

# Data Science

Satoshi Nakamura, Kimito Funatsu,  
Miyao Tomoyuki, Naoaki Ono,  
Hiroki Tanaka, Ai Muto, Katsuyuki Kunida

# Introduction

## ▶ Required subjects of Data Science Program

### – General Subjects (共通科目)

- "Professional Communication I, II, Academic Discussion, Research Presentation, Research Writing, Advanced Research Writing".
- "Technology and Professional Ethics" (技術と倫理)

### – Science and Technology Subjects (序論, 基礎科目, 専門科目)

- "Introduction to Interdisciplinary Research Program" (融合プログラム序論)
- "Data Science" (データサイエンス論)
- "Data Science PBL I, II" (データサイエンスPBL I, II)

### – Research-based Subjects

- "Seminar I, II" (ゼミナールI, II)
- "Thesis Research" (修士論文研究)

For details please take a look at a student guidebook.

# “Data Science” (データサイエンス論)

	Date	Lecturer	Theme	Content
1	5/22 [4]	Tomoyuki Miyao	Basic Statistics (Miyao)	Basic terms and concepts in statistics are reviewed. We will learn descriptive statistics, inference statistics and Bayesian statistics.
2	6/5 [4]	Hiroki Tanaka	Machine Learning: Classification, Clustering (Tanaka)	We study an overview of machine learning. We study differences between supervised and unsupervised learning and basic ideas of them. (e.g. Ward clustering, hierarchical clustering, decision trees, random forests)
3	6/12 [4]	Naoaki Ono	Machine Learning II: PCA, regression, sequential data analysis, deep learning (Ono)	Introduce various popular and basic methods of data science, such as principal component analysis, regression, sequential data analysis, deep learning.
4	6/19 [4]	Naoaki Ono, Ai Muto	Applications of Information Science and Data Science in Biology (Ono, Muto)	Learn about applications of data science in bioscience, i.e., Genomics, phylogenetic analysis, genetic diagnosis, metabolomics, and other related multiomics analysis based on comprehensive, large-scale, multi-modal biological data. Also, we will introduce biological databases that support those analyses.
5	6/26 [4]	Katsuyuki Kunida	Systems Biology (Kunida)	Cell receive extracellular information such as growth factor and hormones as the input signal and drive the intracellular molecular activity and gene expression, controlling the cellular function such as proliferation and metabolism as the dynamical system. In this lecture, we will learn the modeling and data analysis of cellular system based on mathematical biology and system and control engineering.
6	7/3 [4]	Tomoyuki Miyao	Descriptors in Material and Molecular Design (Miyao)	We will learn how to describe objects (molecules and materials) numerically for applying data-science methods to materials and molecular design.
7	7/17 [4]	Kimito Funatsu ・ Tomoyuki Miyao	PCA ・ PLS (development of organic materials) and Pareto solutions (Catalyst design) (Funatsu, Miyao)	To understand methods for constructing a model of material-property relationship by overviewing property and effective parameters. Also to understand the concept of Pareto solutions through catalyst design, considering trade-off among several properties.
8	7/31 [4]	All faculty members in charge	Q&A	All faculty members participate in the Webex and answer questions about the lectures. Questions should be received in advance by e-mail, but additional questions will be accepted during the Q&A time if time allows. Write your name and questions (either in Japanese or English) in the email and send it to the faculty ML from May 22 to July 30.    ds1-ds2-faculty@is.naist.jp

# データサイエンス論

回数	日付 [時間]	担当教員	テーマ	内容
1	5/22 [4]	宮尾 知幸	基礎統計（宮尾）	統計学における基本的な用語・概念を学習する。記述統計、推測統計、ベイズ統計の考え方を例題を通して理解する。
2	6/5 [4]	田中 宏季	機械学習Ⅰ：分類、クラスタリング（田中）	データサイエンス手法の1つである機械学習について概観する。教師あり学習、教師なし学習の違いとそれらの基本的手法について学ぶ。（例、ワード分析、階層的クラスタリング、決定木、ランダムフォレスト）
3	6/12 [4]	小野 直亮	機械学習Ⅱ：PCA、回帰、系列データ学習、深層学習（小野）	データサイエンスで汎用的に使われる機械学習の手法の基本として、主成分分析、回帰モデル、系列データ学習、深層学習などについて説明する。
4	6/19 [4]	小野 直亮、 武藤 愛	バイオサイエンスにおけるビッグデータ解析とデータサイエンス（小野、武藤）	バイオサイエンスにおけるデータサイエンスの応用について学ぶ。ゲノム解析や系統解析、遺伝子診断、代謝解析やその他のマルチオミクス解析など網羅的かつ大規模、多元的な生物データに基づく解析と、それらを支えるバイオデータベースについて紹介する。
5	6/26 [4]	国田 勝行	システムバイオロジー（国田）	細胞は増殖因子やホルモンなどの細胞外情報を入力として受け取り、細胞内の分子活性や遺伝子発現を駆動することにより、増殖や代謝などの細胞機能を動的システムとして制御している。本講義では、数理生物学やシステム制御工学の解析手法を用いた細胞システムのモデリングとデータ解析について学ぶ。
6	7/3 [4]	宮尾 知幸	マテリアル・分子設計における記述子（宮尾）	材料設計、分子設計において、データ科学手法を適用するために必要な記述子（数値化）について、分子構造のコンピュータ上での表現を含め、事例を通して学ぶ。
7	7/17 [4]	船津 公人・宮尾知幸	PCAとPLS（有機材料の開発）、パレート最適解（触媒設計）（船津、宮尾）	材料の特性に影響する種々のパラメータを俯瞰し、材料とその特性との相関関係をモデル化する手法を実例を通して理解する。および材料物性間のトレードオフを対象としたパレート最適解探索を触媒探索を通して理解する。
8	7/31 [4]	担当教員全員	Q&A	授業内容に関するQ&Aタイムとし、担当教員全員がWebexに参加して質問に回答する。 質問は基本的にはメールでの事前受付とするが、時間に余裕があれば当日も受け付ける。 <質問する場合> メールに名前と質問事項（日英どちらでも可）を書いて、教員ML宛に送信。（ds1-ds2-faculty@is.naist.jp） 受付期間は5/22～7/30とする。

# How to get credit

- Assignments and Final report.
  - Weights: Assignments(50%) and Final report (50%)
- The URL for the report upload can be found in Syllabus.

## Lecture related URL

Content
<a href="#">The DropBox to submit your reports, for the 0th class (valid 05/17 - 05/22)</a>

- Password will be announced later.