Machine Learning and Statistical Data Analysis

Sugiyama-Yokoya-Ishida Lab

{sugi, yokoya, ishi}@k.u-tokyo.ac.jp

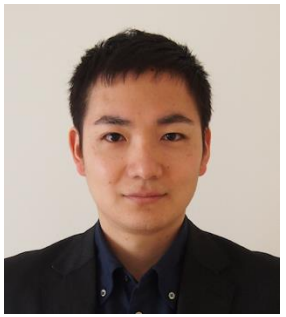
http://www.ms.k.u-tokyo.ac.jp

東京大学
THE UNIVERSITY OF TOKYO

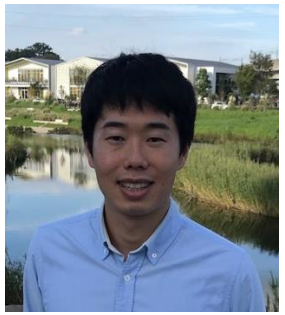
Faculty Members

- Masashi Sugiyama (Professor)
  - Machine learning theory and algorithm (weakly supervised learning, noise robust learning, transfer learning, etc.)

- Naoto Yokoya (Professor)
  - Image processing
  - Remote sensing



- Takashi Ishida (Associate Professor)
  - Weakly supervised learning
  - Limit learning performance estimation



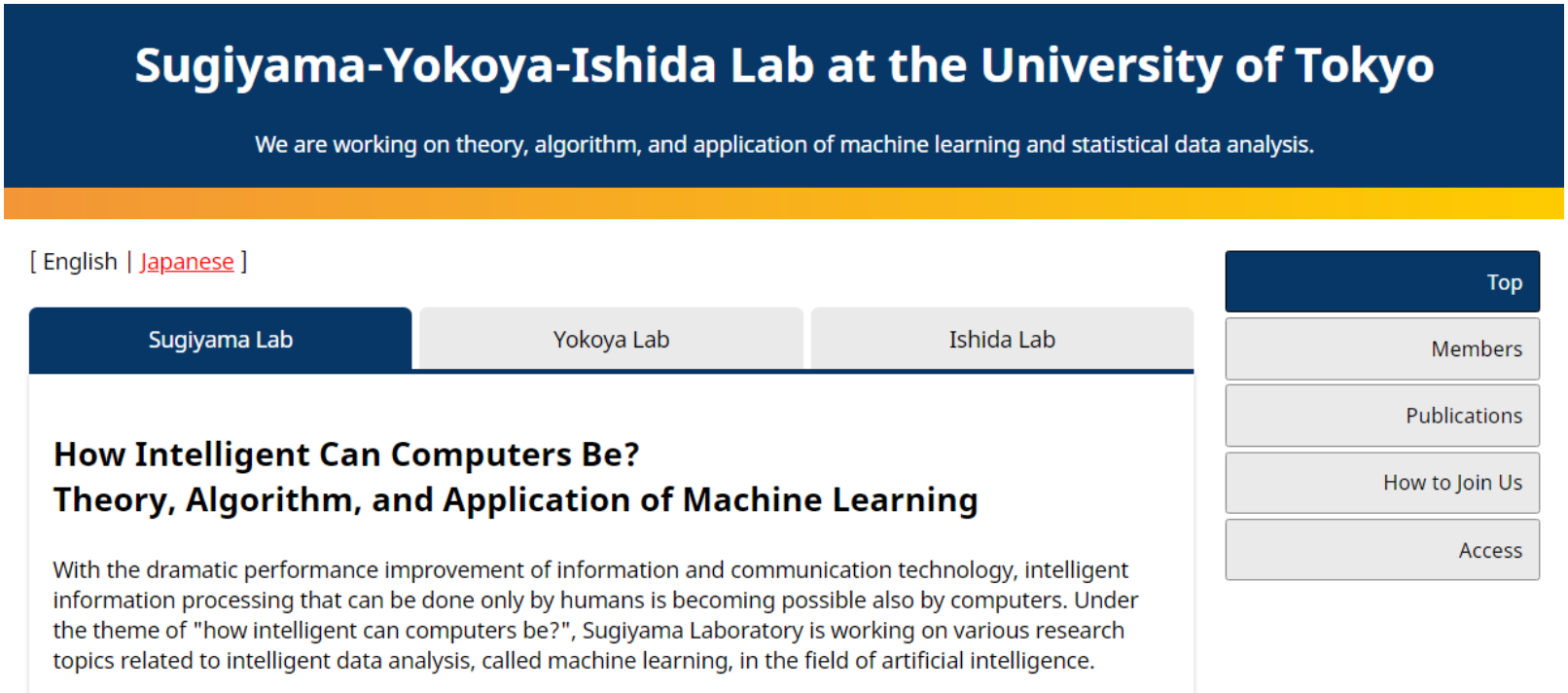
Staffs' Affiliations

- Graduate School of Frontier Sciences (K)
- Graduate School of Information Science and Technology (H)
- Faculty of Science (H)
- The Institute for AI and Beyond (H)
- RIKEN Center for Advanced Intelligence Project (N)

Current members:
Doctoral students: 24
Master students: 21
Research/auditing students: 2

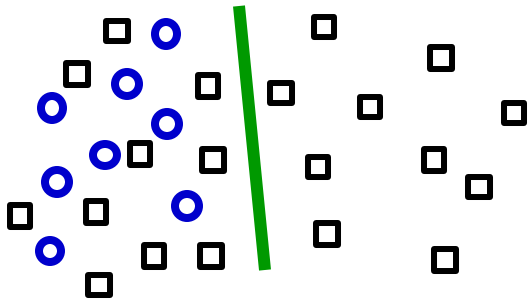
How Intelligent Can Computers Be?

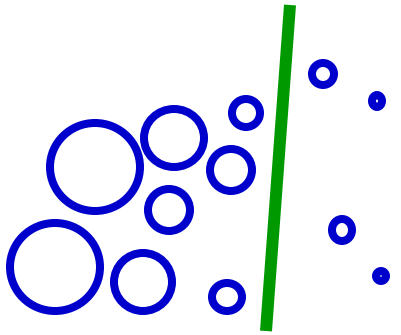
- We are interested in machine learning:
  - Construction of fundamental theories.
  - Development of practical algorithms.
  - Application in real-world problems.

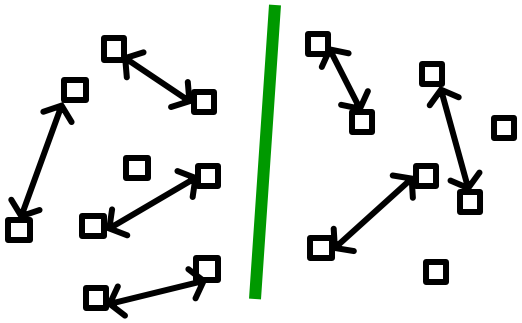


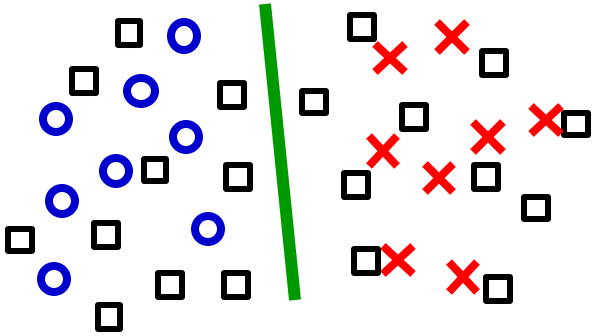
Recent Research Topics

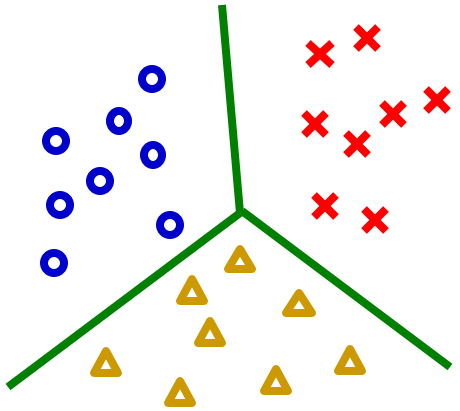
- Weakly supervised learning:
  - Labeling data is costly.
  - But learning from small data is generally difficult.
  - Let's utilize "weak" data that can be collected easily!
  - Examples: PU, PNU, Pconf, UU, SDU, Comp. learning, etc.







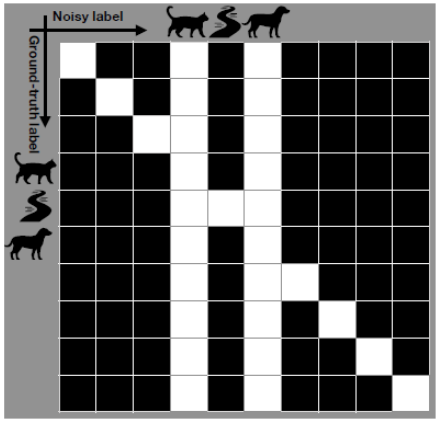


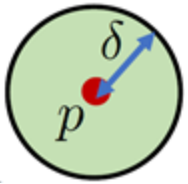


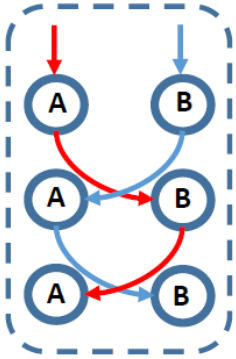
Recent Research Topics

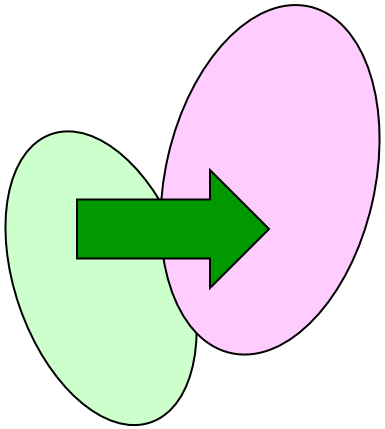
Robust and reliable learning:

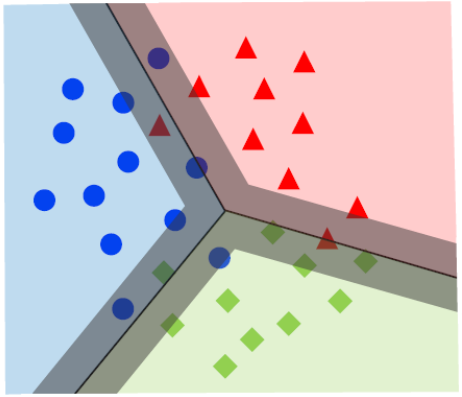
- Data is often contaminated by noise, outlier, non-stationarity, etc.
- Standard machine learning methods do not work well with such data.
- We are developing new technologies that can overcome these difficulties.
- Examples: Co-teaching, label-noise learning, classification with rejection, sequential decision making, transfer learning, etc.





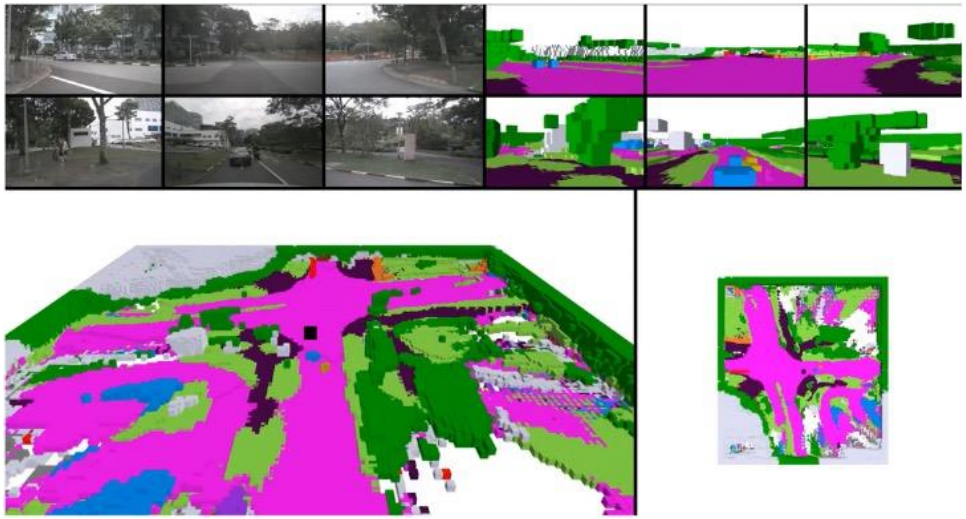






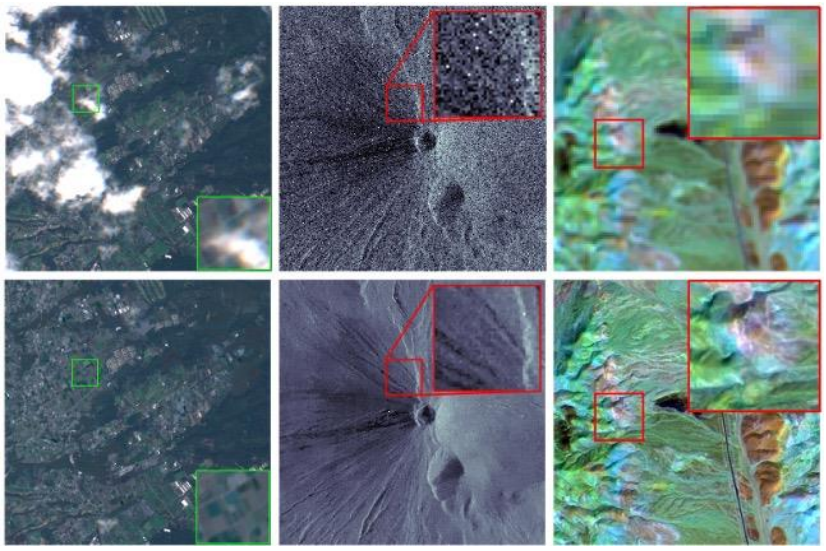
Recent Research Topics

- Understanding complex real-world scenes
  - Environmental variability, scarce labels & resources
  - Developing data fusion & resource efficient algorithms for practical computer vision
  - Examples: Multimodal learning, lightweight model design, few-shot learning, etc.



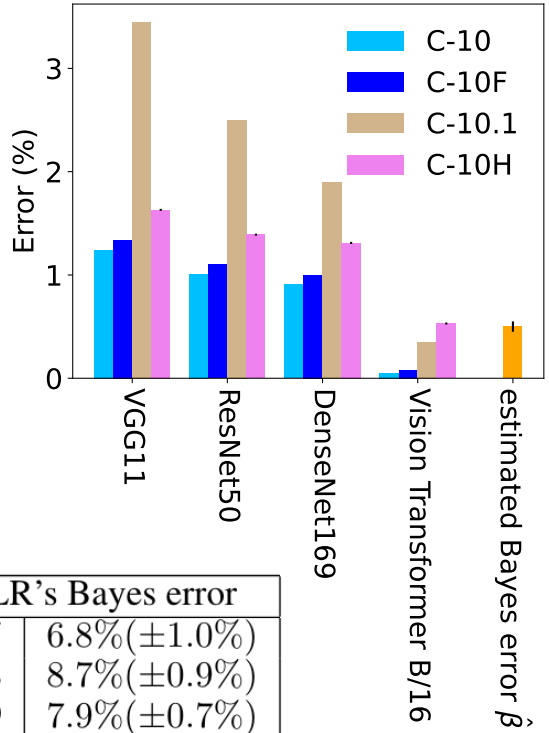
Recent Research Topics

- Remote sensing image processing
  - Resolution is limited & pixel-wise labeling is costly
  - Developing new technologies for super-resolution & map creation with low-cost supervision
  - Examples: Guided super-resolution, learning from synthetic data, mixed supervision learning, etc.



Recent Research Topics

- Estimating the upper bound of prediction performance
  - Useful for comparing with SOTA, detecting test-overfitting, & understanding the difficulty of datasets.
  - Developed a Bayes error estimation method, which represents the best possible performance for classification.



ICLR's Bayes error	
2017	6.8%(±1.0%)
2018	8.7%(±0.9%)
2019	7.9%(±0.7%)
2020	8.8%(±0.5%)
2021	9.3%(±0.5%)
2022	9.6%(±0.5%)
2023	8.0%(±0.4%)

Table 1: Parenthesis is 95% confidence interval.

Sugiyama-Yokoya-Ishida Lab:
Machine Learning and Statistical Data Analysis

- Goal: Develop computers that learn like us.
  - Supervised learning: We directly help computers.
  - Unsupervised learning: We do not help computers.
  - Reinforcement learning: We indirectly help computers.
- Research aspects:
  - Theory: Probability, statistics, optimization, information theory, etc.
  - Algorithm: Effectiveness, efficiency, practicality, etc.
  - Application: signal, image, sensor, web, language, robot, biology, brain, medicine, etc.