

データ解析入門 -大規模データ時代の基礎-

| | | |
|--|---|-----------------------------------|
| 単位数 | ナンバリングコード | |
| 2 | DIF219 | |
|  | 教員名 | 内山 俊郎 |
| | 専門 | 情報システムの設計、データ解析（機械学習、データマイニング、推薦） |
| | 出身校等 | 東京工業大学 博士（工学） |
| | 現職 | 北海道情報大学 経営情報学部 教授 |
| 授業形態 | | |
| 前期印刷授業・後期印刷授業・夏期面接授業 | | |
| 授業範囲 | 試験範囲 | |
| 教科書の1章から5章まで | 教科書の1章から5章まで | |
| | 【印刷授業：試験時参照許可物】 一切自由 ※ただしWebページ（通信教育部POLITEを除く）と生成系AIの参照は不可とする。 【面接授業：試験時持ち込み許可物】 一切不可 | |
| 科目の概要 | | |
| データ解析は、「統計モデル（確率モデル）を用いてデータの背後にある特徴や関係性を明らかにすること」といえます。観測データをモデルにあてはめる時、最適なあてはめを探索するためにコンピュータを活用します。本講義では、データの解析例を通して、解析の具体的なイメージを持ち、抽象化されたモデル表現との対応を理解し、プログラムにより実際に解析するプロセスする流れの理解を目指します。 | | |
| 授業における学修の到達目標 | | |
| この講義は、データ解析の基礎を理解することを目標とする 具体的には、 平方和最小基準のクラスタリングと2乗誤差最小基準のベクトル量子化について、 上記を実現するための、k-meansと競合学習アルゴリズム、 パーセプトロンの学習則、 ナイーブベイズ分類器と、その基礎となる確率論と確率モデル、 上記のモデルパラメータ推定、 を例として理解することを目標とする | | |
| 講義の方針・計画 | | |
| 教科書の2章「データ解析のための基本操作」に目を通してから3章以降に進むことをお勧めしますが、3章「クラスタリング」と5章「確率論と確率モデル」の理解を優先してください。これらの章では、データ解析に関わる有名かつ重要なアルゴリズムを示し、解説します。通教用POLITEに用意された印刷授業用の「情報システムの設計」ページにおいて、全15回の講義資料を用意しています。必要な情報やリンクもありますので、同ページを見てください。 | | |
| 第1回：さまざまなデータ解析（本講義の位置づけの紹介） 第2回：データ解析のための基本操作（演習環境Linuxのファイル構造と基本操作、エディタ） 第3回：データ解析のための基本操作（データの読み込み、読みだしのプログラミング例） | | |

講義の方針・計画

- 第4回：データ解析のための基本操作（データの可視化、グラフ作成の演習）
第5回：データ解析のための基本操作（検定についての考え方の理解）
第6回：クラスタリング（平方和最小基準クラスタリング、手動や乱数によるクラスタリング演習）
第7回：クラスタリング（目的関数「平方和」の計算、k-means法によるクラスタリング）
第8回：クラスタリング（ベクトル量子化との対比）
第9回：クラスタリング（競合学習によるクラスタリング）
第10回：識別関数（パーセプトロンによる分類器の学習）
第11回：識別関数（さまざまな識別関数について）
第12回：確率論と確率モデル（確率論。事象と確率、条件付き確率とベイズの定理）
第13回：確率論と確率モデル（確率モデル。ナイーブベイズ分類器）
第14回：確率論と確率モデル（観測された現象とモデルの関係）
第15回：確率論と確率モデル（平方和最小基準クラスタリングと確率モデル）

準備学習

授業範囲の内容の他に、教科書の内容全体を2単位で90時間かけて学習することを目安としています。
わからない用語や内容は、参考文献等で検索することが準備学習として必要になります。

予習について（1回2時間）

印刷授業は、POLITEにある講義資料および教科書の該当部分を見て予習してください。
わからない用語や内容は、参考文献等で検索することが準備学習として必要になります。

復習について（2時間）

教科書の問題について、自力で何も見ずに解けるか挑戦すること。
特に、間違えたり、正解を見ながら解いた場合は、今一度該当箇所を理解し、その後で、自力で解くこと。

課題(試験やレポート等)に対するフィードバック方法

提出されたレポートについて、正解あるいは不正解に応じてコメントを付与することで、フィードバックを行う。

成績評価の方法およびその基準

科目試験による評価100%

教科書

書名：わかりやすいデータ解析入門-C++による演習-（第3版）
著者名：内山俊郎
発行所：ムイスリ出版
ISBN：9784896413397

参考書

なし

その他

受講にあたっては、 Σ 記号などの数学知識、for文と配列などのプログラム知識、確率のこと、などを知っていると、助けになると思います。

試験期間

シラバス検索画面トップページ (<https://syllabus-tsushin.do-johodai.ac.jp/>) 下部の「2026学年暦」を参照

学習プリント

なし

教職科目

高校数学5の1（必修）、中学数学5の1（必修）、高校数学6の4、中学数学6の4

関連受講科目

なし

担当教員の実務経験

2006年から2012年度まで企業の研究所に所属し、文書データ解析、レコメンダの研究を行うとともに、これら研究の成果を関連事業会社において利用可能なものとする実用化（プロダクト化）を行い、事業会社への技術支援を行った。これら実務において、実用化プロジェクトに関わり、これらの技術を現場において適用する過程を見てきた。これら実務経験を生かす形で、講義の中で示す技術の実応用例などを学生に伝え、技術の先にある現場の様子が想像できるような教育を実施する。