

科目ナンバリング		U-LAS11 10009 LE55							
授業科目名 <英訳>		Basic Data Analysis-E2 Basic Data Analysis-E2				担当者所属 職名・氏名		医生物学研究所 准教授 VANDENBON , Alexis	
群	自然科学科目群			分野(分類)	データ科学(基礎)			使用言語	英語
旧群	B群	単位数	2単位	週コマ数	1コマ	授業形態	講義（対面授業科目）		
開講年度・ 開講期	2024・前期		曜時限	火2		配当学年	全回生	対象学生	全学向
[授業の概要・目的]									
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.									
[到達目標]									
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.									
[授業計画と内容]									
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 and 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 introduce Principal Component Analysis, K-means clustering, and hierarchical clustering.									
Lecture 14. Review of course material.									
-----Basic Data Analysis-E2(2)へ続く-----									

## Basic Data Analysis-E2(2)

< Final examination >

Lecture 15. Feedback

### 【履修要件】

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

### 【成績評価の方法・観点】

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

### 【教科書】

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 (Edition 1). Sections of the book to read in preparation of each class will be announced. This textbook contains theoretical parts as well as practical exercises. Please note that this textbook is also freely (legally) available for download at <https://www.statlearning.com>. )

### 【授業外学修（予習・復習）等】

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

### 【その他（オフィスアワー等）】

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