

633-600 Machine Learning

- Instructor: Yoonsuck Choe
 - Contact info: PETR 327, 845-5466, choe@tamu.edu
 - Zoom: <http://tamu.zoom.us/my/yoonsuckchoe>
- TA: see web page
- Course web page: Canvas

Textbook

- Ethem Alpaydin (2021) “Introduction to Machine Learning”, 4th edition. MIT Press.
- Book webpage: <https://mitpress.mit.edu/books/introduction-machine-learning-fourth-edition>
- Optional (but strongly recommended): Tom M. Mitchell (1997) “Machine Learning”, McGraw-Hill.
- Book webpage: <http://www.cs.cmu.edu/~tom/mlbook.html>
- Text and figures, etc. will be quoted from the textbook without repeated acknowledgment. Instructor’s perspective will be indicated by “YC” where appropriate.

Course Info

- Grading, academic policy, students with disabilities, lecture notes, computer accounts, programming languages.
- See course web page.

Relation to Other Courses

Some overlaps (undergrad courses only):

- Artificial intelligence (420): decision trees (in some courses), neural networks (in some courses).
- Statistics: hypothesis testing
- (Relatively) unique to this course: computational learning theory, genetic algorithms, reinforcement learning, decision trees (in depth treatment), local learning, dimensionality reduction, deep learning

Graduate courses: Deep learning, pattern analysis, reinforcement learning.

Brief Introduction to Machine Learning

- Yoonsuck Choe
Professor
TAMU CSE

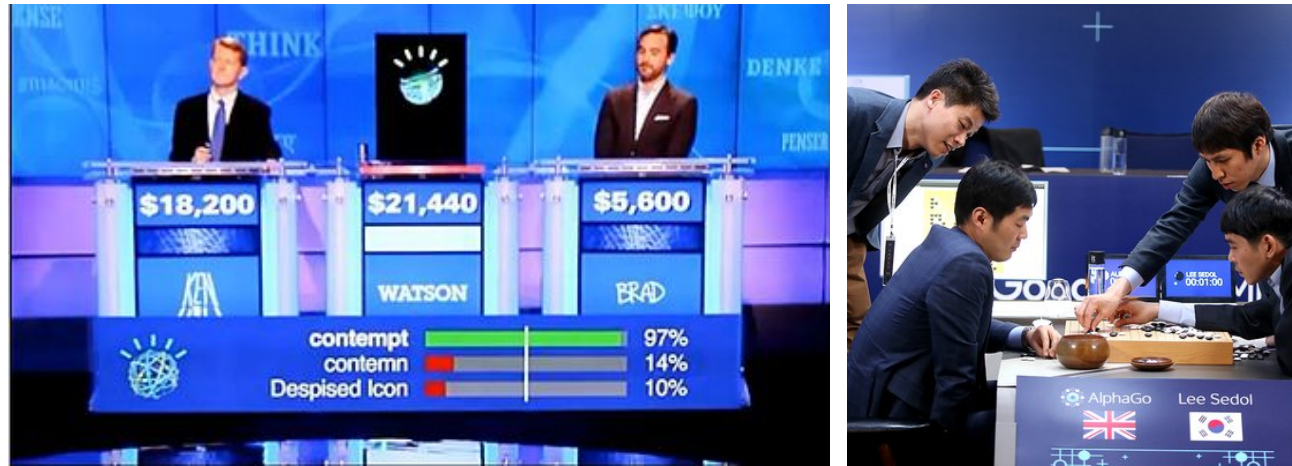
What Is Machine Learning?

- 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

ACM Turing Award 2018: Deep Learning

ACM Awards | Advanced Grades of Membership | SIG Awards | Regional Awards | Nominating Process | Awards

ACM Home | ACM A.M. Turing Award | Turing 50

ACM recognizes excellence

ACM Awards | Advanced Grades of Membership | SIG Awards | Regional Awards | Nominating Process | Awards

Awards Home | ACM A.M. Turing Award | Spotlight on Turing Laureates | ACM Prize in Computing

Home > Latest Awards News > 2018 Turing Award

Fathers of the Deep Learning Revolution Receive ACM A.M. Turing Award

Bengio, Hinton and LeCun Ushered in Major Breakthroughs in Artificial Intelligence



- ACM Turing Award 2018 goes⁹ to Deep Learning pioneers!

What Does It Take to do ML?

A lot of math:

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

Types of Machine Learning

1. Supervised learning

- Input-Target pairs
- $\{\langle \vec{x}_i, \vec{t}_i \rangle | i = 1, 2, \dots, n\}$

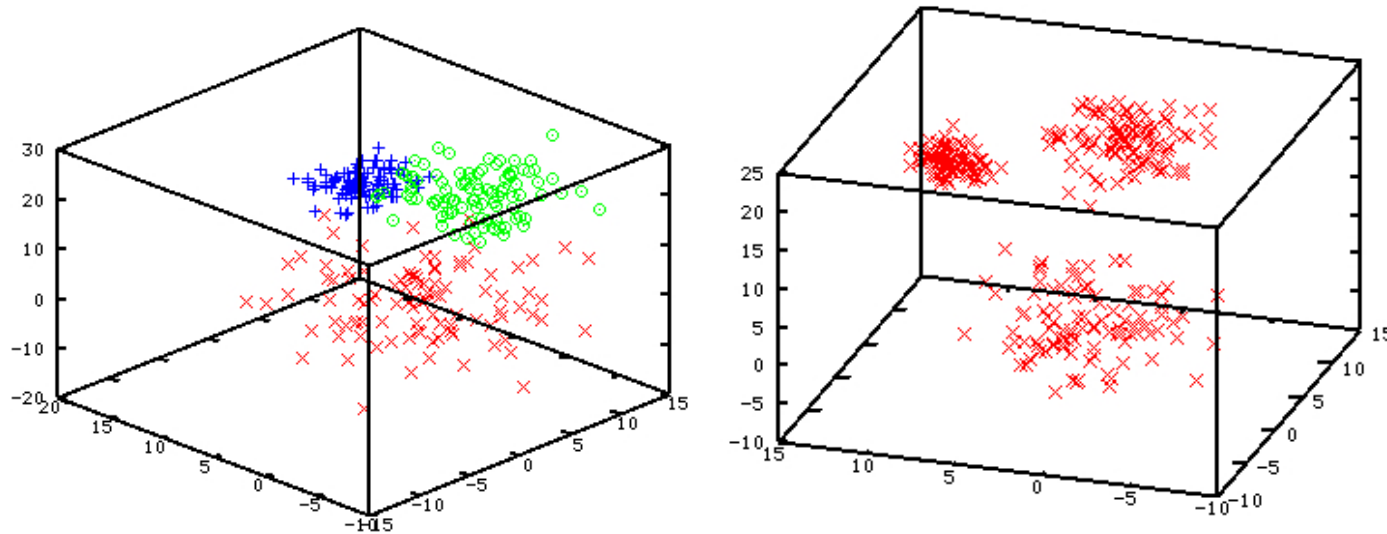
2. Unsupervised learning

- A bunch of inputs (unlabeled)
- $\{\vec{x}_i | i = 1, 2, \dots, n\}$

3. Reinforcement learning

- $\text{state}_1 \xrightarrow{\text{action}_1} \text{state}_2 \xrightarrow{\text{action}_2} \text{state}_3, \dots, \text{reward}$
- $s_{t+1} = \delta(s_t, a_t), r_{t+1} = \rho(s_t, a_t)$

Example Data



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

High-dimensional Data



Terminal

Columns 1 through 13:

240	240	240	240	240	240	240	240	240	240	240	240	240
239	239	239	240	240	240	240	240	240	240	240	240	240
240	240	240	240	239	240	240	240	239	239	240	239	239
240	239	239	239	240	239	239	238	239	240	239	238	235
239	239	239	239	240	240	238	239	239	240	234	198	
239	239	238	239	239	240	240	239	239	235	201	82	
239	239	238	239	238	239	239	239	239	235	205	180	71
239	239	239	239	239	239	239	235	205	94	64	64	
239	239	239	239	239	239	238	218	95	57	61	59	
239	239	238	239	240	239	238	228	131	54	51	59	53
239	239	239	239	239	238	236	175	56	49	45	53	49
239	239	239	239	239	239	229	92	43	43	38	45	48
238	239	239	240	239	240	201	59	38	36	38	43	47
239	238	239	239	239	238	151	46	38	34	38	36	38
239	238	238	239	239	236	112	43	37	36	36	34	34
239	239	239	239	236	133	90	39	37	36	36	32	36
237	238	238	238	239	227	74	34	32	39	38	36	45
238	238	238	238	239	228	66	34	36	40	36	43	75
238	238	238	238	239	228	66	36	45	43	43	76	111
238	238	238	238	239	228	68	36	40	46	66	113	124
237	237	238	236	239	208	60	40	44	53	105	113	182
238	239	237	238	238	234	102	40	34	70	134	134	144
237	237	237	238	237	235	118	40	53	68	143	181	180
237	237	236	237	237	238	158	51	165	64	134	173	147
237	237	237	237	237	239	206	75	288	108	164	151	182
237	237	237	237	236	237	232	147	218	204	191	174	164
237	237	237	237	237	238	223	222	228	202	201	188	
237	237	237	237	237	237	236	209	222	209	205	198	
237	237	237	237	237	237	236	212	208	203	205	212	
237	237	236	237	237	237	238	217	187	188	194	204	
237	236	237	237	237	237	237	218	180	184	197	200	
236	236	237	237	236	236	236	218	182	189	203	209	
236	237	237	237	236	237	237	228	184	190	197	199	
236	236	236	236	237	236	236	237	190	180	183	179	
236	236	236	236	236	236	237	236	237	200	171	184	158
235	236	236	236	236	236	237	236	219	169	155	149	
236	236	235	235	236	236	235	236	233	182	152	154	
235	235	235	235	235	235	236	235	236	236	210	155	165
235	235	235	235	235	235	235	235	235	235	226	158	167
235	235	235	235	235	235	235	235	235	234	173	161	

→ $[240, 240, \dots, 232, \dots]$
2,500-D vector

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

Supervised Learning

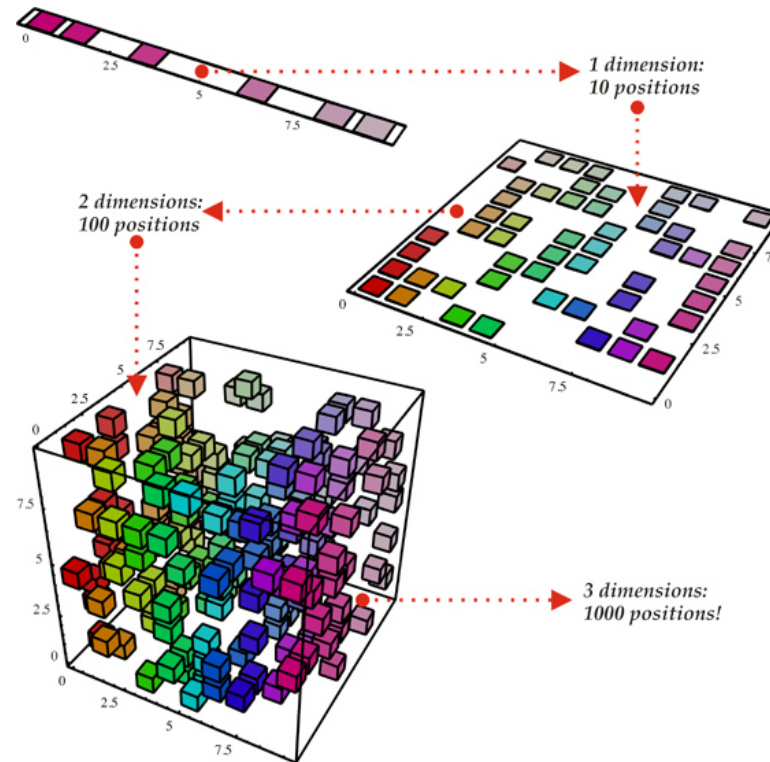
Supervised Learning

- 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

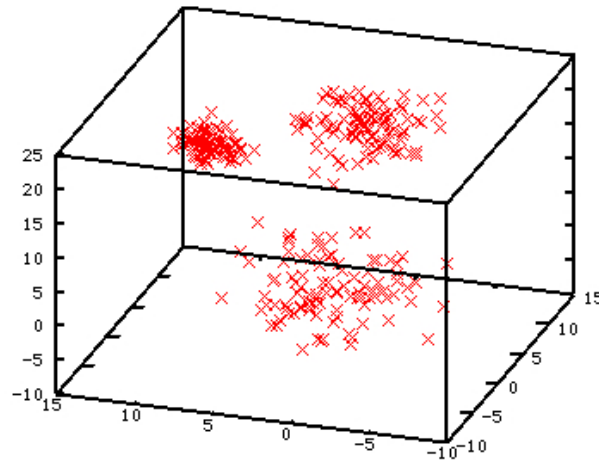


From: Yoshua Bengio's page

- 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).