SC4000/CZ4041/CE4041: Machine Learning

Lecture 1a: Introduction

Kelly KE
School of Computer Science and Engineering
NTU, Singapore

General Information

- Instructors:
  - Dr. Kelly KE (1st half: Weeks 1-6)
  - Dr. Albert Boyang LI (2nd half: Weeks 7-12)
- Lecture time/venue (1st half):
  - Weeks 1-6, Thursdays 15:30 – 17:20
  - LT2A (in person)
- Tutorial time/venue (1st half):
  - Start from Week 2 (Weeks 2, 4, 5, 6), Mondays 15:30 – 16:20
  - LT2A (in person)

General Information (Part-Time)

- Instructors:
  - Dr. Kelly KE (1st half: Weeks 2-7)
  - Dr. Albert Boyang LI (2nd half: Weeks 8-13)
- Lecture & Tutorial time/venue (1st half):
  - Weeks 2-7, Thursdays 18:30 – 21:20
  - TR+3 (in person)
- Notes:
  - Lecture and tutorial start from Week 2.
  - Teaching mode: lecture video watching (online learning) + "review lecture (in person) + tutorial (in person)"
  - Students are highly encouraged to watch lecture recordings from previous week before attending the in-person session.

General Information (cont.)

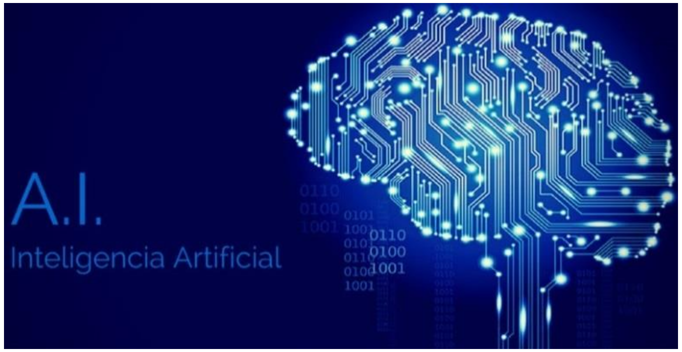
- Q&A (about 1st half teaching content & course project)
  - Send questions via email ypke@ntu.edu.sg or Teams
  - Make an appointment via email
  - After lectures or tutorials
- Course Webpage
  - SC4000/CE4041/CZ4041 @ NTULearn (official course webpage)

Evaluation

- Course project (40%)
  - Group-based (**<u>maximal size</u>**: 5 students)
  - Course report (30%) + presentation video (10%)
  - A Kaggle competition. Detailed information including assessment criteria about the course project will be released in Week 2 tutorial session
- <u>Open book</u> final exam (60%)
  - No restrictions on the number of hardcopies
  - Date: 24 April 2023
  - Duration: 2 hours

Hot Keywords in the IT Sector

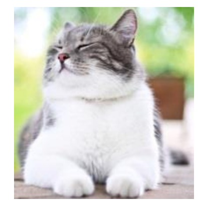
Machine Learning



What is Machine Learning?

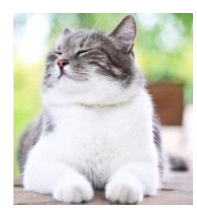
- Motivated by how human beings learn from examples/experience/exercise



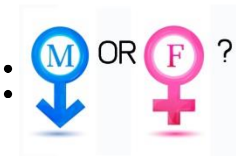


- Focuses on the development of computer programs that can teach themselves to grow from data and change when exposed to new data

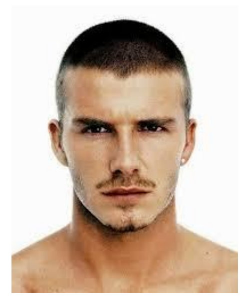




A Motivating Example

- Given a face image, to classify the face gender: 






- Once upon a time, to develop an AI system to solve such a task, developers or domain experts need to provide rules and implement them in the system

If the face has long hair and does not have moustache, then this is a "female" face;
If the face has short hair and moustache, then this is a "male" face.

A Motivating Example (cont.)

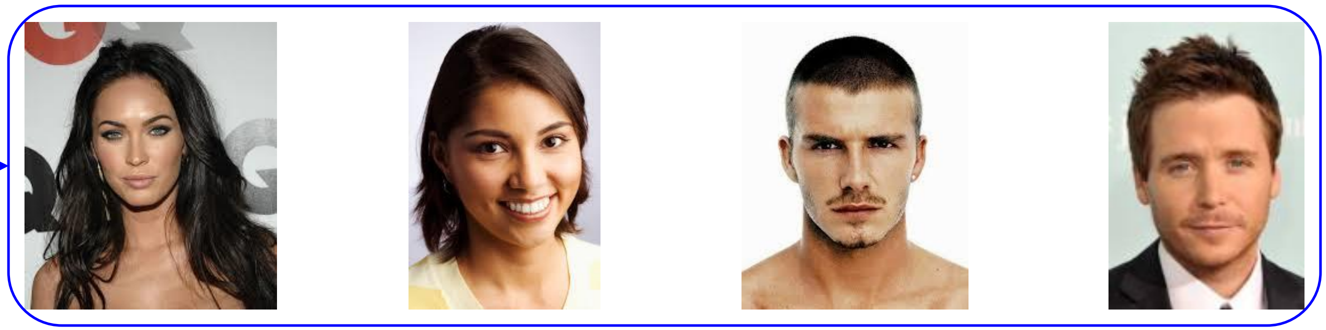
- Limitations:
  - Time consuming
  - The defined rules may not be complete
  - Not able to handle uncertainty

If the face has long hair and does not have moustache, then this is a "female" face;
If the face has short hair and moustache, then this is a "male" face.

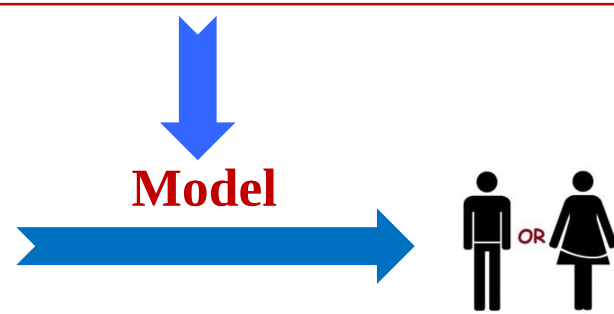
A Motivating Example (cont.)

- How about letting the machine learn the rules by itself?
  - The computer is presented with example inputs and their desired outputs, and the goal is to "learn" a set of general rule or "model" that maps inputs to outputs

Example inputs →

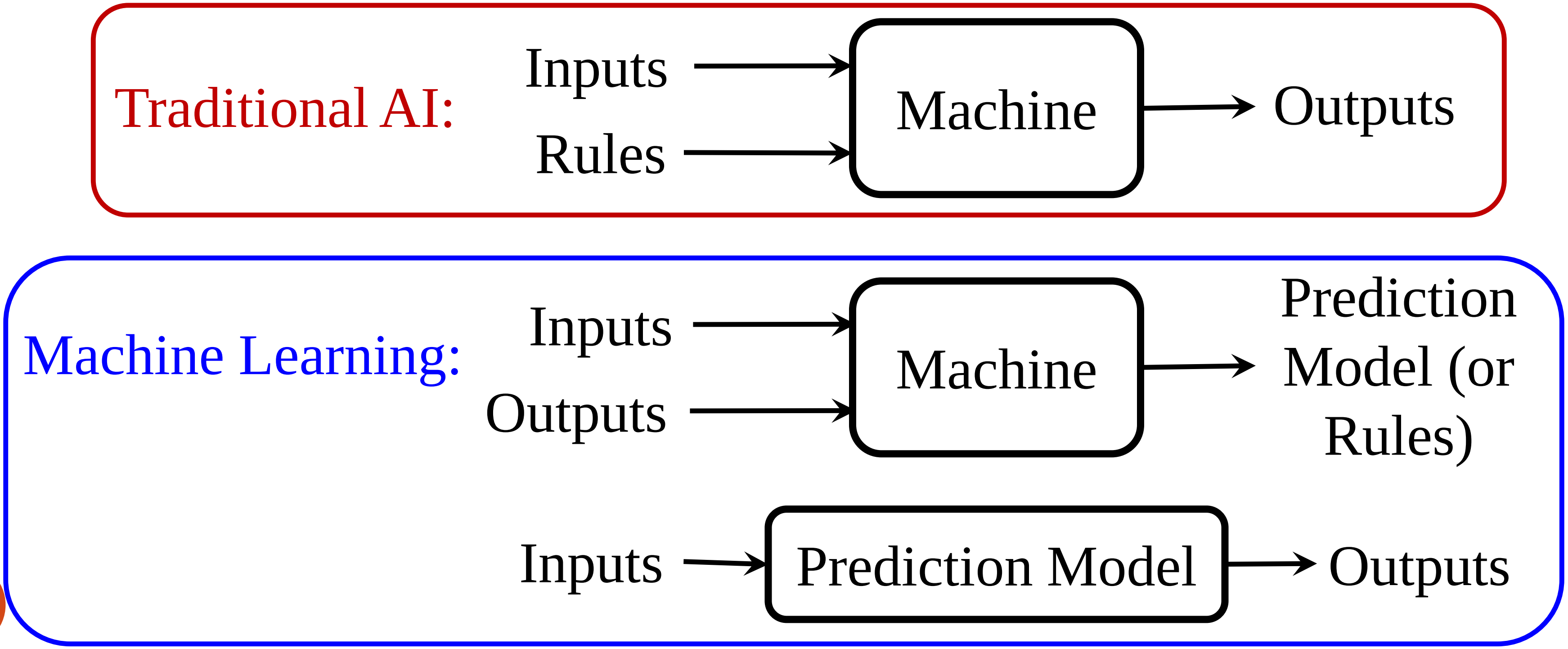


Outputs → Female Female Male Male

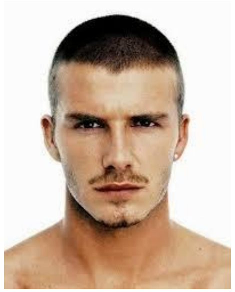


Machine Learning Definition

- Machine learning is a type of artificial intelligence that provides computers with the ability to learn from examples/experience without being explicitly programmed



How to Represent an Example?

- Feature engineering (not machine learning focus)
- Representation learning (one of the crucial research topics in machine learning)
    - Deep learning is the current most effective approach to representation learning

Machine Learning $\stackrel{?}{=}$ Deep Learning $\stackrel{?}{=}$ AI

- Machine learning is a field of AI – many other fields
- Deep learning is a type of methodologies of machine learning – many other methodologies in machine learning
- Machine learning has become a primary mechanism for data analytics (key in data science)
- Nowadays, machine learning is more and more interdisciplinary:
  - Distributed/parallel computing + machine learning → Distributed/parallel machine learning
  - Machine learning + hardware → AI chips

Different Paradigms/Settings

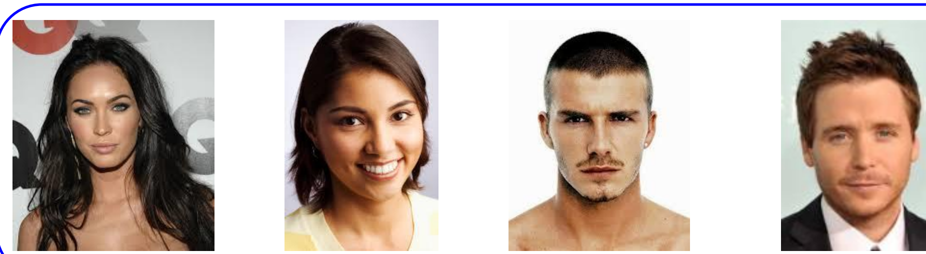
- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning
- Advanced paradigms:
  - Semi-supervised learning
  - Active learning
  - Transfer learning

Supervised Learning

- The examples presented to computers are <u>pairs of inputs and the corresponding outputs</u>, the goal is to "learn" a mapping or model from inputs to labels

Labeled training data

Inputs: Face images



Outputs: Female or Male

Female Female Male Male

$$f: \textbf{label} = f(\textbf{input})$$

Outputs are discrete (i.e., categorical) values → classification

Labels are continuous values → regression

Supervised Learning – Regression

Different Paradigms/Settings

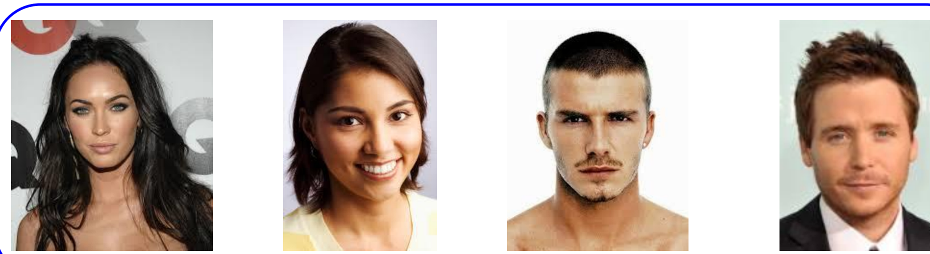
- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning
- Advanced paradigms:
  - Semi-supervised learning
  - Active learning
  - Transfer learning

Unsupervised Learning

- The examples presented to computers are a set of inputs without any outputs, the goal is to “learn” an intrinsic structure of the examples, e.g., clusters of examples, density of the examples

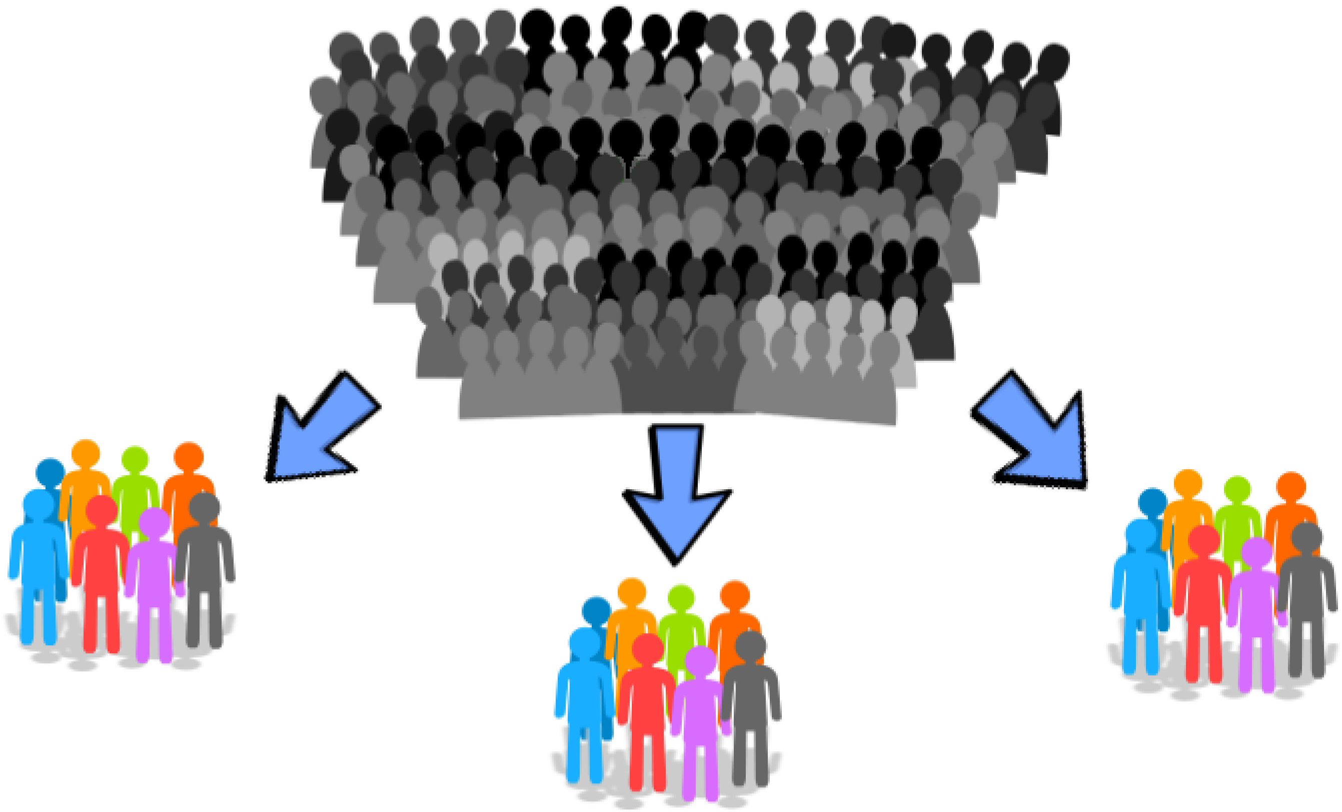
Unlabeled training data

Inputs: Face images



Groups of similar faces

Unsupervised Learning – Clustering



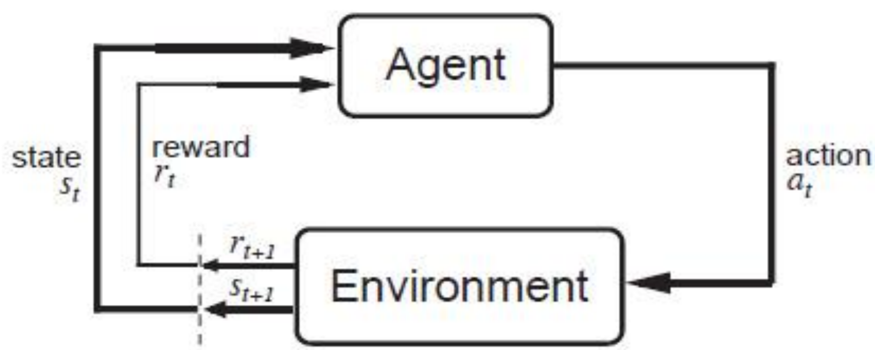
User Segmentation

Different Paradigms/Settings

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning
- Advanced paradigms:
  - Semi-supervised learning
  - Active learning
  - Transfer learning

Reinforcement Learning

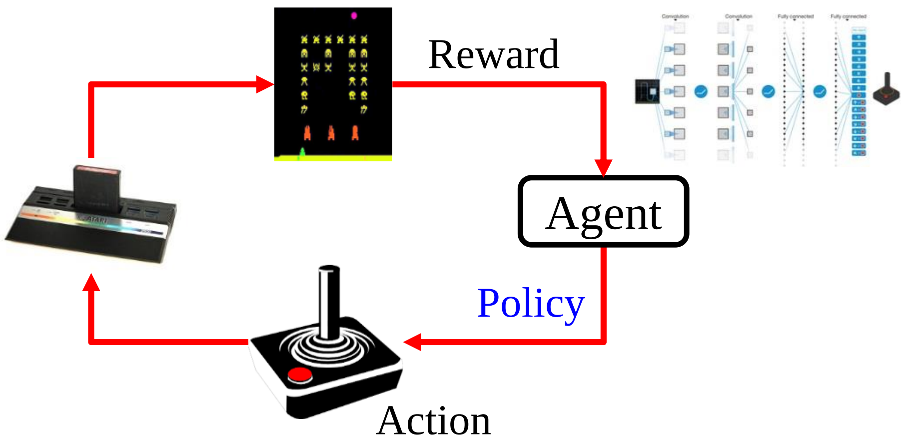
- Learning by interacting with an environment to achieve a goal
- Objective: to learn an optimal policy mapping states to actions

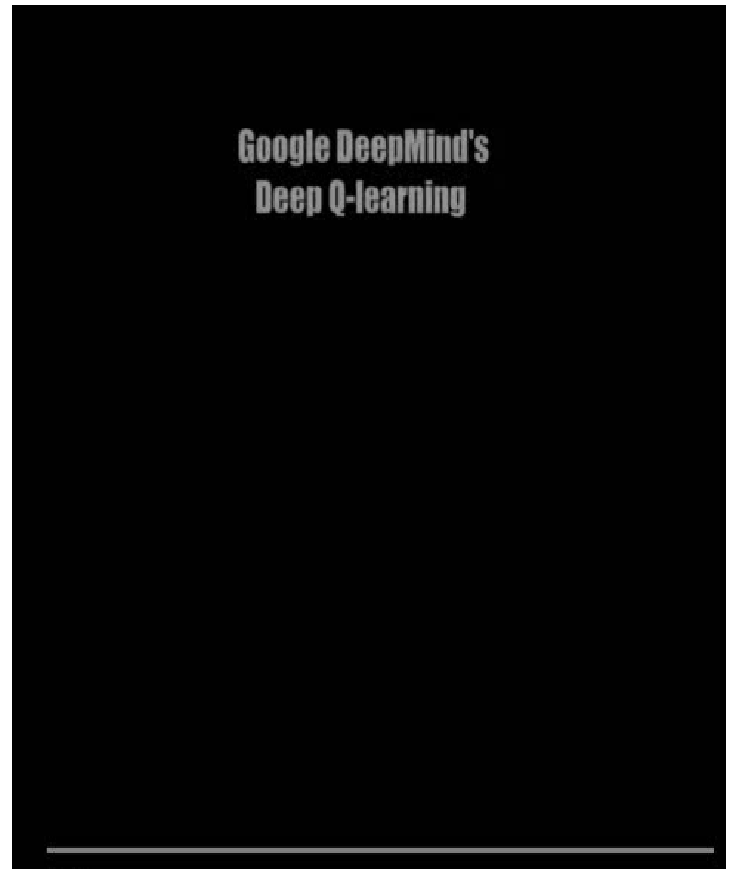


$$s_0 \xrightarrow[r_0]{a_0} s_1 \xrightarrow[r_1]{a_1} s_2 \xrightarrow[r_2]{a_2} s_3$$

Reinforcement Learning (cont.)

- Deep Q-Network (DQN) [1]
  - Play Atari 2600 Games





[1] Mnih et al, Human-level control through deep reinforcement learning. *Nature*, 2015

Different Paradigms/Settings

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning
- Advanced paradigms:
  - Semi-supervised learning
  - Active learning
  - Transfer learning

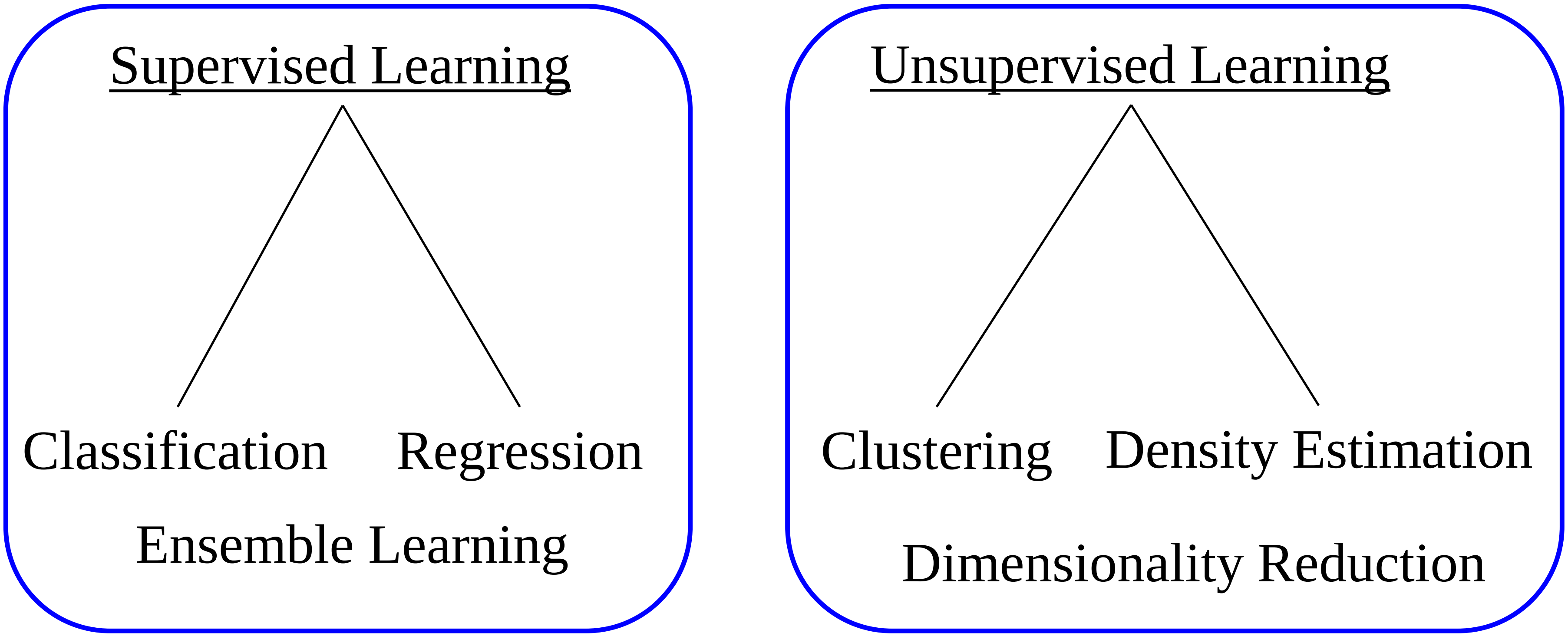
Limitation of Supervised Learning

- Require sufficient labeled data to train a precise model (i.e., a model with good prediction performance)
  - Sufficiency of labeled data is context-aware, depending on different kinds of applications and specific datasets
- When there is insufficient labeled data, can we still train a precise model?
  - Advanced machine learning paradigms

Additional Self-Readings

- Reinforcement Learning:
  - *Reinforcement learning: a survey*
- Semi-supervised learning:
  - *Semi-supervised learning literature survey*
- Active learning:
  - *Active learning literature survey*
- Transfer learning:
  - *A survey on transfer learning*

Course Scope



1st Half Schedule

Lectures		
Week 1	12th Jan.	L1: Introduction and Overview of Supervised Learning
Week 2	19th Jan.	L2: Bayesian Classifiers & Bayesian Decision Theory
Week 3	26th Jan.	L3: Naïve Bayes Classifiers
Week 4	2nd Feb.	L4: Bayesian Networks
Week 5	9th Feb.	L5: Decision Trees
Week 6	16th Feb.	L6: Generalization & Nearest-Neighbor Classifiers

Tutorials		
Week 2	16th Jan.	Course project introduction
Week 4	30th Jan.	Solutions to L2 & L3 questions
Week 5	6th Feb.	Solutions to L4 questions
Week 6	13th Feb.	Solutions to L5 & L6 questions

1st Half Schedule (Part-Time)

Lectures & Tutorials (In Person)		
Week 2	19th Jan.	Review of L1 & course project intro
Week 3	26th Jan.	Review of L2 & Solutions to L2
Week 4	2nd Feb.	Review of L3 & Solutions to L3
Week 5	9th Feb.	Review of L4 & Solutions to L4
Week 6	16th Feb.	Review of L5 & Solutions to L5
Week 7	23rd Feb.	Review of L6 & Solutions to L6

Watch Lecture Recordings (Online Learning)	
Weeks 1-2	L1: Introduction and Overview of Supervised Learning
Weeks 2-3	L2: Bayesian Classifiers & Bayesian Decision Theory
Weeks 3-4	L3: Naïve Bayes Classifiers
Weeks 4-5	L4: Bayesian Networks
Weeks 5-6	L5: Decision Trees
Weeks 6-7	L6: Generalization & Nearest-Neighbor Classifiers

Course Objective

- To provide students with essential concepts and principles of machine learning algorithms
- To enable students to understand how to revise or design (beyond how to use) various machine learning algorithms to solve supervised learning and unsupervised learning problems

Breadth and Depth

- Through lectures:
  - Supervised learning techniques
    - Classic classification and regression algorithms, ensemble learning methods
  - Unsupervised learning techniques
    - Classic clustering, density estimation and dimensionality reduction algorithms
- Self-learning through doing a course project:
  - Real-world Applications

Breadth and Depth (cont.)

- Focus on introducing well-known concepts and fundamental methodologies of machine learning
  - Motivations
  - Induction of mathematical models (mathematics)
- For those who want to learn more, some up-to-date techniques and advanced issues will be mentioned
  - Details cannot be covered in lecture, some additional materials for reading will be suggested (*optional*)

Relationships to Other Modules

SC4000/CZ-CE4041: Machine Learning

Modern AI approaches:

- Classification:
  - Bayesian Decision Theory
  - Bayesian Classifiers (Naïve Bayes & Bayesian Networks)
  - Decision Trees
  - Nearest-Neighbor Classifier
  - Artificial Neural Networks
  - Support Vector Machines
- Regression:
  - Regularized Linear Regression
- Clustering:
  - K-means and its variants
  - Hierarchical clustering
- Density Estimation
- Ensemble Learning
- Dimensionality Reduction

Objective: Deeply understand principles

- CZ3005: Artificial Intelligence

Classic AI approaches:

- Search
- First Order Logic

Reinforcement learning

- CZ4042/CE4042: Neural Networks and Deep Learning

Various Architectures of Neural Networks

- CZ4032/CE4032: Data Analytics and Mining

Objective: Understand how to use

Mathematics Background

Various machine learning applications:
Face recognition, object recognition, text mining, activity recognition, stock price prediction, etc.

Various learning paradigms:
Supervised learning, unsupervised learning, reinforcement learning, other advanced learning.

Various types of methodologies:
Graphical models, deep learning, empirical risk minimization, entropy-based models, kernel methods, etc.

Various mathematical techniques:
Probability theory, linear algebra, calculus, optimization, information theory, functional analysis, etc.

NTU Confessions

January 23, 2017 ·

"There are a lot of year 4 CS modules that require a very solid math foundation to the extent that I think if math majors try taking them, most of them will score better than actual CS students themselves. I believe NTU math graduates will also perform better if they are to take CS graduate courses than actual NTU CS graduates too. This is because we're not exposed to linear algebra / statistics / calculus / number theory / functional analysis / optimization as deeply, if at all. We mostly are only taught about coding and how to software project management in year 2-3. The only math we do in year 1 is way too basic. I dont see how most of us have the foundation necessary to learn more advanced topics in CS and survive pursuing Masters / PhD in many interesting specializations in CS. It's like we are limited to only those areas that require little to no math at all despite us having an actual bachelors degree in CS.

Then again, most CS majors dont care about more specialized topics in CS and have no interest in pursuing further education in CS, because most of us are qualified to become software engineers once we receive our bachelors degree already which allow us to earn quite a lot already. But I think this issue shouldnt be neglected. We need more math in our CS course, whether you like it or not."

#NTUConfessions20807

Textbook and Reference

- Textbook:
  - **Introduction to Machine Learning (2nd Ed.)**, by Ethem Alpaydin, The MIT Press, 2010.
- Reference:
  - **Pattern Classification (2nd Ed.)**, by Richard Duda, Peter Hart, and David Stork, Wiley-Interscience, 2000.
  - **Introduction to Data Mining**, by Pang-Ning Tan, Michael Steinbach, and Vipin Kumar, Addison Wesley, 2005.
  - **Pattern Recognition and Machine Learning**, by Christopher M. Bishop, Springer, 2006.
- Regarding Mathematics:
  - Part I of the MIT Press book "*Deep Learning*" http://www.deeplearningbook.org/

Useful Resources: Datasets

- UCI Repository:
  - http://www.ics.uci.edu/~mlearn/MLRepository.html
- Kaggle:
  - http://www.kaggle.com/

Useful Resources: Libraries

- scikit-learn (Python) – <u>recommended</u>:
  - http://scikit-learn.org/stable/
- MALLET (Java)
  - http://mallet.cs.umass.edu/
- Weka (Java)
  - http://www.cs.waikato.ac.nz/ml/weka/
- Tensorflow:
  - https://www.tensorflow.org/
- Pytorch:
  - https://pytorch.org/
- Many other libraries on deep learning
  - http://deeplearning.net/software_links/

scikit-learn: machine learning in

scikit-learn.org/stable/

Install User Guide API Examples More

Go

scikit-learn

Machine Learning in Python

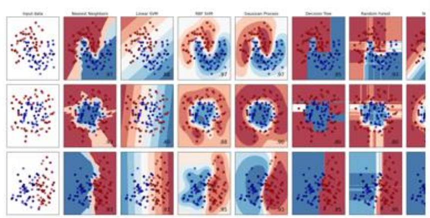
Getting Started | Release Highlights for 0.24 | GitHub

- Simple and efficient tools for predictive data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

Classification

Identifying which category an object belongs to.

Applications: Spam detection, image recognition.
Algorithms: SVM, nearest neighbors, random forest, and more...

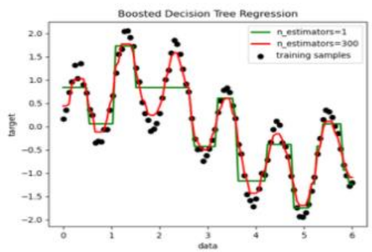


Examples

Regression

Predicting a continuous-valued attribute associated with an object.

Applications: Drug response, Stock prices.
Algorithms: SVR, nearest neighbors, random forest, and more...

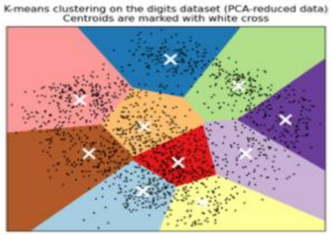


Examples

Clustering

Automatic grouping of similar objects into sets.

Applications: Customer segmentation, Grouping experiment outcomes
Algorithms: k-Means, spectral clustering, mean-shift, and more...



Examples

Dimensionality reduction

Reducing the number of random variables to consider.

Applications: Visualization, Increased efficiency
Algorithms: k-Means, feature selection, non-negative matrix factorization, and more...

Model selection

Comparing, validating and choosing parameters and models.

Applications: Improved accuracy via parameter tuning
Algorithms: grid search, cross validation, metrics

Preprocessing

Feature extraction and normalization.

Applications: Transforming input data such as text for use with machine learning algorithms.
Algorithms: preprocessing, feature extraction, and more...

Useful Resources: Conferences

- International Conference on Machine Learning (ICML)
- Neural Information Processing Systems (NIPS)
- Conference on Learning Theory (COLT)
- Uncertainty in Artificial Intelligence (UAI)
- International Conference on AI & Statistics (AISTATS)
- International Joint Conference on Artificial Intelligence (IJCAI)
- AAAI Conference on Artificial Intelligence (AAAI)
- International Conference on Learning Representations (ICLR)

Useful Resources: Journals

- Journal of Machine Learning Research (JMLR)
- Machine Learning (MLJ)
- IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)
- IEEE Transactions on Neural Networks and Learning Systems (TNNLS)
- Artificial Intelligence (AIJ)
- Journal of Artificial Intelligence Research (JAIR)

Thank you!