

科目名	データ駆動科学概論と実践
講義題目(テーマ)	データ駆動科学概論と実践
担当教員	熊本大学 赤井一郎 計算科学振興財団 伊藤 聡
年度・学期	2021年 集中
単位数	2

学修成果とその割合	
1.高度な専門的知識・技能及び研究力	70
2.学際的領域を理解できる深奥な教養力	30
3.グローバルな視野と行動力	0
4.地域社会を牽引するリーダー力	0
その他	0

使用言語	「日本語」による授業
教科書・資料の言語	「日本語」のテキスト
実務経験を活かした授業	非該当
授業の形態	講義
対面・遠隔の別	遠隔形式
授業の方法	Zoomを用いた遠隔授業と、オンデマンド受講

授業の目的	データ駆動型社会を担う人材育成プログラムの導入。情報統合型研究開発実践例の紹介。Python/Rの実践演習
授業の概要	以下の事柄について講義を行う。 (1).本プログラムについて、(2).復習・誤差論、(3).ベイズ推定の基礎、(4).スパースモデリングの基礎、(5).データ駆動科学の適用事例、(6).pymc3を用いたベイズ推定、(7).scikit-learnを用いた線形モデル解析、(8)情報量基準と交差検定、(9)情報統合型物質・材料開発事例

学修目標	
A水準 (到達すれば「優」に相当)	(1).本プログラムについて、(2).復習・誤差論、(3).ベイズ推定の基礎、(4).スパースモデリングの基礎、(5).データ駆動科学の適用事例、(6).pymc3を用いたベイズ推定、(7).scikit-learnを用いた線形モデル解析、(8)情報量基準と交差検定、(9)情報統合型物質・材料開発事例について、十分に理解し、講義内容を他人に正確に説明できる。
C水準 (到達すれば「可」に相当)	(1).本プログラムについて、(2).復習・誤差論、(3).ベイズ推定の基礎、(4).スパースモデリングの基礎、(5).データ駆動科学の適用事例、(6).pymc3を用いたベイズ推定、(7).scikit-learnを用いた線形モデル解析、(8)情報量基準と交差検定、(9)情報統合型物質・材料開発事例について、概ね理解し、講義内容の要点をまとめることができる。
評価方法・基準	Moodleで提出されたレポートの到達度(100%) から評価する。

各回の授業内容		
回	授業テーマ (5文字以上100文字以内)	内容概略(10文字以上200文字以内)

1	本プログラムについて	データ駆動型社会を担う人材育成プログラムの趣旨、概要について解説する。
2	Python/R環境の整備	データ駆動型社会を担う人材育成プログラムで利用するPythonとR言語の環境整備について解説する。
3	復習・誤差論	理工系で行われる計測データ解析で用いる誤差論を復習し、その問題点について解説する。
4	ベイズ推定の基礎1	因果律の原因と結果の同時確率にベイズの定理を適用する、データ駆動科学で用いるベイズ推定について解説する。
5	ベイズ推定の基礎2	データ駆動科学でベイズ推定を行う際に用いるマルコフ連鎖モンテカルロ法について解説する。
6	スパースモデリングの基礎1	データ解析における不良設定問題とスパースモデリングの定式化について解説する。
7	スパースモデリングの基礎2	多項式回帰におけるスパースモデリングと過学習について解説する。
8	データ駆動科学の適用事例	物性研究にベイズ推定やスパースモデリングを適用した実践列とメリットについて解説する。
9	pymc3を用いたベイズ推定1	pymc3の概要と、Google Colaboratry / Anacondaディストリビューションにおけるpymc3環境の整備について解説する。
10	pymc3を用いたベイズ推定2	pymc3による線形回帰 / スペクトル分解解析と事後確率分布の評価法について解説する。
11	scikit-learnを用いた線形モデル解析	scikit-learnを用いた線形モデル解析法、LASSO法、多項式回帰・フーリエ変換解析におけるLASSO解析について解説する。
12	情報量基準と交差検定	データの正則化と、正則化の規準となる情報量規準や、交差検定法について解説する。
13	情報統合型物質・材料開発事例1	データ駆動型科学による蓄電池材料や磁石材料などの開発事例を通して、実際の材料開発におけるデータ駆動型科学アプローチの活用のあるあり方を紹介する。
14	情報統合型物質・材料開発事例2	データ駆動型科学による材料の熱物性設計等を通して、物性の理解への展開について紹介する。
15	情報統合型物質・材料開発事例3	データ駆動型科学を行うために必要なデータベースの動向と、データ処理について解説する。

授業外学修時間の目安	本科目は、90時間の学修が必要な内容で構成されている。授業は30時間分（2h×15コマ）となるため、60時間分相当の事前・事後学修（課題等含む）が、授業の理解を深めるために必要となる。
------------	--

キーワード	データ駆動科学、Python、R言語、誤差論、ベイズ推定、マルコフ連鎖モンテカルロ法、スパースモデリング、pymc3、scikit-learn、線形モデル、LASSO法、正則化、情報量規準、交差検定
テキスト	授業の際に資料を配布する。

参考文献	<ol style="list-style-type: none"> <li>1. 植村誠, 天文屋のための HOW TO スパースモデリング, 連立1次方程式を解く, <a href="http://home.hiroshima-u.ac.jp/uemuram/How2SM/01_demo.pdf">http://home.hiroshima-u.ac.jp/uemuram/How2SM/01_demo.pdf</a></li> <li>2. 青木繁伸, Rによる統計解析, オーム社, 2009</li> <li>3. 村井潤一郎, Rによる統計解析: ごく初歩の操作から統計解析の導入まで, 北大路書房, 2013</li> <li>4. K. P. Murphy, Machine Learning: A Probabilistic Perspective (Adaptive Computation and Machine Learning series), The MIT press, 2012</li> <li>5. John V. Guttag, 久保幹雄監訳, Python言語によるプログラムイントロダクション第2版: データサイエンスとアプリケーション, 近代科学社, 2017</li> <li>6. 池内孝啓, 片柳薫子, 岩尾エマはるか, @driller, PythonユーザーのためのJupyter[実践]入門, 技術評論社, 2017</li> <li>7. 吉田拓真、尾原 颯, 現場で使える! NumPyデータ処理入門 機械学習・データサイエンスで役立つ高速処理手法, 翔泳社, 2018</li> <li>8. 岩波データサイエンス刊行委員会, 岩波データサイエンス Vol.1, ベイズ推論とMCMCのフリーソフト, 岩波書店, 2015</li> <li>9. Osvaldo Martin (原著), 金子武久 (翻訳), Pythonによるベイズ統計モデリング: PyMCでのデータ分析実践ガイド, 共立出版, 2018</li> <li>10. Cameron Davidson-Pilon (原著), 玉木徹 (訳), Pythonで体験するベイズ推論 PyMCによるMCMC入門, 森北出版, 2017</li> </ol>
オフィス・アワー	データ駆動型社会を担う人材育成プログラム事務室に連絡をとること
担当教員への連絡方法	データ駆動型社会を担う人材育成プログラム事務室に連絡をとること
担当教員からのメッセージ	<p>本科目は、「データ駆動型社会を担う人材育成プログラム」の修了認定を希望する場合、必修科目に設定されています。そのため、この科目が、皆さんの研究・開発において、機械学習の適用のきっかけになることを期待して、機械学習のファストトライが可能となるように、ベイズ推定やスパースモデリングの概要の理解と、入門として実践を行います。また、それらを適用した実践例についても紹介します。</p>