An Introduction on Machine Learning and Its Applications in Networking

Lecturers: Sharare Zehtabian & Siavash Khodadadeh

Spring 2020

Outline

- Machine learning
- Machine learning applications
- Types of machine learning
  - Supervised learning setup
    - classification
    - regression
  - Unsupervised learning
  - Reinforcement learning
- Neural networks and deep learning

Machine learning definition

Definition by Tom Mitchell (1998): Machine Learning is the study of algorithms that

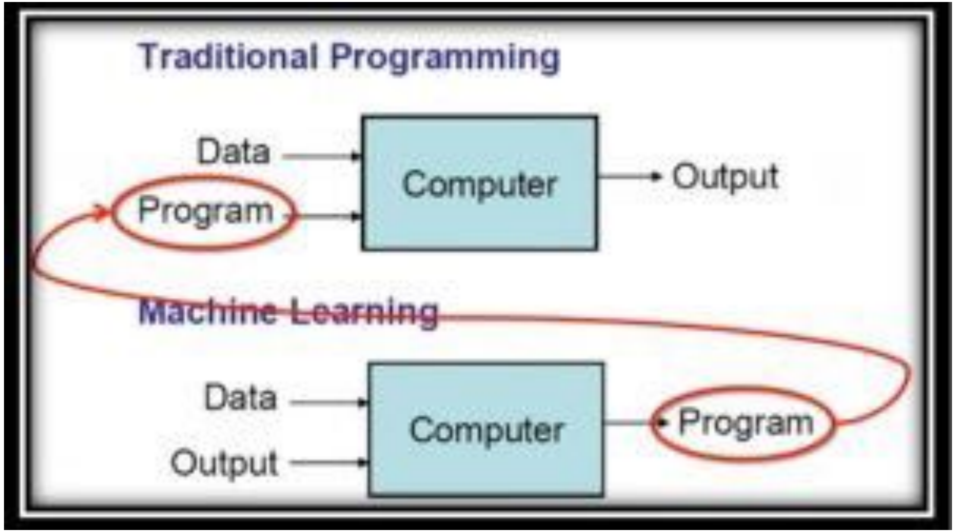
- Improve their performance P
- At some tasks T
- With experience E

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

Why machine learning?

For many problems, it’s difficult to program the correct behavior by hand

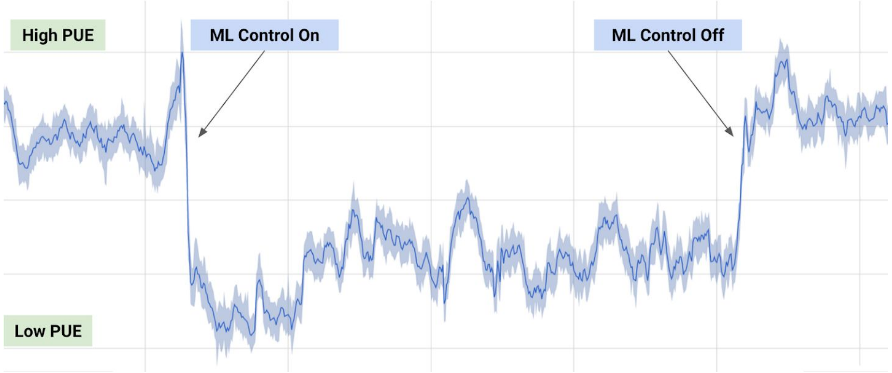
- Recognizing people and objects
- Understanding human speech



Source: https://mc.ai/machine-learning-1100101b-lets-learn-about-learning/

Applications in networking

- Pattern recognition:
  - Identifying patterns in networks traffic (e.g during a day or a week)
- Anomaly detection:
  - Using AI to detect anomalies in the way applications are being accessed (e.g. outlier detection at Netflix using a clustering algorithm)
- Network optimization
  - DeepMind AI reduced Google data centre cooling bill by 40% (PUE: Power Usage Effectiveness)



Applications in networking (cont'd)

- Forwarding path simplification
  - Could ML find a better way CRUD (Create/Read/Update/Delete) operations in networking?
- Coordinating ML across edge and cloud
  - Predictive caching
- Intent based networking: Intelligent automation and assurance
  - Let's watch this video:

Types of machine learning

- **Supervised learning:** have labeled examples of the correct behavior
- **Unsupervised learning:** no labeled examples – instead, looking for interesting patterns in the data
- **Reinforcement learning:** learning system receives a reward signal, tries to learn to maximize the reward signal

Supervised learning setup

- We have a bunch of (x,y), where $x \in R^d$ is the input instance and y is label
- Training dataset $D=\{(x_1,y_1),\ldots,(x_n,y_n)\} \subseteq R^d \times C$

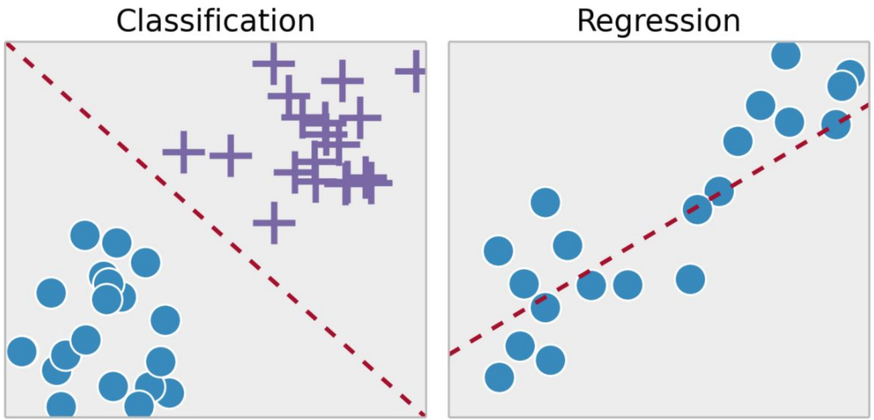
- R^d is the d-dimensional feature space
- x_i is the input vector of the ith sample
- y_i is the label of the ith sample
- C is the label space
- The training data pairs are drawn from some unknown distribution P

Supervised learning setup (cont’d)

- Try to predict properties of unseen data
  - Given a new sample, can we predict its properties?
- Learning problem:
  - Learn function h such that
  - for a new pair $(x, y) \sim P$, we have $h(x) \approx y$
- Example:
  - You are given the data of 900 passengers on Titanic. ($n = 900$)
  - For each passenger, we know some information like name, age, ticket number, cabin, etc
  - We want to learn from this data if there is a correlation between these features (x) and whether the passenger survived the disaster (labels)
  - Now we are given a new passenger’s data and we want to predict whether he/she survives
  - Label space? {survived, not survived}

Classification vs regression

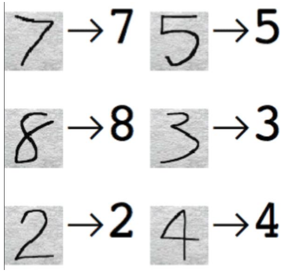
- What can be our labels?
  - Classification (discrete value)
    - Binary classification (e.g. spam or not spam)
    - Multi-class classification (e.g. dog or cat or horse or ..)
  - Regression (continuous value e.g. price of a house)



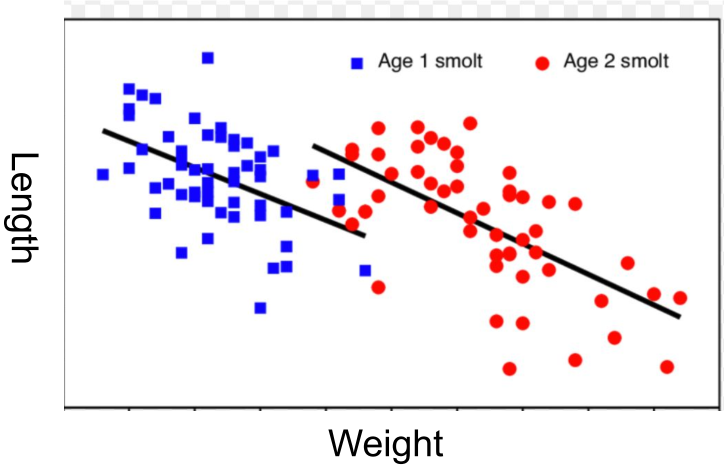
ref:https://scorecardstreet.wordpress.com/2015/12/09/is-machine-learning-the-new-epm-black/

Classification vs regression (examples)

- Classification example:
  - Handwritten digit recognition



- Regression example:
  - Prediction of the length of a salmon as a function of its age and weight.



Other Classification Tasks

Classification: given inputs x, predict labels (classes) y

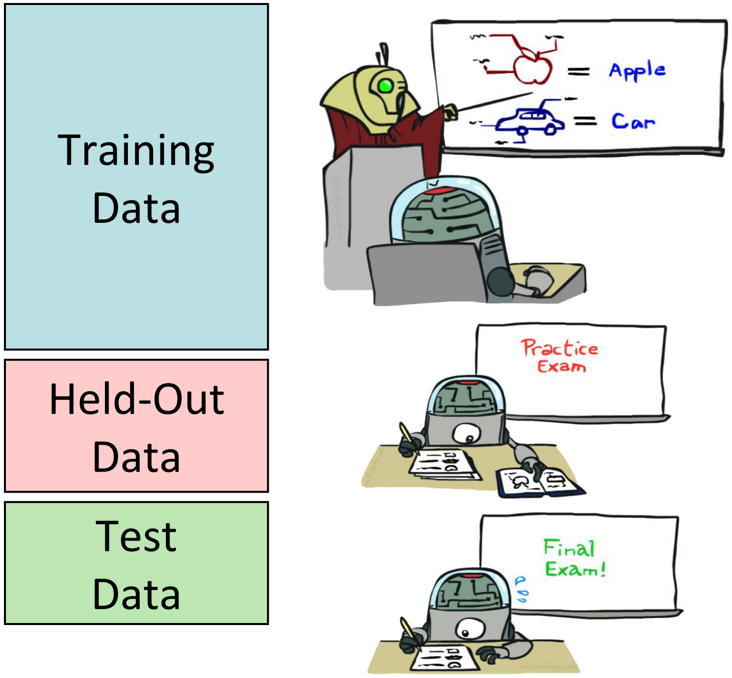
Examples:

- Spam detection (input: document,
  classes: spam / ham)
- OCR (input: images, classes: characters)
- Medical diagnosis (input: symptoms,
  classes: diseases)
- Automatic essay grading (input: document,
  classes: grades)
- Fraud detection (input: account activity,
  classes: fraud / no fraud)
- Customer service email routing
- … many more

Classification is an important commercial technology!

Training held-out and test data

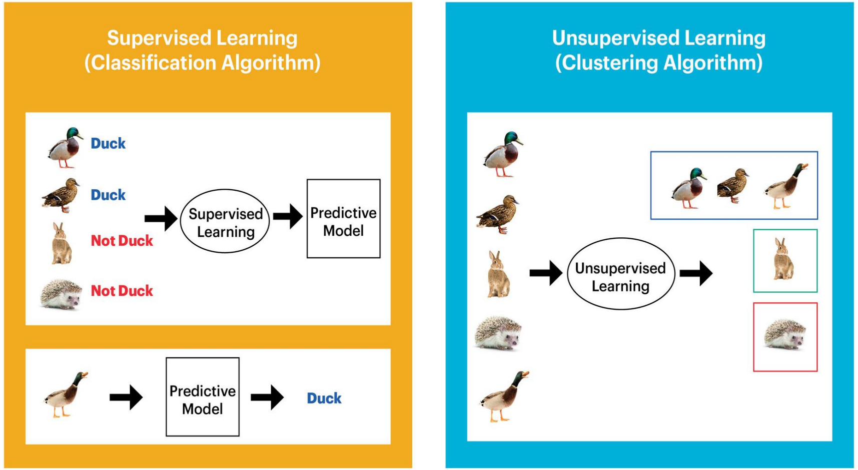
- How can we **evaluate** our machine learning algorithm?
    - Machine learning is about learning some properties of a data set (train) and then testing those properties against another data set (test).
    - The test data set is used **only** for evaluation and you should not use it except for that. (Do not use this data set for making any decision about the model).



[This image is adapted from the ones created by Dan Klein and Pieter Abbeel for CS188 Intro to AI at UC Berkeley, available at http://ai.berkeley.edu.]

Supervised learning vs unsupervised learning

- **Supervised learning:** in which the data comes with additional attributes that we want to predict.
  - classification
  - regression
- **Unsupervised learning:** in which the training data consists of a set of input vectors x without any corresponding target values.
  - clustering
  - density estimation



Western Digital.

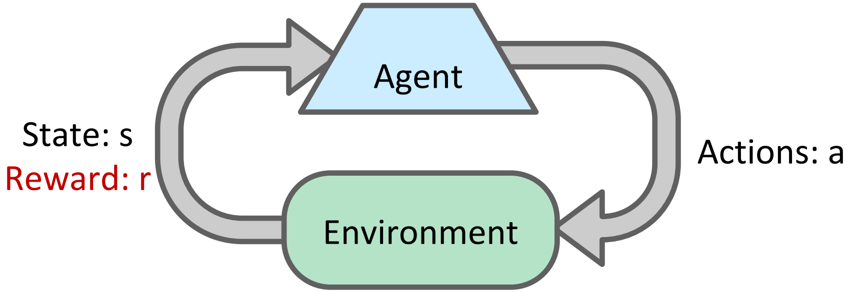
More about unsupervised learning

- Why unsupervised learning is important?
  - Labeling data costs time and resources
    - 300 hours of video are uploaded to youtube every minute
- What are different approaches for it?
  - Auto encoders
    - Encode input to a latent space and reconstruct it from there
  - Generative models
    - Two agents (neural networks) play a min-max game against each other

Reinforcement learning

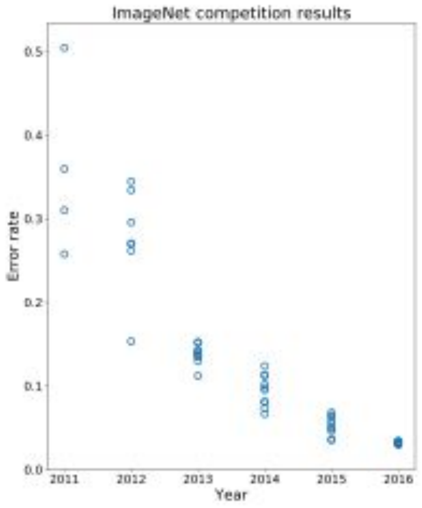
Basic idea:

- Receive feedback in the form of rewards
- Agent's utility is defined by the reward function
- Must (learn to) act so as to maximize expected rewards
- All learning is based on observed samples of outcomes!



Artificial neural networks

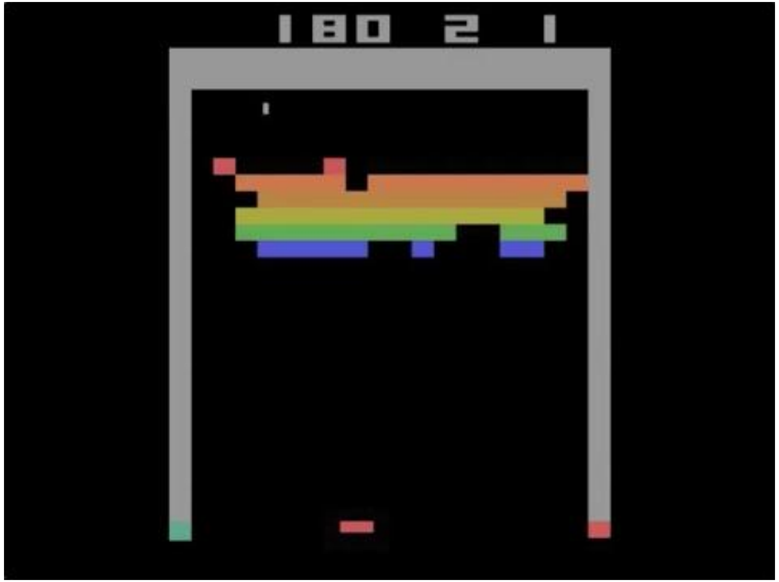
- History
  - 1957 Perceptron (Frank Rosenblatt)
  - 1969 Minsky & Papert (Perceptron book) (Funding for AI research collapsed)
  - 1980s Machine learning emerges (Find patterns in data, bottom-up)
  - 1980s Conferences for neural networks emerged
  - 1990 (SVM) (No paper was accepted by conferences)
  - 2006 (Geoffrey Hinton, Yunn LeCun, Yoshua Bengio)
    - Rephrase neural networks to deep learning
  - 2012 Imagenet-competition (Industry-wide artificial intelligence boom)
- Deep learning success
  - Computational power: Data, GPUs
  - Research: ReLU activations, Batch normalization, SGD
- Deep learning example
  - https://playground.tensorflow.org/

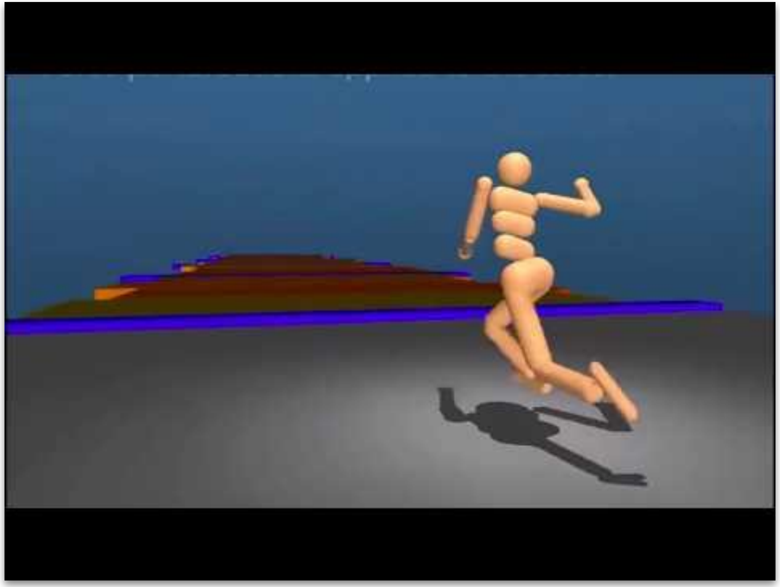


https://en.wikipedia.org/wiki/ImageNet

Deep reinforcement learning

- Atari games
- Robot locomotion





Conclusion

- No free lunch theorem
  - We can use different functions for our learning algorithm
    - Decision tree
    - Perceptron
    - SVM
    - Neural network
    - etc
  - We have to make assumption about the function which we use
  - There is no single solution for all ML problems
- Deep learning is used in many different domains
  - A function
  - Hard to find the rules
  - Can have a good amount of data