科目ナンバリング	U-LAS11 20003 SJ55						
授業科目名 <英訳>	データ分析演習Ⅰ Data Analysis Practice I			担当者所属 職名・氏名	非常勤講師 木村　真之		
群	自然科学科目群	分野(分類)	データ科学(発展)		使用言語	日本語	
旧群	B群	単位数	2単位	週コマ数	1コマ	授業形態	演習（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]

今日では、コンピュータやネットワーク、様々なセンサなどの技術の進歩により、日々膨大なデータが蓄積されるようになった。これらのデータの活用への期待は大きく、データを適切に分析し、その結果から適切な判断を下すことが重要である。
「データ分析演習I」は、ICT（情報通信技術）の進展とビッグデータ、さらにデータ表現の基礎等を確認したうえで、原則としてプログラミング言語を用いた経験のない学生を対象として、データ解析の基礎を習得する実践科目である。
本講義は、文部科学省のモデルカリキュラム（応用基礎レベル）の内、データサイエンス基礎とデータエンジニアリング基礎、AI基礎をバランスよくカバーする形で構成されている。
具体的には、データを扱うために必要なプログラミングスキル、データの取得・整形、種々の可視化手法、機械学習の基礎について学ぶ。まず、データ科学の分野で広く用いられているPythonの導入法や基礎文法を修得し、データの取得方法や整形、可視化など、データ科学に必要不可欠な技術を学ぶ。続いて回帰分析や教師あり/教師なし機械学習の基礎理論と実装法を修得する。具体的には、重回帰、ロジスティック回帰、Ridge、Lasso、サポートベクターマシン、ランダムフォレストなどの決定木、ニューラルネットワーク、k近傍法、k平均法、多様体学習、主成分分析などを学ぶ。
本講義の単位(2単位)を修得することで、文部科学省が定める数理・データサイエンス・AI教育プログラム応用基礎レベル(MDASH Advanced Literacy)修了証の取得が可能である。修了証取得の手続きについては、講義内で担当教員より指示がある。

[到達目標]

1. Python 言語等を用いたデータの取り扱いを理解し，統計処理や可視化を行うことができる
2. 機械学習の各種手法についてアルゴリズムの基本を理解し、Python等を用いて実装できる
3. 実データに対して解析手法を適切に選択し，統計処理・可視化・機械学習などを用いた分析が実行できる

[授業計画と内容]

フィードバックを含め全15回の授業で、Python言語を用いてデータ分析を実践する。本演習の前半ではデータ科学分野で広く使用されているPython言語の基礎プログラミングと外部モジュールを用いたデータの整形や可視化について、後半では機械学習モジュールを用いた教師あり/教師なし機械学習の実装と演習を予定している。
なお開講にあたっては、受講生の所属するキャンパスの配置や受講形態にも配慮し、一部メディア授業も取り入れる。
第1回 データサイエンス概観とPythonの導入（データ駆動型社会・データサイエンス活用事例、ビッグデータ、プライバシー保護、個人情報の取り扱いの概説を含む）
第2-4回 Pythonプログラミング入門
第5回 NumPy入門
第6回 pandas入門
第7回 データの入出力・整形
第8回 データの可視化
第9回 時系列データの解析

データ分析演習Ⅰ(2)へ続く↓↓↓

データ分析演習Ⅰ(2)

第10回 AI・機械学習の基礎
第11回 回帰分析 (重回帰、 ロジスティック回帰、 Ridge、 Lasso)
第12回 教師あり機械学習 (サポートベクターマシン、ランダムフォレストなどの決定木、ニューラルネットワーク、k近傍法)
第13回 教師なし機械学習 (k平均法、多様体学習、主成分分析)
第14回 テキストデータの解析
第15回 レポート解説およびフィードバック

なお、講義の進度・文科省のモデルカリキュラム等を反映して内容順序の変更や省略・追加を行うことがある。

[履修要件]

「統計入門」あるいは同等の科目を履修していることがのぞましい。
また、共通教育における線形代数の基礎知識(ベクトルや行列など)や高校レベルの数学の知識を必要とする。

[成績評価の方法・観点]

講義中に与える課題とレポートの内容によって到達目標への到達度を評価する。詳細は授業中に指示する。

[教科書]

使用しない
講義資料のPDF版を配布する。

[参考書等]

（参考書）
Wes McKinney 『Pythonによるデータ分析入門 第2版 ― NumPy、pandasを使ったデータ処理―』（オライリージャパン, 2019）ISBN:978-4-87311-845-1
Jake VanderPlas 『Pythonデータサイエンスハンドブック ―Jupyter、NumPy、pandas、Matplotlib、scikit-learnを使ったデータ分析、機械学習』（オライリージャパン, 2019）ISBN:978-4-87311-841-3
Bill Lubanovic 『入門 Python 3』（オライリージャパン, 2017）ISBN:978-4-87311-738-6
京都大学　データ科学イノベーション教育研究センター 『講義実録　統計入門』（現代図書, 2023）ISBN:978-4-434-31857-3

[授業外学修（予習・復習）等]

復習として，講義で解説した内容を自らプログラムを組んで実装し，様々なデータに対して適用してみることを期待する.

[その他（オフィスアワー等）]

講義中に教員との連絡方法について指示する。