

データサイエンス基礎概論

必修 2単位

城戸 隆

1. 授業の概要(ねらい)

情報が重要である現代においては、必要な情報を的確に入手し、情報を正しく分析し、問題発見・解決を的確に行うことが求められる。このために、データを正しく活用する基本的な知識・方法を身につけることは不可欠である。本コースでは、実社会において、どの進路に進むにも必要となる、基礎・基本的な情報リテラシー、考える力、データを正しく活用する力を身につけることを目的とする。

特に本コースでは、「将来、必ず役に立つデータサイエンスの知識とスキルを学びたい学生」を念頭に、体系的な知識と実践的なスキルを学ぶ場を提供する。データサイエンスは、これからもどんどん変化していくので、現在はやっている領域のみを学ぶよりも、「本質を見抜く力」と「学び続ける力」を養うことが大切である。(1)データサイエンスの歴史に沿って、統計学や機械学習などの重要な概念の変遷を体系的に学ぶこと、(2)具体例を用いて「本質を見抜く力」を磨くこと (3)オンライン教材を用いて、手を動かしながら、目的に応じて必要なスキルを学んでいく方法を身につけていくことに重点を置く。研究やビジネスに生かせる実践的なデータサイエンス入門。

2. 授業の到達目標

- ・ 現代のデータ社会に対して主体的な問題意識とパースペクティブを育むこと。
- ・ 「これからのAI時代に本当に必要なスキルは何か?」について、自らの進路にとって必要な目標を設定できること。
- ・ データを読み解くことの面白さ、分析の大切さを知る。
- ・ データサイエンスの歴史に沿って統計学や機械学習などの重要な概念の変遷を体系的に理解できる。
- ・ データサイエンス、統計解析、機械学習に関する重要語句を適切に理解することが出来る。
- ・ データの意味合いを理解し、本質を見抜く力を身につける。
- ・ データサイエンスの数理的および分析的思考、データリテラシーの実践的な基本を身につける。
- ・ データサイエンスの一連の流れを理解し、どんなスキルを身につける必要があるのか、最低限の基礎を知る。
- ・ データサイエンスの専門分野への応用について、事例や課題を理解できる。

3. 成績評価の方法および基準

授業中に行われるディスカッションやグループワークへの参加、小テスト(30%・LMS上で実施)、期末レポート(70%)によって評価します。

4. 教科書・参考文献

教科書

竹村彰通、姫野哲人、高田聖治 (編) データサイエンス入門(データサイエンス大系) 学術図書出版社

参考文献

ハンス・ロスリング et al, FACT FULNESS 10の思う巨海を乗り越え、データを基に世界を正しく見る習慣、日経BP

統計でウソをつく法—数式を使わない統計学入門 ブルーバックス

データサイエンス講義 オライリー・ジャパン

ネイト・シルバー シグナル&ノイズ 天才データアナリストの「予測学」 日経BP

revor Hastie (著), Robert Tibshirani (著), Jerome Friedman (著), 杉山 将 (翻訳) 統計的学習の基礎—データマイニング・推論・予測— 共立出版

C.M. ビショップ (著), 元田 浩 (監訳), 栗田 多喜夫 (監訳), 樋口 知之 (監訳), 松本 裕治 (監訳) パターン認識と機械学習 上下 丸善出版

黒滝 紘生 (翻訳), 河野 慎 (翻訳), 味曾野 雅史 (翻訳), 保住 純 (翻訳), 野中 尚輝 (翻訳), 冨山 翔司 (翻訳) 深層学習 KADOKAWA

David M. Diez, Mine Cetinkaya-Rundel, Christopher D. Barr 著, 国友直人, 小暮厚之, 吉田端 (訳) データ分析のための統計学入門 一般財団法人 日本統計協会

5. 準備学習の内容

- ・ コンピュータの基本操作(起動終了、文字入力、ファイル操作など)を習得していること。
- ・ 高校1-2年程度の数学の知識、スキル、Excelで基本的なことができる程度の素養。
- ・ 専門領域に関する具体的な課題や問題意識をもっていることが望ましい。(必須ではない。)

6. その他履修上の注意事項

7. 授業内容

- 【第1回】 時代背景と本コースの位置づけ、事前アンケート:本授業のゴールと現代社会におけるデータサイエンスの役割について理解すること。
- 【第2回】 データサイエンスの歴史と概念の変遷:データサイエンスの歴史を統計学、人工知能の分野の発展にそって概念の変遷を理解すること。
- 【第3回】 データの取得と管理(実習):インターネット上で利用可能なデータをダウンロードし、Excelもしくはテキストエディタに読みこみ確認すること。
利用可能なオンライン教材を用いた学習方法を学ぶこと。
- 【第4回】 データ分析の基礎 (1): ヒストグラム、箱ひげ図、平均、分散
- 【第5回】 データ分析の基礎 (2): 散布図、相関係数、回帰直線
- 【第6回】 Excel実習: インターネット上で利用可能なデータをExcelを用いて分析すること。
- 【第7回】 データ分析で注意すべき点: 相関・因果、観察・実験研究、標本調査
- 【第8回】 コンピュータを用いた分析: 統計解析ソフトR、プログラミング言語 Python。
- 【第9回】 データサイエンスの手法 (1): ベイズ推論、アソシエーション分析、クラスタリング
- 【第10回】 データサイエンスの手法 (2): 決定木、ニューラルネットワーク
- 【第11回】 データサイエンスの手法 (3): 機械学習、深層学習
- 【第12回】 R、(またはPython)実習:インターネット上で利用可能なデータをRもしくはPythonを用いて分析すること。
- 【第13回】 データサイエンスの応用事例(1): マーケティング、金融、品質管理

【第14回】 データサイエンスの応用事例(2): 画像処理、音声処理、医学

【第15回】 より進んだ学習のために: オンライン学習教材やKaggleなどコンテストの紹介

※各回の内容には、授業の進行ペースや皆さんの関心の傾きに依じて調整を加えることがあります。