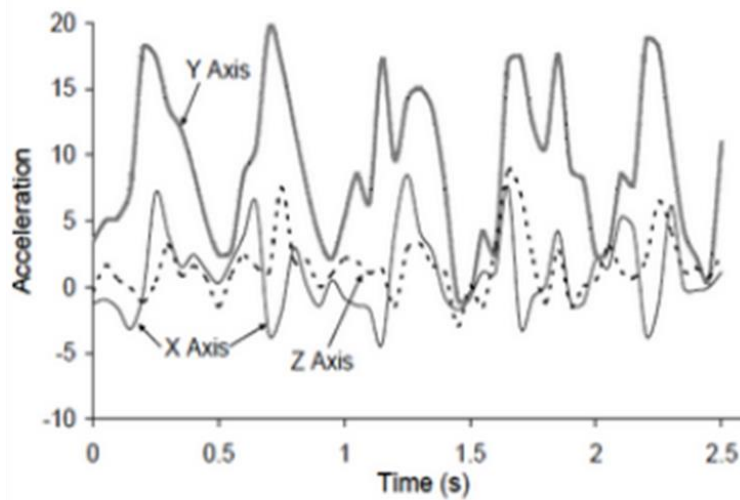
- Everyday exercise progress monitor and motivator
- Provide reliable feedback about how much they move. (People often overestimate!)
- Provide instant and constant feedback about activity levels.
- Gamify to encourage individuals to compete in getting fit and losing weight.

Inertial Sensors: Accelerometer

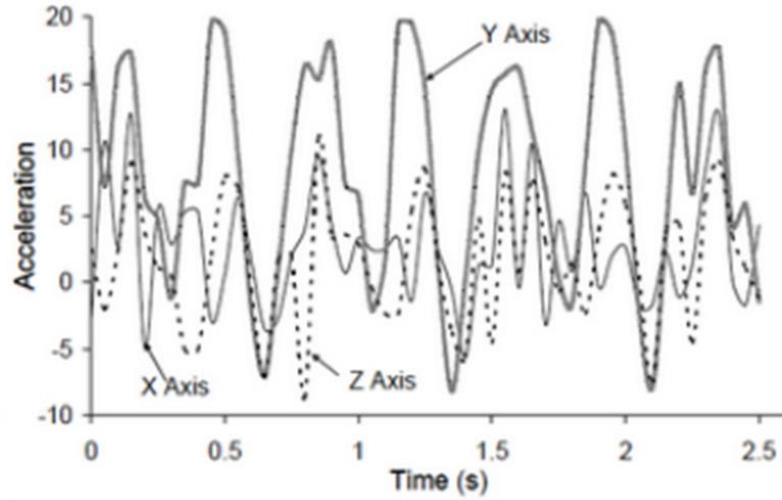
- All commodity smartphones have accelerometers.
- Measure linear acceleration (m/s^2) in three different directions.



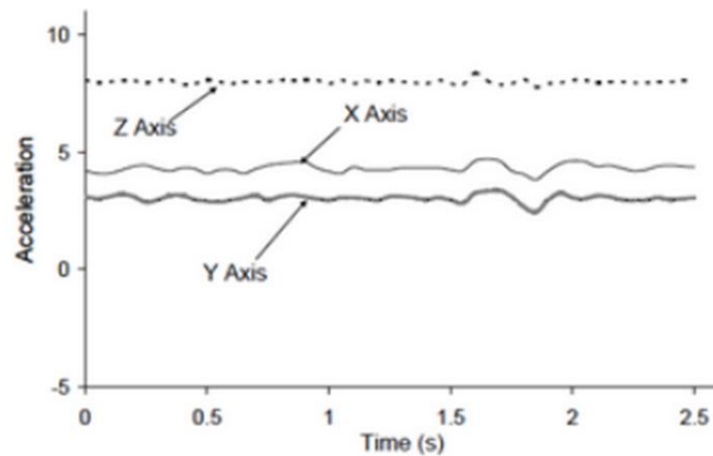
Signal Patterns



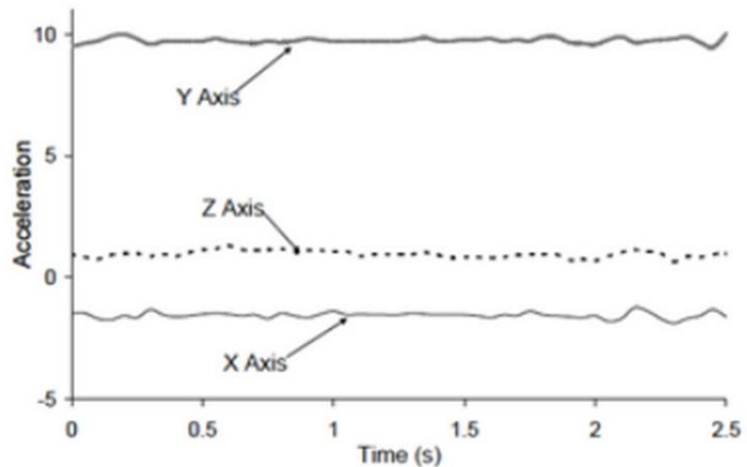
Waking



Jogging



Sitting



Standing

Solution 1: Heuristic

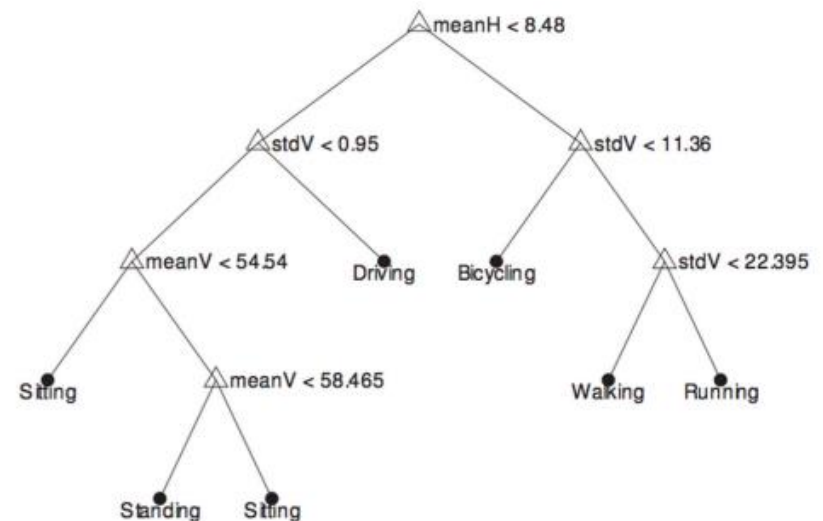
- If $\text{STDEV}(\text{y-axis samples}) < C_{\text{Threshold1}}$
 - ✓ If $\text{AVG}(\text{y-axis samples}) > C_{\text{Threshold2}}$
 - output standing
 - ✓ Else
 - output sitting
- Else
 - ✓ If $\text{FFT}(\text{y-axis samples}) < C_{\text{Threshold3}}$
 - output walking
 - ✓ Else
 - output jogging

Are We Good?

- How do we determine good features and good thresholds?
 - ✓ How do we know STDEV is better than MAX?
 - ✓ How do we know AVG is better than Median?
 - ✓ How do we know the right values for $C_{\text{threshold}}$?
- What if a user puts her phone in her bag, not in her front pocket?
 - ✓ The Y-axis of the phone is not anymore the major axis of movement.
- How do we solve these problems? A better heuristic?

Solution 2: Decision Tree

- A simple but effective ML classifier.
- This tree can be built by the C4.5 algorithm.
- Given sufficient training data, the algorithm can automatically determine the important features and their thresholds.



How to Build a Decision Tree?

- Pseudocode

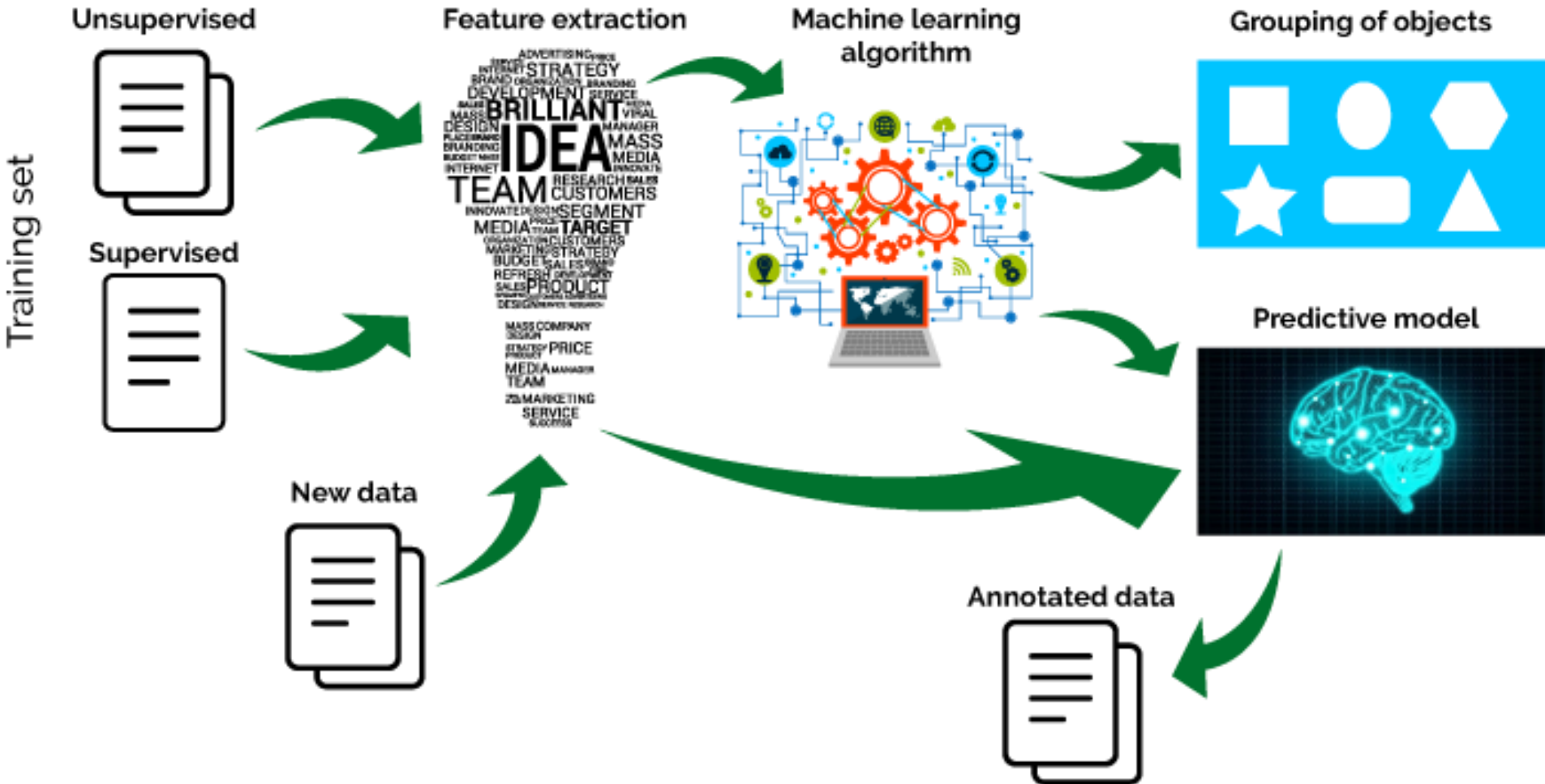
1. For each feature f , find the normalized information gain (a metric to effectively split data into classes) from splitting on f
2. Let f_best be the attribute with the highest normalized information gain
3. Create a decision node that splits on f_best
4. Recurse on the sublists obtained by splitting on f_best , and add those nodes as children of node

- More to be covered in Section 7 (Tree-based Approach) of our textbook.

Other ML Techniques

- Naïve Bayes classifier
- Decision tree
- Random forest
- Support vector machine
- Linear regression
- Hidden Markov model
- Gaussian mixture model
-

ML Techniques Flow



ML Techniques: Limitations

- Expert knowledge required for feature extraction
- Not easy to improve accuracy after a certain point (even with a large volume of data)
- Not easy to model non-linear relations between an input and output.

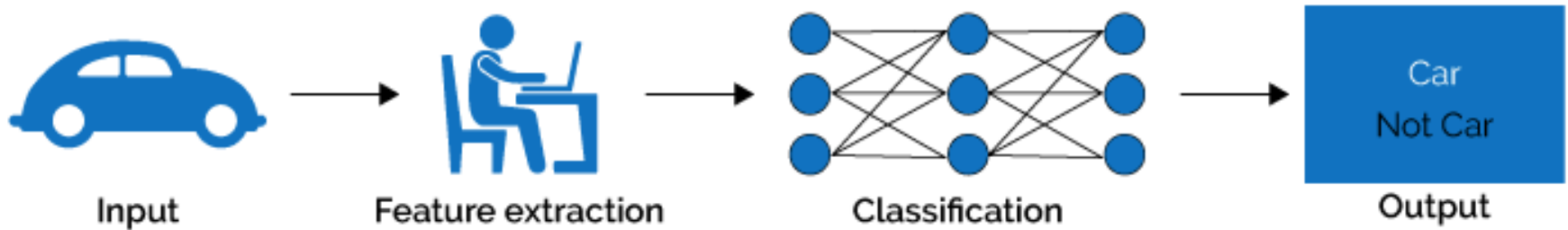
ML Techniques: Limitations

- Linear regression?
 - ✓ Why is it **linear**?
- Bayesian?
 - ✓ What is the **prior**?
- SVM?
 - ✓ What are the **features**?
- Decision tree?
 - ✓ What are the **nodes/variables**?
- KNN?
 - ✓ Cluster on what **features**?

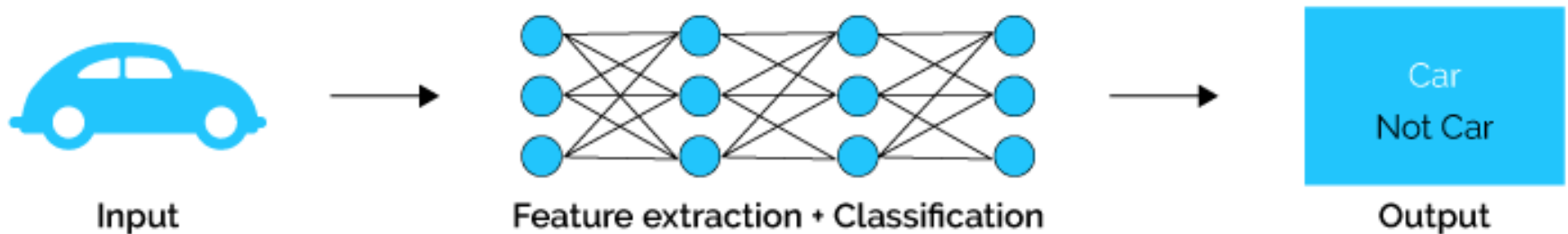
These methods do not suit well with very complex models.

Deep Learning

Machine Learning

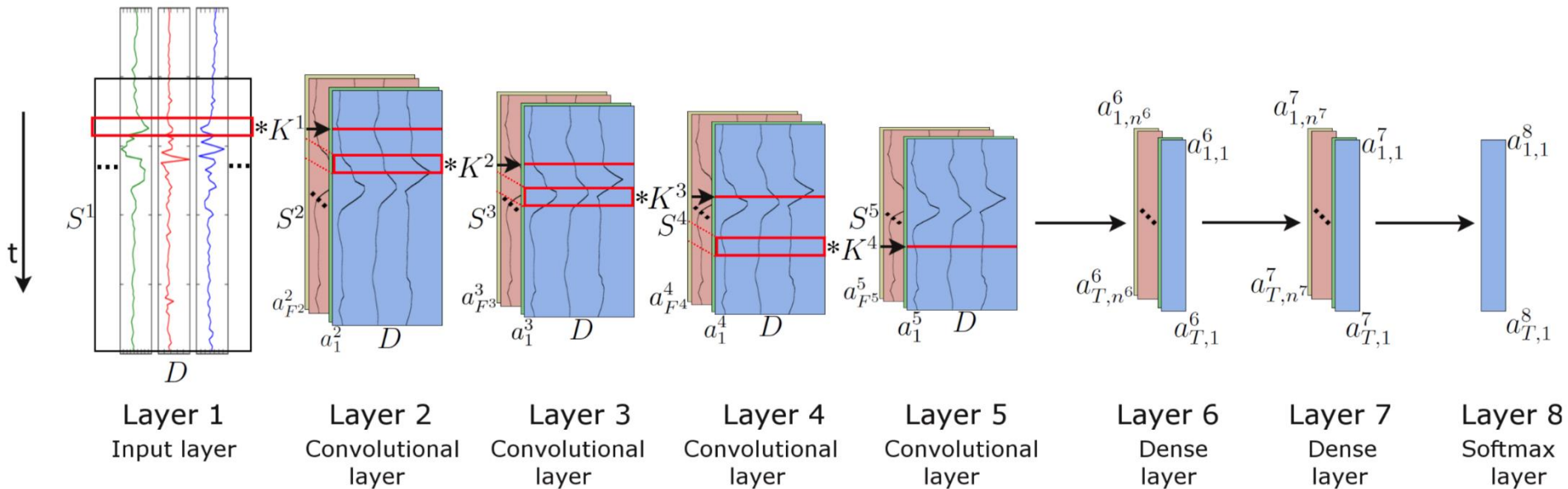


Deep Learning



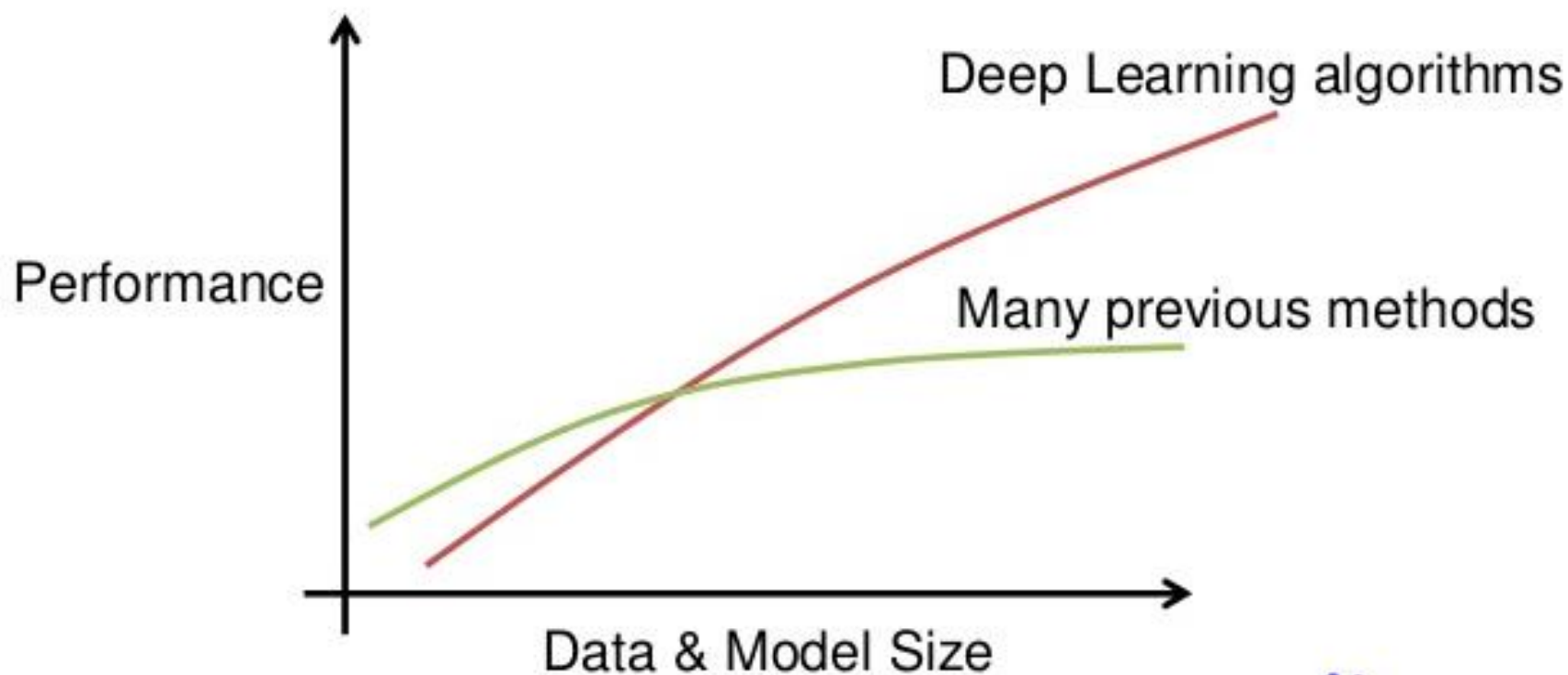
Deep Learning for Activity Recognition

- Example of applying a convolutional neural network



Machine Learning vs. Deep Learning

- Deep learning: the more data, the higher accuracy



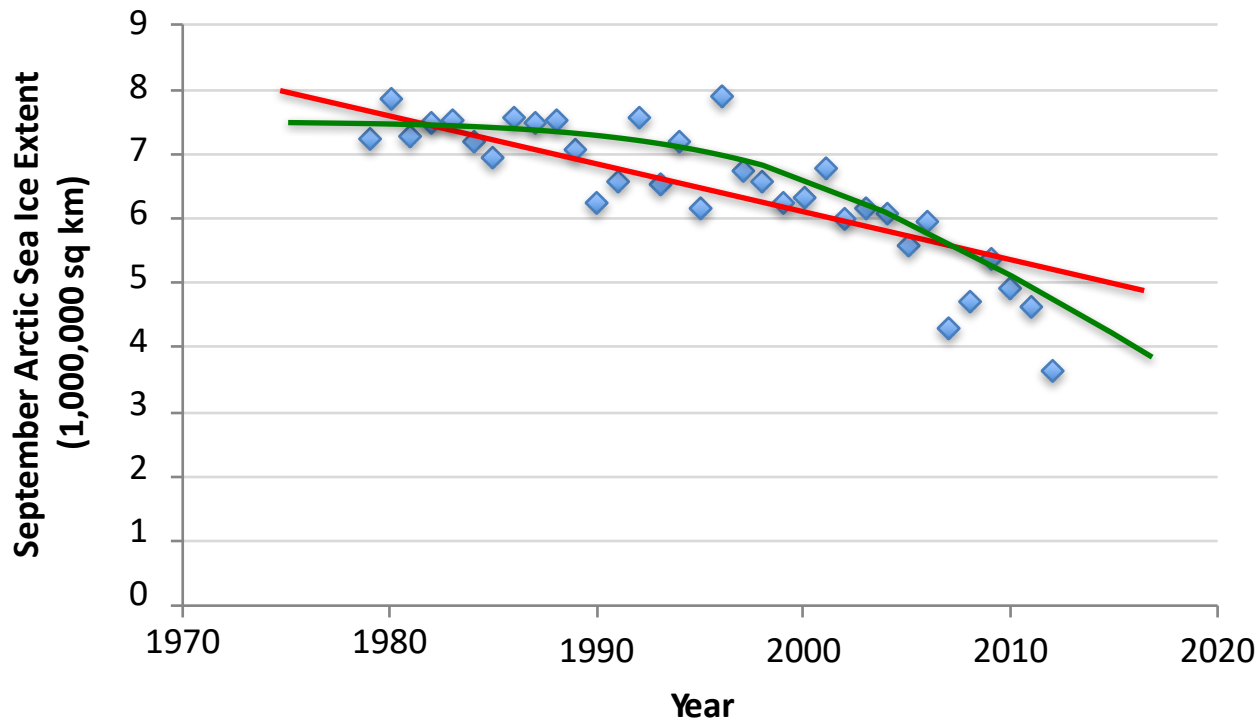
Types of Learning

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

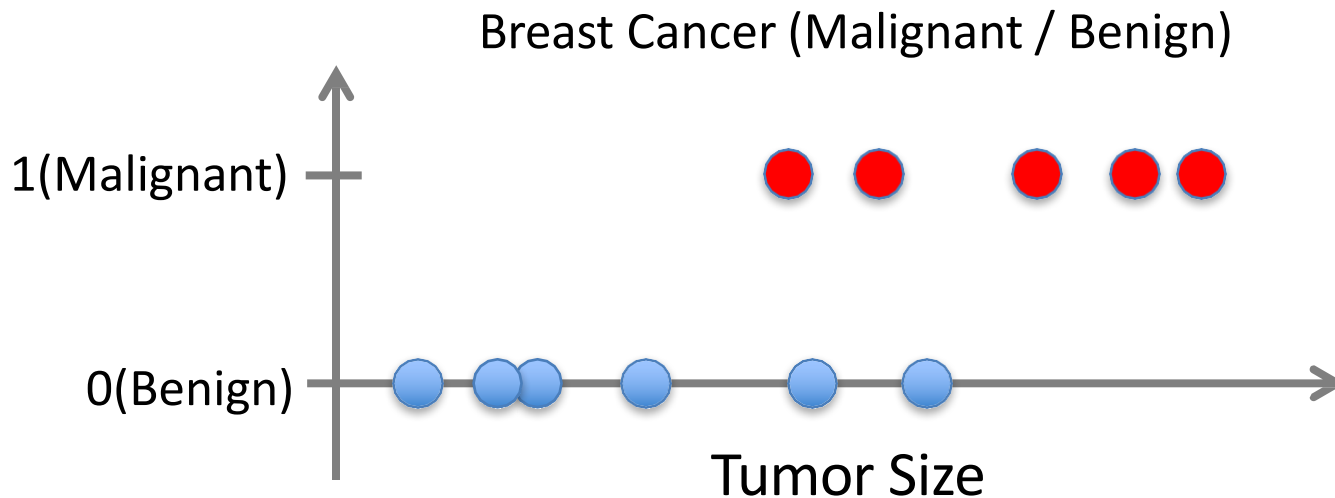
Supervised Learning: Regression

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



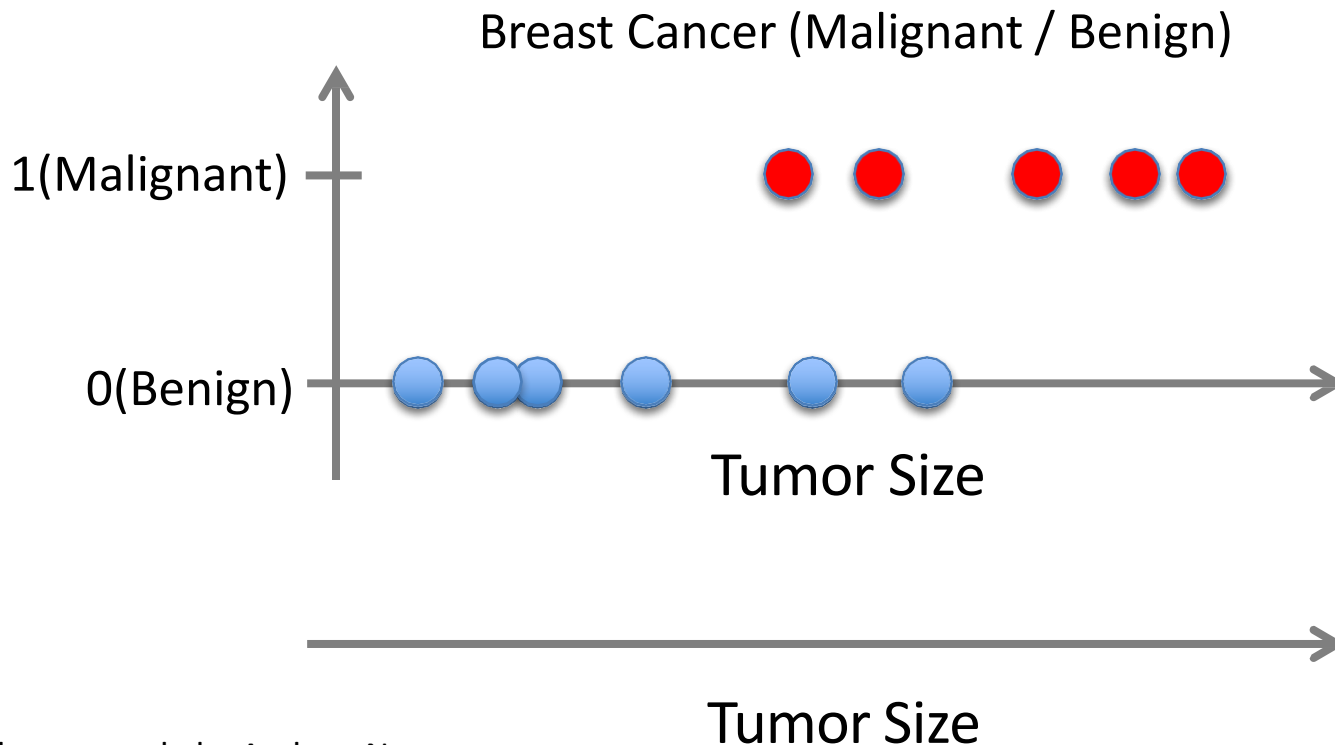
Supervised Learning: Classification

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



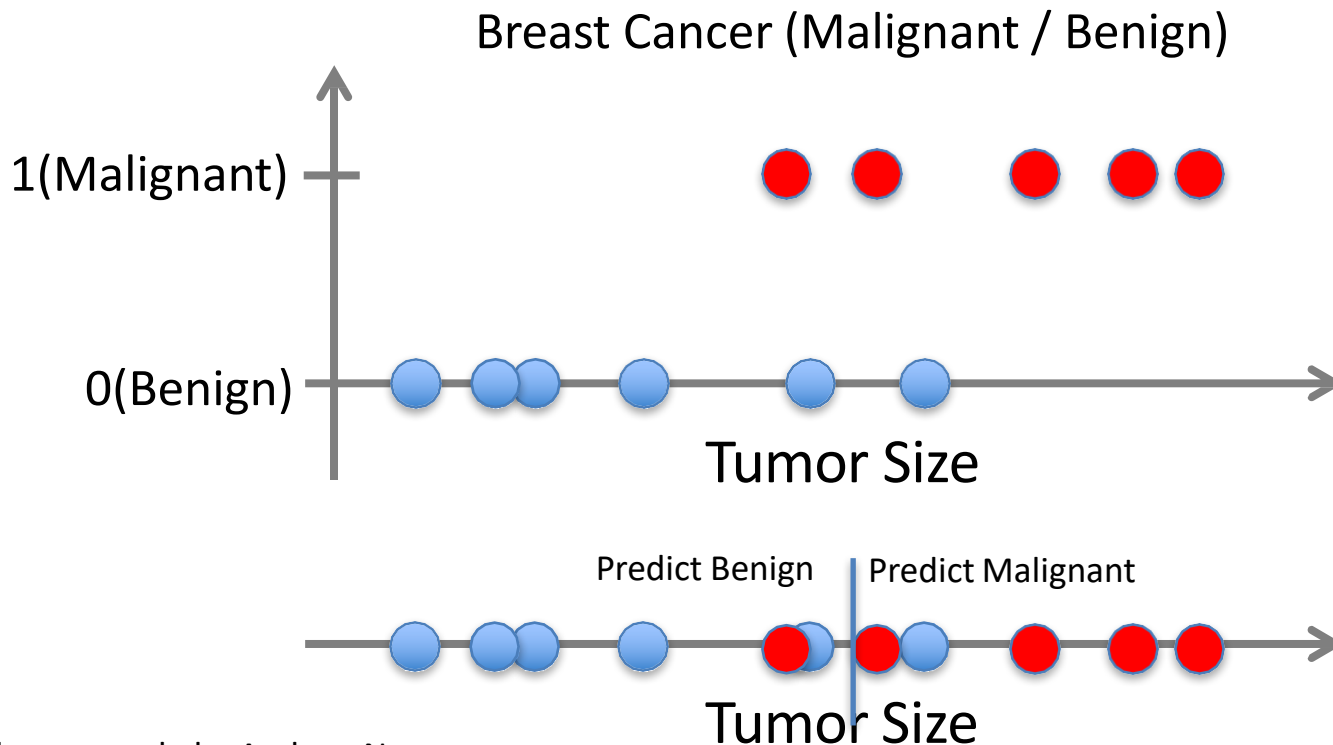
Supervised Learning: Classification

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



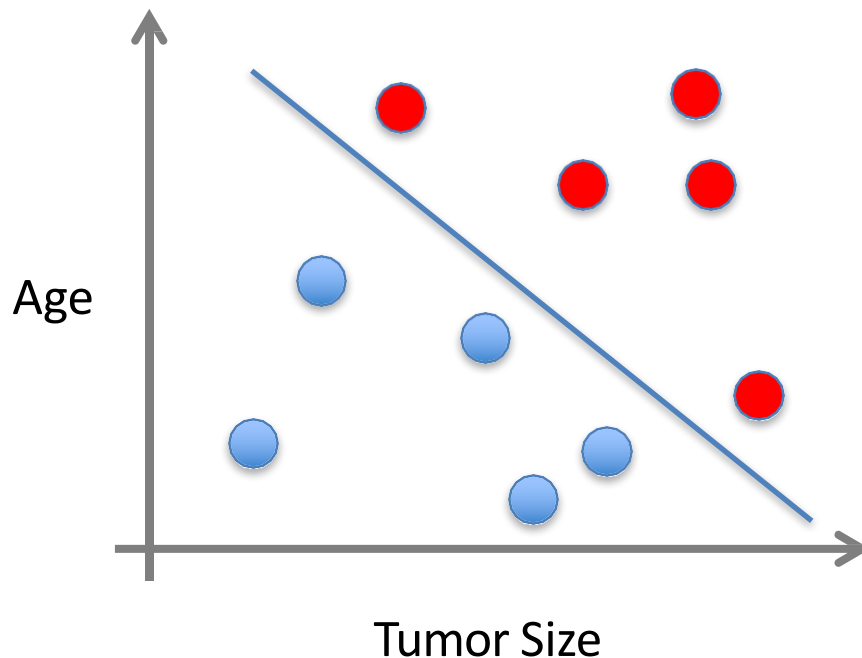
Supervised Learning: Classification

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



Supervised Learning

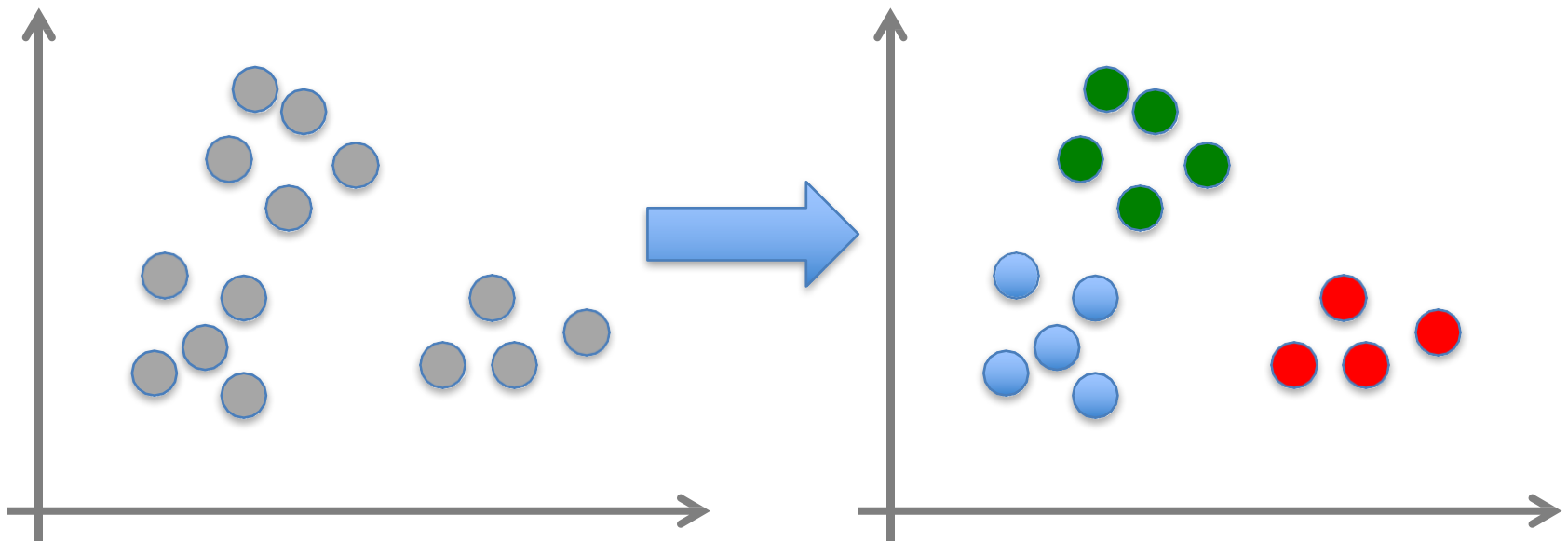
- x can be multi-dimensional
 - Each dimension corresponds to an attribute



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

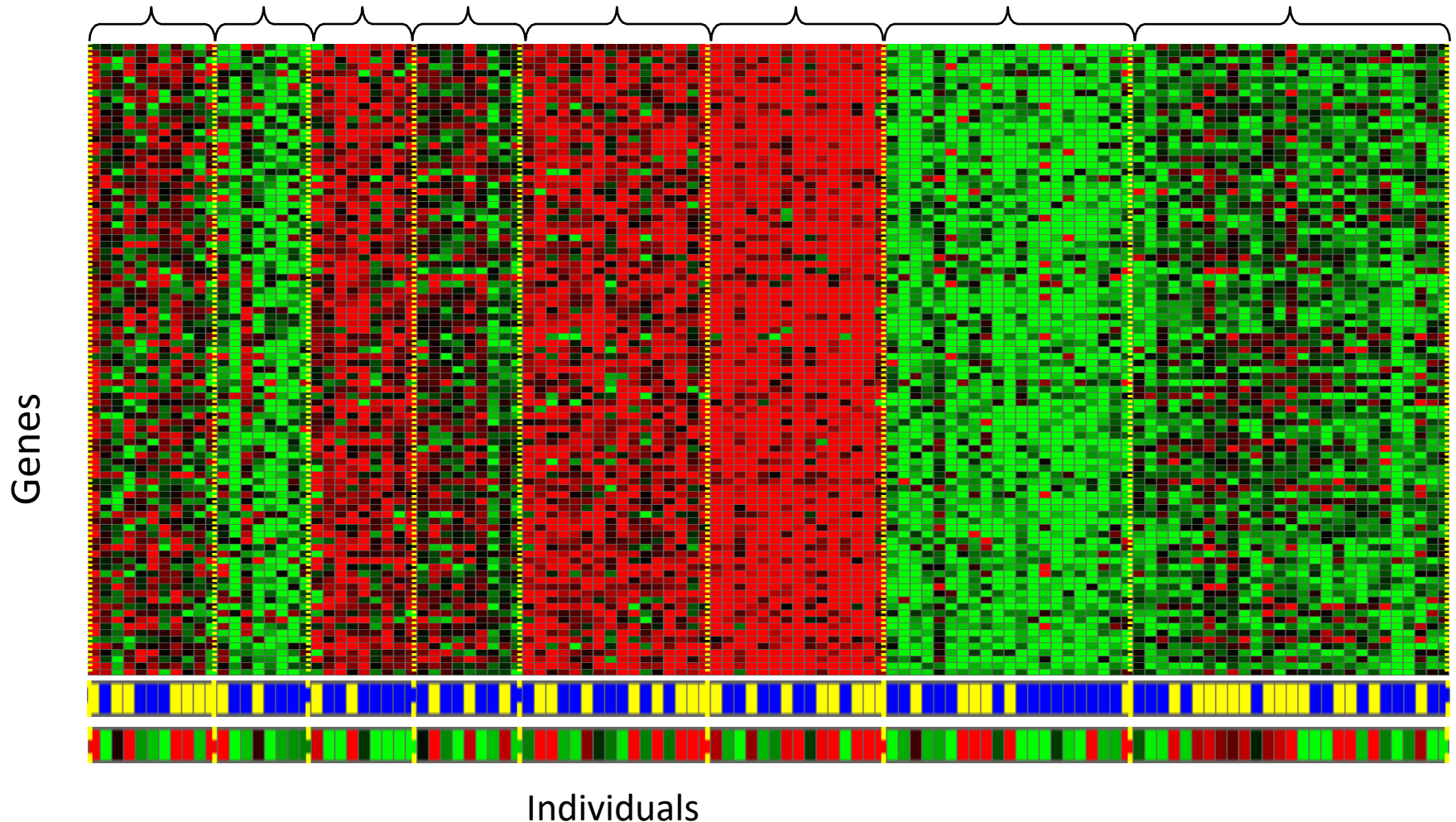
Unsupervised Learning

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



Unsupervised Learning

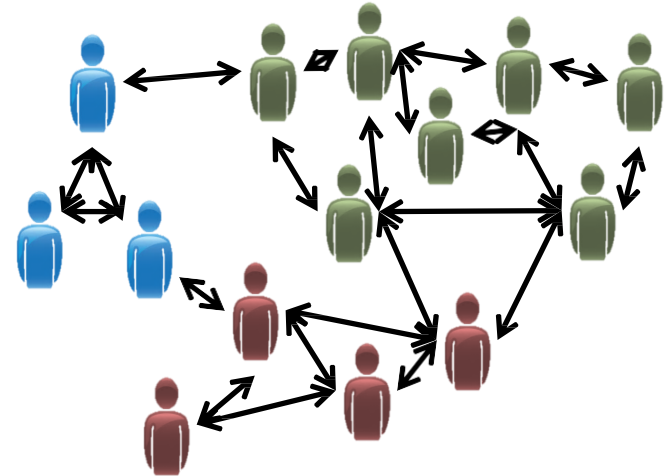
Genomics application: group individuals by genetic similarity



Unsupervised Learning



Organize computing clusters



Social network analysis



Market segmentation

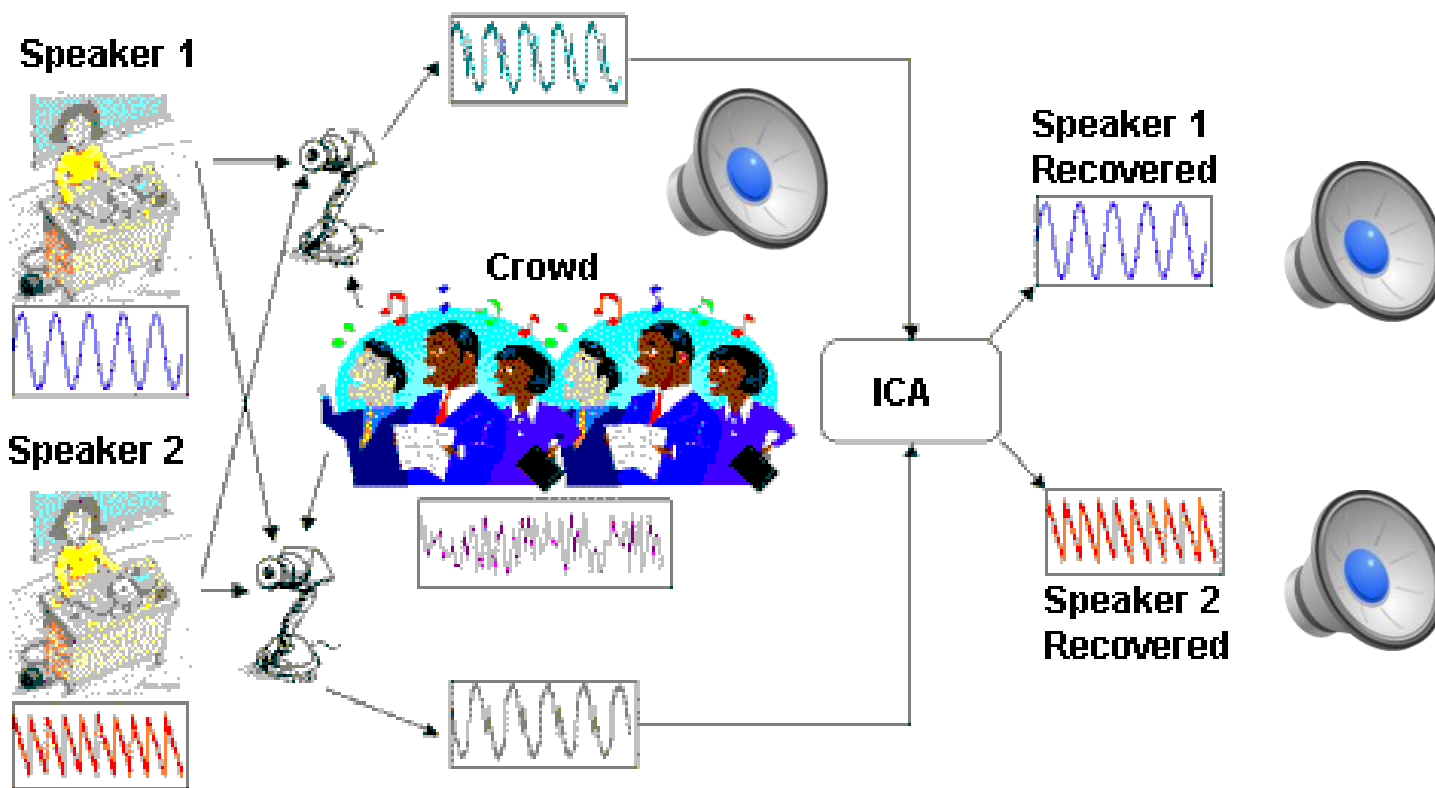


Image credit: NASA/JPL-Caltech/E. Churchwell (Univ. of Wisconsin, Madison)

Astronomical data analysis

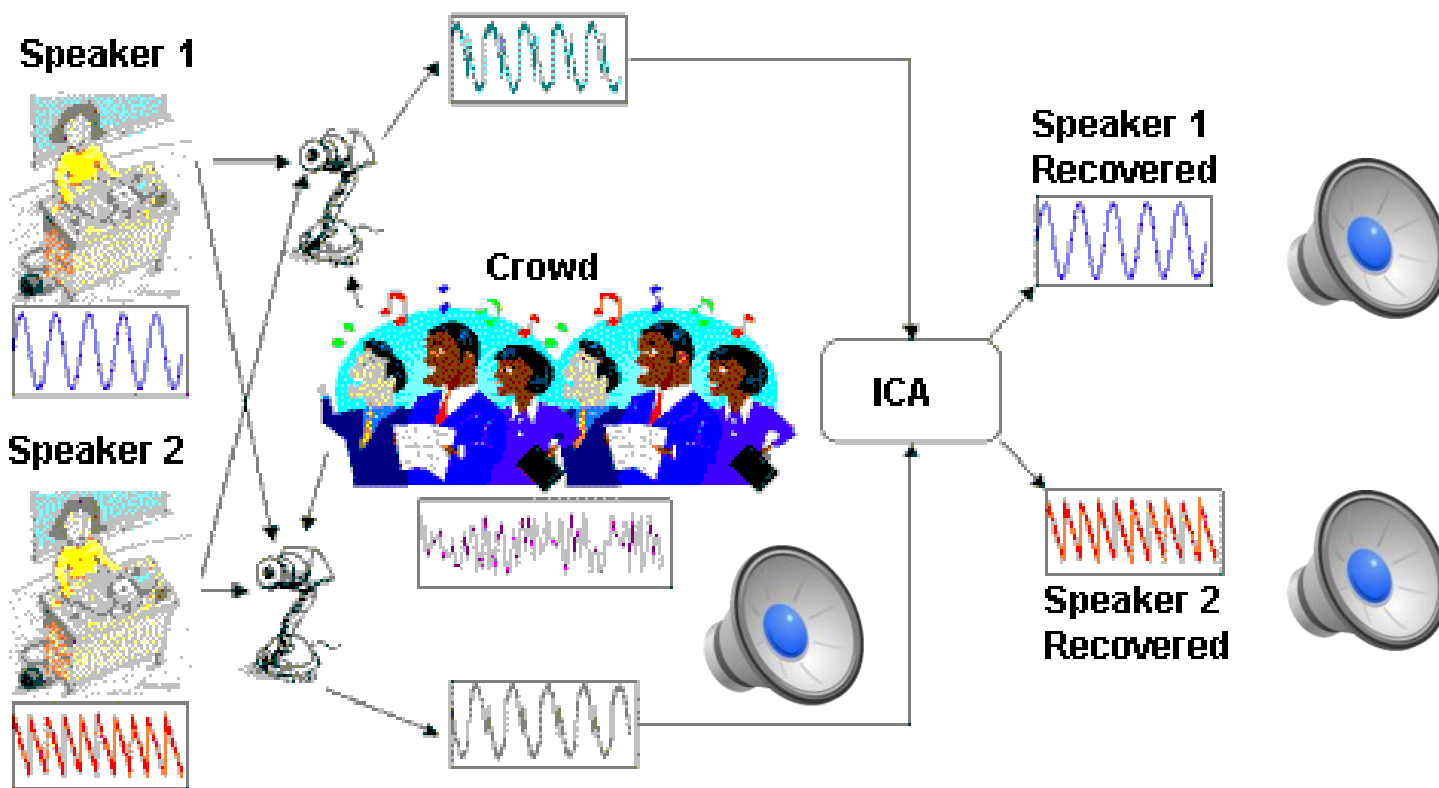
Unsupervised Learning

- Independent component analysis – separate a combined signal into its original sources



Unsupervised Learning

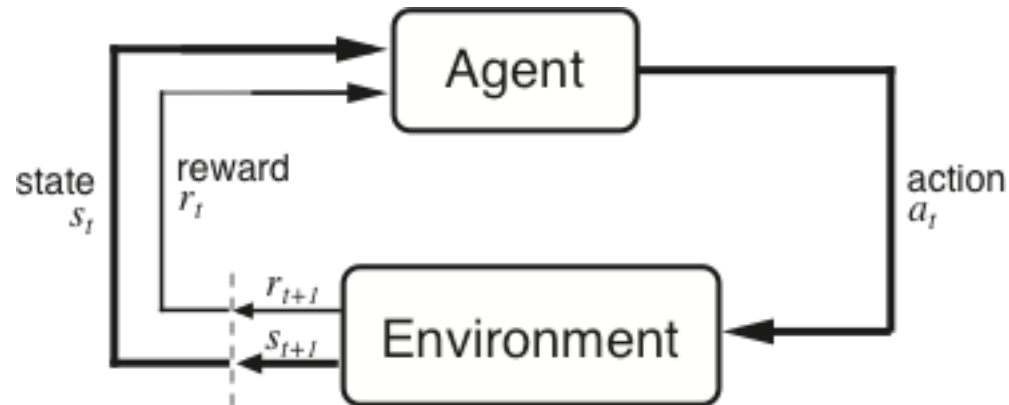
- Independent component analysis – separate a combined signal into its original sources



Reinforcement Learning

- Given a sequence of states and actions with (delayed) rewards, output a policy
 - Policy is a mapping from states \rightarrow actions that tell 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



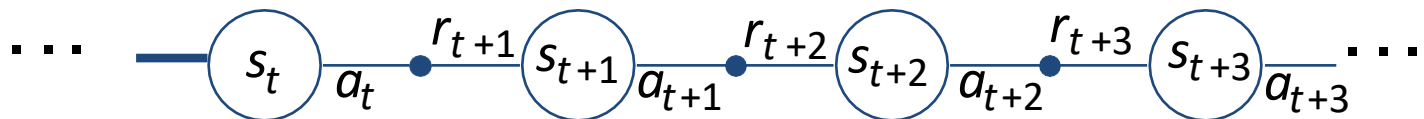
Agent and environment interact at discrete time steps : $t = 0, 1, 2, \mathbb{K}$

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

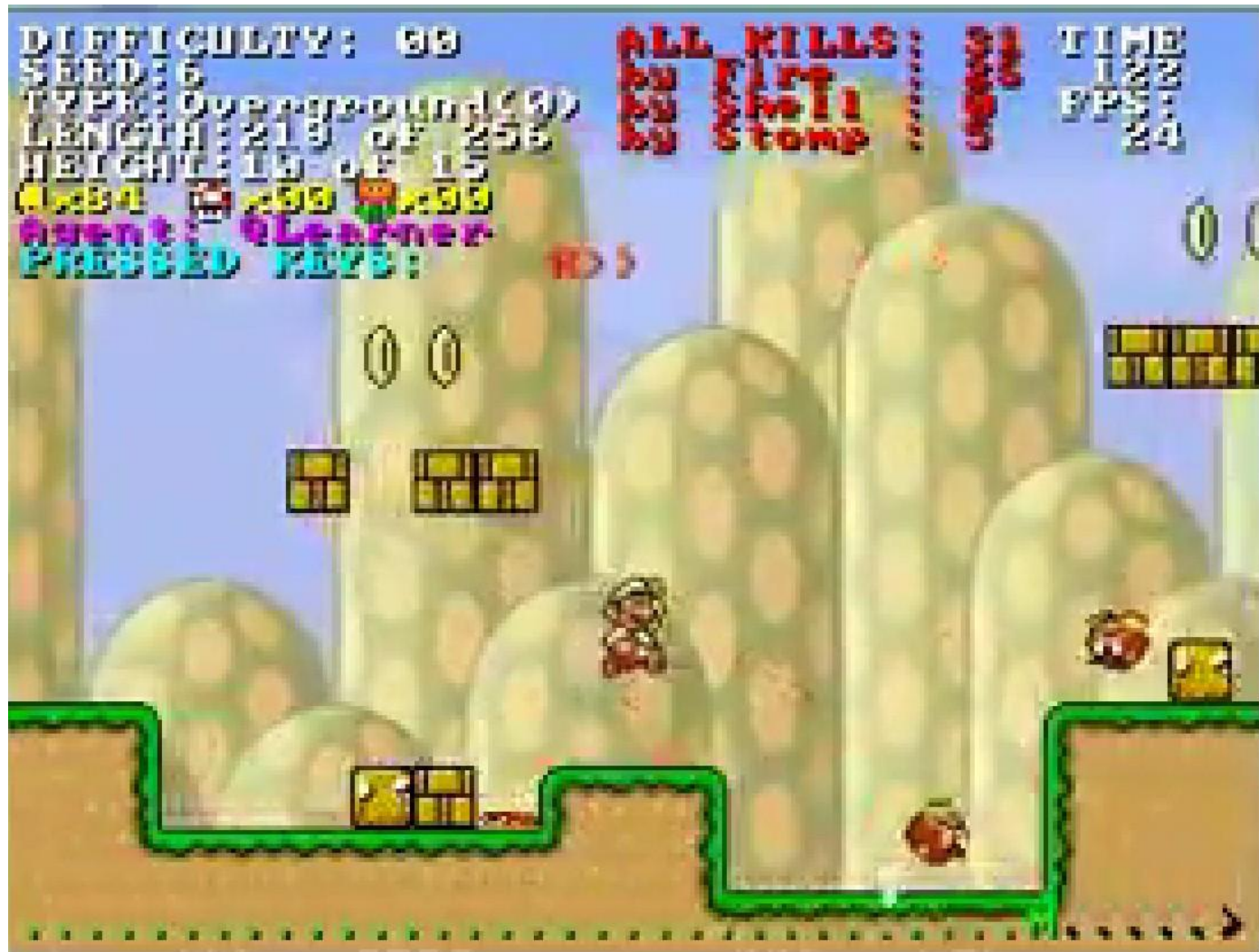
produces action at step t : $a_t \in A(s_t)$

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

and resulting next state : s_{t+1}



Reinforcement Learning



<https://www.youtube.com/watch?v=4cgWya-wjgY>

What We'll Cover in this Course

- **Supervised learning**
 - Linear regression
 - Logistic regression
 - Discriminant Analysis
 - Tree-based Methods
 - Support Vector Machines
- **Unsupervised learning**
 - Clustering
 - Dimensionality reduction
- **Evaluation**
- **Applications**

Our focus is = learning various ML techniques and their applications