

Machine Learning for Bio-data Mining

Lecturer: Kyu-Baek Hwang

Affiliation: Soongsil University

Day: 15-16th of Jan, 2020

Lecture Outline:

This course covers machine learning and R programming for biomedical data mining. Basic concepts in machine learning such as supervised learning, unsupervised learning, and the bias-variance trade-off are taught. We focus on two main problems in supervised learning, i.e., regression and classification. Linear methods for regression and classification are intensively covered. Then, methods for estimating the generalization error of a machine learning model are taught. Finally, we learn shrinkage methods for linear regression, which alleviate the sparse data problem, frequently encountered in biomedical applications. In lab, we learn to use RStudio for mining biomedical data using machine learning methods.

***Prerequisite: Basic programming abilities (in any languages such as C and Java) and undergraduate-level knowledge of probabilities and statistics.

Schedule

	Time	Content	Lecturer	Lecture Type
	Session 1 (09:00-10:30)	✓ Introduction ✓ Bias-variance trade-off	Kyu-Baek Hwang	Lecture
	(10:30-10:45)	Break		
Day 1	Session 2 (10:45-12:15)	✓ Lab for basics in R programming ✓ Linear regression	Kyu-Baek Hwang	Lab/Lecture
		Lunch		
	Session 3 (13:30-15:00)	✓ Linear regression (cont'd)	Kyu-Baek Hwang	Lecture
	(15:00-15:15)	Break		
	Session 4 (15:15-16:45)	✓ Lab for linear regression	Kyu-Baek Hwang	Lab

	Session 5 (09:00-10:30)	Linear classification	Kyu-Baek Hwang	Lecture
	(10:30-10:45)	Break		
Day 2	Session 6 (10:45-12:15)	Lab for linear classification	Kyu-Baek Hwang	Lab
		Lunch		
	Session 7 (13:30-15:00)	✓ Resampling Shrinkage in linear regression	Kyu-Baek Hwang	Lecture
		Break		
	Session 8 (15:15-16:45)	Lab for resampling/shrinkage	Kyu-Baek Hwang	Lab