Course number	U-LAS11 10009 LE55		
Course title (and course title in English)	Basic Data Analysis-E2 Basic Data Analysis-E2	Instructor's name, job title, and department of affiliation	Institute for Life and Medical Sciences Associate Professor,VANDENBON, Alexis
Group	Natural Sciences	Field(Classification)	Data Science(Foundations)
Language of instruction	English	Old group	Group B
Number of credits	2		
Number of weekly time blocks	1	Class style	Lecture (Face-to-face course)
Year/semesters	2025・First semester		
Days and periods	Mon.4	Target year	All students
Eligible students	For all majors		

[Overview and purpose of the course]

Nowadays, research in many fields of science is increasingly dependent on large amounts of data. The key problem is how to turn this data into new knowledge. This course covers a wide variety of data analysis and machine learning approaches. The course starts with an introduction of the basic concepts in machine learning. After that, we will introduce regression and classification methods, including linear models, tree-based methods, support vector machines, neural networks, deep learning, and principal component analysis.

[Course objectives]

Students will learn about basic concepts in data analysis and statistical learning, such as regression and classification problems, and supervised and unsupervised machine learning. Students will become familiar with strengths and weaknesses of several approaches, and learn how to apply them on real datasets.

[Course schedule and contents)]

Lectures 1 to 3. Introduction to data analysis and machine learning: We will discuss data analysis in the context of scientific investigation. Using several examples, the concepts of supervised and unsupervised learning, regression and classification problems, and assessment of model accuracy will be introduced.

Lectures 4 and 5. Linear regression: Introduction to linear regression as a simple supervised learning approach. We will cover simple and multiple linear regression, discuss how to interpret models, and compare linear regression with K-nearest neighbors.

Lectures 6 and 7. Classification methods. We will introduce classification methods, including logistic regression, linear discriminant analysis, and quadratic discriminant analysis. We will discuss the differences between them, and their strong and weak points.

Lectures 8 and 9. Model assessment: We will introduce several approaches for evaluating the accuracy of models, including cross-validation and bootstrapping.

Lecture 10. Tree-based methods: Focussing on decision trees, we will introduce tree-based methods for regression and classification. After that, we will cover more advanced methods, such as Bagging, Random Forests, and Boosting.

Lecture 11. Support Vector Machines (SVMs): We will introduce maximal margin classifiers, and use this as a base to exploring SVMs.

Continue to Basic Data Analysis-E2(2)↓↓↓

Basic Data Analysis-E2(2)

Lecture 12: Neural networks and deep learning.

Lecture 13: Unsupervised learning: Introduction to unsupervised learning problems. We will introduce Principal Component Analysis, K-means clustering, and hierarchical clustering.

Lecture 14. Review of course material.

＜Final examination＞

Lecture 15. Feedback

[Course requirements]

The course is intended for students who have a basic understanding of statistics.

[Evaluation methods and policy]

Grading will be based on a final examination (50%) and small quizzes (50%).

[Textbooks]

James, Witten, Hastie and Tibshirani 『An Introduction to Statistical Learning: with Applications in R』（Springer）ISBN:978-1071614174（The course lectures will follow the content of this textbook (Edition 2). Please note that this textbook is also freely (legally) available for download at https://www.statlearning.com.）

[Study outside of class (preparation and review)]

The course will follow the textbook. I will specify the sections that will be covered and the sections that will be skipped.

[Other information (office hours, etc.)]

No fixed office hours. Students are requested to make appointments directly or by email.