Brief Introduction to Machine Learning

- Yoonsuck Choe
  Professor
  TAMU CSE

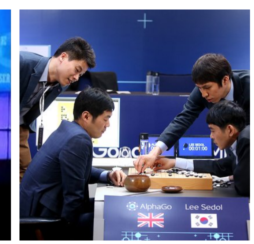
What Is Machine Learning?

- A subfield of AI that is rapidly growing in importance.
- Performance of a system is improved based on learning experience.
- Learning from data.

Why Machine Learning?

- Abundance of data: the data deluge.
  - Scientific instruments.
  - Data acquisition devices.
  - Internet and the web.
  - All sectors of human society producing and digitizing data (e.g., your cell phone).
- Not enough human expertise or human power to make sense of such huge amounts of data.

Machine Learning in the News



IBM's Watson Google DeepMind's AlphaGo

- IBM's Watson beats human champions: Jeopardy (game show)
- Google detects cats from YouTube videos.
- Google Glass app recognizes people it sees.
- Legal, medical, financial applications.
- Google DeepMind: Atari 2600 game playing, AlphaGo, AlphaStar

What Does It Take to do ML?

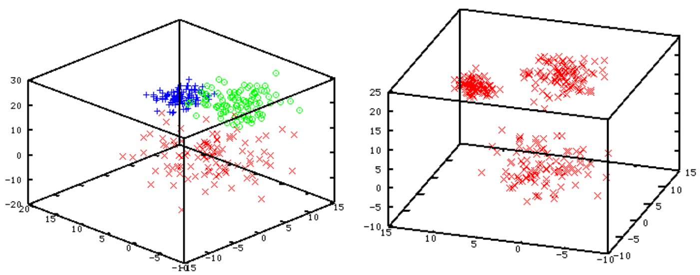
A lot of math (but not too deep):

- Linear algebra
- Calculus
- Probability and statistics
- Differential geometry
- Numerical methods

Types of Machine Learning

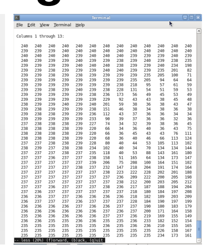
1. Supervised learning
   - Input-Target pairs
   - $\{\langle \vec{x}_i, \vec{t}_i \rangle | i = 1, 2, .., n\}$
2. Unsupervised learning
   - A bunch of inputs (unlabeled)
   - $\{\vec{x}_i | i = 1, 2, ..., n\}$
3. Reinforcement learning
   - state$_1$ $\xrightarrow[\text{action}_1]{}$ state$_2$ $\xrightarrow[\text{action}_2]{}$ state$_3$, ... , reward
   - $s_{t+1} = \delta(s_t, a_t), r_{t+1} = \rho(s_t, a_t)$

Example Data



- Left: supervised
- Right: unsupervised
- Typically very high dimensional (10,000, 1 million [or more]).

High-dimensional Data



$\longrightarrow \underbrace{[240, 240, ..., 232, ...]}_{\text{2,500-D vector}}$

- Images: these are 2D images, but ...
- These are 50×50 = 2,500-dimensional vectors.
  - Each such image is a single point in 2,500-dimensional space.

Supervised Learning

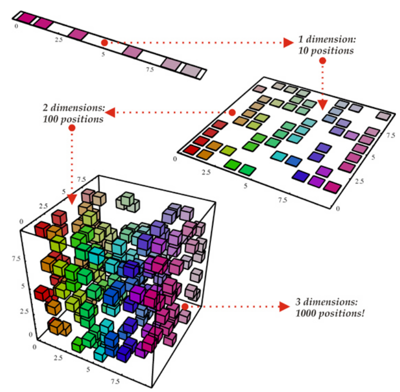
Supervised Learning

- Regression: approximating $y = f(x)$
- Classification: face recognition, hand-written character recognition, credit risk assessment, etc.
- Techniques:
  - Neural networks
  - Decision tree learning
  - Support vector machines
  - Radial basis functions
  - Naive Bayes learning
  - k-nearest neighbor

Supervised Learning Issues

- How well will it do on training inputs?
- How well will it do on novel inputs?
  - Generalization.
- How many samples needed for sufficient performance and generalization?
  - Sample complexity
  - Curse of dimensionality
  - Computational learning theory
- Catastrophic forgetting (online learning hard).

Addendum: Curse of Dimensionality

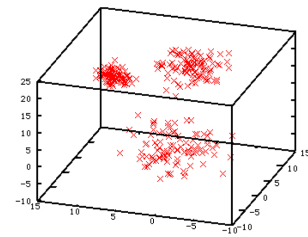


From: Yoshua Bengio's page

- Exponentially many points needed to achieve same density of training samples.

Unsupervised Learning

Unsupervised Learning



Clustering, feature extraction, blind source separation, dimensionality reduction, etc.

- Principal Component Analysis (PCA)
- Self-Organizing Maps (SOM)
- Independent Component Analysis (ICA)
- Multi-Dimensional Scaling (MDS)
- ISOMAP, Locally Linear Embedding (LLE)
- t-distr. Stochastic Neighbor Embedding (t-SNE)

Unsupervised Learning Issues

- Discovering structure.
- Discovering features.
- Removing redundancy.
- How many clusters?
- What distance measures to use?

Reinforcement Learning

Reinforcement Learning

- Very different from supervised and unsupervised learning.
- Multi agent control, robot control, game playing, scheduling, etc.
- Techniques:
  - Value function-based: Q-learning, Temporal difference (TD) learning
  - Direct policy search: Neuroevolution, genetic algorithms.

Reinforcement Learning Issues

- Discrete states and actions was a norm.
- Scalability an issue.
- Certain assumptions: state-action pair visited infinitely often.
- Online learning, safety, transfer, imitation, etc.
- Deep reinforcement learning disrupted a lot of the traditional assumptions.

Wrap Up

Summary

- Machine learning is a rapidly developing field with great promise:
  - Big data
  - Deep neural networks
  - Fast computing: GPGPU, cloud, etc.
- Three types of ML:
  - Supervised learning
  - Unsupervised learning
  - Reinforcement learning
- Need to look beyond ML:
  - ML good at solving problems, but not posing problems (Choe and Mann 2012).