Machine Learning from Weak Supervision:
An Empirical Risk Minimization Approach

Masashi Sugiyama

RIKEN and The University of Tokyo, Japan.

https://www.ms.k.u-tokyo.ac.jp/sugi/   sugi@k.u-tokyo.ac.jp

Machine learning (ML) is a sub-field of artificial intelligence (AI), which is aimed at investigating computer algorithms that improve themselves automatically through experience. ML has been one of the most evolved and deepened research topics in science and technology in the early 21st century, and it is the technology that boosted the use of AI in the real world. Nowadays, ML-based AI systems have been deployed vitally in pioneering new business as well as advancing scientific research and technological development.

So far, success of ML has mainly been in the virtual world, such as e-commerce, social networks, and gaming, since ML-based AI systems need to be trained with big data that contain rich supervised information. However, once we try to apply ML to physical-world problems in our real world, such as medical diagnosis, natural disaster, and education, it is extremely difficult or even impossible to collect such a huge amount of fully supervised data. Thus, data collection is one of the critical bottlenecks for AI to be further penetrated in our society. Thus, there is an urgent need to develop novel theory and algorithm of ML that allow us to train ML-based AI systems from limited supervision.

This lecture is aimed at providing basics and practical algorithms of weakly supervised classification, based on the monograph we published in 2022 [1]. By weakly supervised classification, we do not mean that we try to train a classifier from small training data, which is mathematically not possible without imposing strong assumptions on data and models. Instead, we try to train a classifier with a large amount of data that can be easily collected. Such easily collectible data usually contain weaker supervised information than expensive fully supervised data, but we will show that it is possible to train a classifier only from such weakly supervised data as if we have fully supervised data.

[1] Sugiyama, M., Bao, H., Ishida, T., Lu, N., Sakai, T., & Niu, G. Machine Learning from Weak Supervision: An Empirical Risk Minimization Approach, MIT Press, Cambridge, Massachusetts, USA, 2022.