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	Tue.2	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, and principal component analysis. Practical applications will be demonstrated using the statistical programming language R.

[Course objectives]

Students will learn about basic concepts in data analysis and machine 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)]

In principle, the course will be offered according to the following plan. However, depending on the progress of the course the order or the number of lectures for each topic may change.
Lectures 1 and 2. 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 3 and 4. 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 5 and 6. 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.
Lecture 7: Mid-term exam.
Lecture 8. Model assessment: We will introduce several approaches for evaluating the accuracy of models, including cross-validation and bootstrapping.
Lectures 9 and 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.
Lectures 12 and 13: Unsupervised learning: Introduction to unsupervised learning problems. We will

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

Basic Data Analysis-E2(2)

introduce Principal Component Analysis, K-means clustering, and hierarchical clustering.
Lecture 14. Review of course material.
Lecture 15. Final examination
Lecture 16. Feedback

[Course requirements]

The course is intended for students who have at least a basic understanding of statistics. Programming experience is useful but not required.

[Evaluation methods and policy]

Grading: attendance and active participation (10%), mid-term exam (30%), quizzes/assignment (20%), and final exam (40%)

[Textbooks]

James, Witten, Hastie and Tibshirani 『An Introduction to Statistical Learning: with Applications in R』 （Springer）ISBN:978-1461471370（The course lectures will follow the content of this textbook. Sections of the book to read in preparation of each class will be announced. This textbooks contains theoretical parts as well as practical exercises. Please note that this textbook is also freely (legally) available for download at http://faculty.marshall.usc.edu/gareth-james/ISL/）

[Study outside of class (preparation and review)]

The course will follow a textbook. At the end of each lecture I will specify the sections to read before the next lecture.

[Other information (office hours, etc.)]

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

[Essential courses]