

専攻名	創造技術専攻	必修・選択	選択	単位	2	学期	1Q
科目群	デジタル技術科目群	科目名 (英文表記)	データサイエンス特論 Data Science			教員名	橋本 洋志 大久保 友幸

概要	データサイエンスは、「データを科学的に扱う」学問分野である。本講義では、様々なデータの収集、可視化、解析、マイニング、評価、などの手法に関する知識とスキルを学ぶ。この際、統計学、コンピュータ科学、システム工学論、信号処理論などの観点から、データに対して仮説発見、仮設検証が行えるよう、客観的・定量的評価を行うことのできる資質を身に付ける。さらに、データサイエンスの実習を通して、この体系の理解を実践的に学ぶ。			
目的・狙い	各種データ（自然現象、社会システム、物理システム、経済システムなど）の持つ有用な特徴を見出すための仮説を立てる能力、およびそれを実証するスキルを適切に使いこなせる能力を修得することを目的とする。スキルを実践するためのツールは Python（スクリプト言語、世界的に定評があり、IT 業界、科学・数理計算で広く使われている）とそのパッケージ（SciPy, pandas, scikit-learn, OpenCV など）を用いる。ただし、プログラミングはほとんど行わず、それを使った得られる結果を適切に評価できる能力を身に付けることを主とする。 この目的のもとで、本講義の狙いは、次にある。 1.データが有する意味や特徴を、その背景から体系的に推定し、意味ある仮説を立てられる能力を修得する。 2.他者にわかるようなデータの可視化技法を修得する。 3.データの分析手法を修得する。 4.分析結果の評価に関して、その表現法と説明スキルを修得する。			
前提知識 （履修条件）	自立した学習態度が必要である。このため、プログラミングは実質的に行わないが（プログラムは提供する）、わずかでもいいのでプログラミングの経験があることが望ましい。Python の文法説明はまず行わないので、これについては自学習に委ねる。 コンピュータ操作について基礎的なスキル（ファイル操作、Excel と Word 操作など）があること。基本的な数式は幾つか出現する。数式を用いた解析は行わないが、記号や変数を読むこと。 自学習のため、テキスト・プログラムを提供するので USB メモリ（USB2.0, 1GB 以上）を初回講義から持参すること。			
到達目標	上位到達目標			
	・データ解析から仮説発見、仮設検証を行える。 ・データ解析結果から、科学的、社会的、経済的に意味ある説明ができるようになる。			
	最低到達目標			
	・与えられたデータに対して適切な解析手法を適用、実行できるようになる。 ・データの可視化が行える。			
授業の形態	形態		実施	特徴・留意点
	録画・対面混合授業		○	授業参加が原則であるが、仕事上の欠席は個別対応のため事前相談されたし。
	対面 授 業	講義（双方向）	○	講義の後に直ちに演習・実習を行う。
		実習・演習（個人）	○	質問は随時受け付ける。
		実習・演習（グループ）	—	
	サテライト開講授業		—	
その他		—		
授業外の学習	毎回、予習は必須である。第 1 回目講義に臨む以前に次のテキストを熟読しておくこと。 http://hhlab.org/ ⇒（左欄）“担当授業” ⇒ “データサイエンス特論” ⇒ “授業の進め方”をクリックして現れるページの説明文に従い、第 1 回目講義の予習と準備を行っておくこと。さらに、 ・データサイエンススキルを獲得するための反復練習を自習で行う。 ・データを社会でどのように利用・活用し、どのような価値があるかの背景調査、文献調査を行う。			
授業の内容	講義は配布する資料を使用して行う。授業前半で Python を用いた分析の実習を行い、分析結果に対する評価について学ぶ。また、本学の授業支援システム(LMS)を利用した資料配布、レポート提出を行う。なお、仕事などの都合でやむを得ず出席できない場合は、事前にメールで申請し、教員の指示に従うこと。			

	回数	内容	サテライト 開講	対面/録画
授業の計画	第 1 回	講義ガイダンス・ソフトウェア使い方とレポート作成作法 データサイエンスの意義、データの収集、入出力、変換などに関する基本操作を習得する。さらに、レポート作成作法について説明する。	—	対面
	第 2 回	データの取得と操作 様々なデータを収集の仕方、この加工の仕方を学ぶ。さらに、データ操作として pandas のデータベース操作やクロス集計について学ぶ。	—	対面
	第 3 回	データの視覚化 様々なデータに適するプロット技法を学ぶ。これを通して、視覚化の意義について考える。	—	対面
	第 4 回	統計分析の基礎 初めに確率論の基礎としての基本統計量を学ぶ。次に、区間推定、仮説検定などについて学ぶ。	—	対面
	第 5 回	回帰分析 1 単回帰分析、多項式モデルなどについて学ぶ。	—	対面
	第 6 回	回帰分析 2 前回の回帰分析では表現しきれないデータに対する回帰分析法として、重回帰分析や一般化線形モデルを用いた分析を学ぶ。	—	対面
	第 7 回	パターン認識 1 データの識別手法として、線形判別法などについて学ぶ。	—	対面
	第 8 回	パターン認識 2 データの識別手法として、クラスタリングの手法、その表現法などについて学ぶ。	—	対面
	第 9 回	動的システムの波形分析 微分方程式で表される動的システムの 1 次応答、2 次応答の見方を説明して、その分析に関する知識とスキルを学ぶ。	—	対面
	第 10 回	時系列データの推定と予測 時系列データの特徴を知ったうえで、統計的な推定論を通した予測モデルの求め方を学ぶ。	—	対面
	第 11 回	周波数分析 データが有する特徴量を周波数成分やパワースペクトルで表されることを学んだ後に、周波数領域でのデータ処理について学ぶ。	—	対面
	第 12 回	画像処理 画像処理の基礎（2 値化、エッジ処理）、動画の取得・処理などについて学ぶ。	—	対面
	第 13 回	メディア情報処理 マイク・カメラから得られる音声情報や画像情報に対するデータ処理法と応用例を学ぶ。	—	対面
	第 14 回	位置情報処理 測位方法の基礎を知った上で、得られる位置情報・空間情報に対するデータ処理の応用例を学ぶ。	—	対面
	第 15 回	センサデータ処理 製造装置や生体などに設置するセンサから得られるセンサデータに対する処理法や IoT などの応用例を学ぶ。	—	対面
	試験	データサイエンスに関する基本的知識、及び分析結果に対する評価法に関する内容などの試験を実施する。	—	対面
成績評価	次の 2 つのポイントで評価する(合計 100 点満点) ・レポート評価 80 点：上数回のレポートで評価。 ・最終試験 20 点：講義終了後に、これまで獲得した知識・スキルを総合的に評価 上記、レポート全ての提出と試験を受けることが単位認定の必要条件である。			
教科書・教材	教科書・教材とも、配布資料及びサーバー上にアップしたコンテンツとして提供する。			
参考図書	・橋本、牧野：データサイエンス教本、オーム社 ・東京大学教養学部統計学教室：統計学入門 (基礎統計学 I)、東京大学出版会 ・R で学ぶデータサイエンスシリーズ、20 巻、共立出版 ・Wes McKinney: Python によるデータ分析入門 NumPy、pandas を使ったデータ処理、オライリージャパン			