

<<最終更新日：2025年02月26日>>

基本情報

時間割コード	290724
開講区分(開講学期)	集中
曜日・時間	他
開講科目名	データ科学特論 I
開講科目名(英)	Advanced Course in Data Sciences I
定員	999
教室	
ナンバリング	29CHEN5M204,29SSAI5M204,29MASC5M204,29MSS5M204
授業形態	講義科目
単位数	2.0
年次	1,2年
担当教員	鈴木 讓,内田 雅之
メディア授業科目	該当（学部学生がメディア授業科目を卒業要件に算入できるのは60単位が上限です。）

※メディア授業科目について
授業回数の半数以上を、多様なメディアを高度に利用して教室等以外の場所で行う授業を「メディア授業科目」としています。
学部学生が「メディア授業科目」を卒業要件に算入できるのは60単位が上限です。
なお、非該当の場合であっても、メディアを利用した授業を実施する場合があります。

詳細情報

授業サブタイトル	スパース推定とそのモデリング	
開講言語	日本語	
学習方法	聴講・視聴：講義・教材・実演を視聴して学ぶ（例：講義の対面受講、オンデマンド教材視聴）	
授業の目的と概要	本講義では、スパース推定の理論と応用を学ぶ。スパース推定は、高次元データ解析において、不要な変数を選択し、モデルの解釈性と予測精度を向上させるために不可欠な手法である。本講義では、線形回帰、ロジスティック回帰、グループLasso、グラフィカルモデルなどの基本的なスパース推定手法を数学的に理解するとともに、Pythonを用いた実装を通じて実践的なスキルを身につける。 本講義は、以下の3つの側面に焦点を当てる： 1. スパース推定の理論的基礎 Lasso、Ridge、Elastic Net、グループLasso、Fused Lassoの定式化 近接勾配法、ADMM、情報量規準を用いたハイパーパラメータ選択 スパース推定における統計的性質の解析 2. スパース推定の応用と実装 グラフィカルLasso、多変量解析（因子分析、クラスタリング、行列分解） 産業応用（医療、金融、空間データ解析） Pythonを用いた実データの解析とスパース推定モデルの構築 3. 発展的手法と最新動向 スパースベイズモデリング、縮小事前分布、情報量規準に基づく変数選択 近似確率伝搬法、非線形スパース推定（HSIC Lasso） 高次元データ解析におけるスパース推定の展開 この講義を通じて、受講者はスパース推定の理論的背景を理解し、実データに適用できるスキルを習得することを目指す。	
学習目標	1	スパース推定手法を数学的に理解するとともに、Pythonを用いた実装を通じて実践的なスキルを身につける。
履修条件・受講条件	学部1年程度の統計学の知識が必要である。	
出欠席及び受講に関するルール	zoomで出席が確認できるよう、学籍番号を表示してください。	
授業計画	第1回	題目:2025 年9 月2 日(火) 13:30- スパース推定入門(1): 線形回帰とロジスティック回帰

	2 日目以降の講義の準備として、下記項目について最低限のことを説明する：線形回帰、劣微分、Lasso、Ridge、elastic ネット、λ の値の設定、線形回帰のLasso の一般化、2 値のロジスティック回帰、多値のロジスティック回帰、ポアソン回帰、生存時間解析。 担当教員：鈴木讓 (阪大) 教室：オンライン 授業時間外学習：講義に関する疑問点をまとめて、質問をする。
第2回	題目:15:10- スパース推定入門(2): グループLasso とFused Lasso 2 日目以降の講義の準備として、下記項目について最低限のことを説明する： グループ数が1 の場合、近接勾配法、グループLasso、スパースグループLasso、オーバーラップグループLasso、目的変数が複数個ある場合のグループLasso、ロジスティック回帰におけるグループLasso、一般化加法モデルにおけるグループLasso、Fused Lasso の適用事例、動的計画法によるFused Lasso、LARS、Lasso の双対問題と一般化Lasso、ADMM 担当教員：鈴木讓 (阪大) 教室：オンライン 授業時間外学習：講義に関する疑問点をまとめて、質問をする。
第3回	題目:16:50- スパース推定入門(3): グラフィカルモデル、行列分解、多変量解析 2 日目以降の講義の準備として、下記項目について最低限のことを説明する：グラフィカルモデル、グラフィカルLasso、疑似尤度を用いたグラフィカルモデルの推定、Joint グラフィカルLasso、特異値分解、Eckart-Young の定理、低ランク近似のスパースの適用、主成分分析、K-meansクラスタリング、凸クラスタリング。 担当教員：鈴木讓 (阪大) 教室：オンライン 授業時間外学習：講義に関する疑問点をまとめて、質問をする。
第4回	題目:9 月3 日(水)13:30- スパース因子分析 因子分析モデルは、多変量データの相関構造から背後にある潜在因子を推定するモデルである。このモデルに含まれる主要なパラメータである因子負荷行列は、通常、2 段階推定法によって求められる。具体的には、まず、最尤法などにより推定を行い、その後、因子回転を施すことで解釈しやすい因子負荷行列を得る。本講義では、この2 段階推定を一般化したスパース因子分析について解説する。とくに、非凸ペナルティであるMinimax Concave Penalty (MCP)や因子負荷行列の各行ベクトルの要素の積から構成されるProduct-based Elastic Net (Prenet)を用いたスパース因子分析を紹介する。さらに、それらを用いた事例についても説明する。 担当教員：廣瀬慧 (九大) 教室：オンライン 授業時間外学習：講義に関する疑問点をまとめて、質問をする。
第5回	題目:15:10- 周辺スクリーニングに基づく変数選択 スクリーニングとは、モデルを推定する前にあらかじめ重要な変数を選択するための手法であり、高次元データ解析において特に有用なものである。実際、Lasso などのスパース正則化法を直接適用せずに、スクリーニング後にモデルを推定することで数値的に予測精度が改善されることがしばしばある。本講義では、周辺モデルに基づくスクリーニングについて解説する。 担当教員：梅津佑太 (長崎大) 教室：オンライン 授業時間外学習：講義に関する疑問点をまとめて、質問をする。
第6回	題目:16:50- スパースグループLasso の適用例 本講義では地理的加重回帰とGMANOVA に対するスパースグループLasso の適用例を紹介する。地理的加重回帰は空間データに対する局所推定法で、観測地点ごとに重み付き推定を行う。そのため、説明変数の選択問題として地点ごとの選択と全地点で共通の選択の2 種類が考えられる。GMANOVA は例えば経時測定データに対して用いられる多変量回帰の手法である。経時トレンドに多項式を当てはめたとすると、説明変数の変数選択と多項式の次

	数選択が考えられる。これらのモデルにスパースグループLassoを適用することで、それぞれの2つの選択問題を同時に行うことができる。 担当教員：大石峰暉 (東北大) 教室：オンライン 授業時間外学習：講義に関する疑問点をまとめて、質問をする。
第7回	題目:9月4日(木) 13:30- スパース推定の活用事例と手法開発 スパース推定の活用事例と発展的手法の開発について紹介する。特に、スパース推定は産学のデータ科学の現場で広く活用されており、欠損データや定期解析といった現場特有の課題に対して新たなスパース推定法が開発されていることを示す。さらに、手法やハイパーパラメータごとに興味深い漸近的挙動が現れ、理論的に解析できることを示す。 担当教員：高田正彬 (東芝) 教室：オンライン 授業時間外学習：講義に関する疑問点をまとめて、質問をする。
第8回	題目:15:10- スパース推定における情報量規準 スパース推定における正則化パラメータの選択問題を扱う。現状では、たとえスパース推定のためのAICがシンプルな形で得られていて、かつ目的が良い予測をしようというものであっても、必ずしもそのAICは用いられておらず、つまり標準手法は定まっていないように見受けられる。本回では、そういった標準手法が定まっていないケース、正確には正規線形回帰分析でLASSOを用いるケースで、LASSOとAICを組み合わせて推定したモデルの性能評価を数値実験でおこなう。具体的には、LASSOとAICを組み合わせたとき、リッジ正則化法とAICを組み合わせたとき、最尤法と通常のAICを組み合わせてベストサブセット回帰を用いたとき、およびLASSOと交差検証法を組み合わせたときを、予測二乗誤差の評価を通じて比較する。また、その交差検証法で、データの分割の仕方によって推定結果がどのくらい違ってくるのかを、予測二乗誤差や選ばれたモデルの自由度のばらつきを数値評価することで確認する。また、時間の許す範囲で、スパース推定のためのAICを導入したり、BICなど他の情報量規準を紹介したりする。 担当教員：二宮嘉行 (統計研) 教室：オンライン 授業時間外学習：講義に関する疑問点をまとめて、質問をする。
第9回	題目:16:50- 縮小事前分布を用いたスパースベイズモデリング ベイズ的枠組みでは縮小事前分布を用いることで柔軟なスパースモデリングが可能となる。本講義では、Bayesian Lassoや馬蹄事前分布などの代表的な縮小事前分布を紹介し、それらがスパース性をどのように誘導するかを解説する。また、これらのモデルを適用するためのマルコフ連鎖モンテカルロ法について解説する。また、縮小事前分布を活用したモデル例や応用例を紹介する。 担当教員：菅澤翔之助 (慶應大) 教室：オンライン 授業時間外学習：講義に関する疑問点をまとめて、質問をする。
第10回	題目:9月5日(木)13:30- 非線形高次元特徴選択 Lasso法は、近年、医療データをはじめとする高次元データの予測や特徴の解釈に広く活用されている。しかし、Lassoは線形モデルであり、非線形な関係性を捉えるのが難しいという課題がある。本講義では、この問題を克服するための非線形拡張手法であるHilbert-Schmidt Independence Criterion (HSIC)Lassoについて、アルゴリズムの導出から実際のアプリケーションへの応用まで詳しく紹介する。 担当教員：山田誠 (OIST) 教室：オンライン 授業時間外学習：講義に関する疑問点をまとめて、質問をする。
第11回	題目:15:10- Parsimonious statistical analysis for multivariate models of high dimension Inherent to most multivariate models, the curse of dimensionality refers to the explosive increase in the number of parameters to be estimated as the dimension of the underlying vector of observations grows. Within this framework, balancing a model that is sufficiently rich in parameters to capture complex relationships while remaining parsimonious enough to avoid overfitting is a key challenge. This

		challenge can be addressed using sparsity-based methods, which focus on variable selection to recover the true signal when the model has a sparse representation. Penalization techniques are typically applied to the objective function (e.g., likelihood, least squares) during estimation to produce an estimator with many coefficients set to zero. The following points will be considered: - Examples of multivariate models suffering from the curse of dimensionality - Properties of sparsity-based estimators - Implementation issues.			
	担当教員：Poignard Benjamin Michel Claude (阪大)				
	教室：オンライン				
	授業時間外学習：講義に関する疑問点をまとめて、質問をする。				
	第12回	題目:16:50- 近似確率伝搬法によるスパース推定			
		スパース推定におけるアルゴリズムの一つである近似確率伝搬法を解説し、これを用いた研究例を紹介する。具体的には、グラフィカルモデルの一般的表現から始め、近似確率伝搬法の概要や導出方法を説明し、スパース推定への適用法を解説する。また、近似確率伝搬法を用いた漸近的な推定性能評価法や、交差検証誤差の評価法についても説明する。			
		担当教員：坂田綾香 (統数研)			
		教室：オンライン			
		授業時間外学習：講義に関する疑問点をまとめて、質問をする。			
	第13回	題目:9 月6 日(金) 8:50- The Generalized Lasso: Theory and Applications			
		In this talk, I'll review the generalized lasso framework, some key special cases revolving around trend filtering, and applications.			
		担当教員：Ryan Tibishrani (Berkeley)			
		教室：オンライン			
		授業時間外学習：講義に関する疑問点をまとめて、質問をする。			
	第14回	題目:10:30- 関数データ解析におけるスパース推定の応用			
		スパース推定は、その汎用性の高さからさまざまな統計・機械学習モデルの推定に用いられており、得られる効果も多岐にわたる。その一例として、観測個体それぞれが経時的に観測値を得たデータに対して分析を行う方法の一つである関数データ解析においても、スパース推定は利用される。本講義では特に、関数データに対する回帰分析にスパース推定を適用することで、関数データとして与えられた説明変数を選択したり、目的変数に寄与する関数データとしての説明変数の定義域を選択する方法について紹介する。			
		担当教員：松井秀俊 (滋賀大)			
		教室：オンライン			
		授業時間外学習：講義に関する疑問点をまとめて、質問をする。			
	第15回	題目:13:30- 次元削減法に対するスパース推定			
多次元データの情報をなるべく損失することなく低次元データに変換する方法は次元削減法と呼ばれ、統計学における基礎的な手法の一つである。次元削減法では多次元データの線形結合に基づき射影軸を構成するが、多次元データの中には射影軸の構成に寄与しない変数が含まれている可能性がある。そこで、射影軸の構成に必要な変数を選択するスパース推定が考えられている。本講義では、主成分分析を中心として、次元削減法に対するスパース推定の理論と方法論を解説する。					
担当教員：川野秀一 (九大)					
教室：オンライン					
授業時間外学習：講義に関する疑問点をまとめて、質問をする。					
教科書・指定教材	講義資料を配布する				
参考図書・参考教材	各講義内で紹介する				
成績評価 ※学習目標の番号にカーソルをあてると、その学習目標の全文が表示されます。	評価方法	レポート・論文			
	学習目標1	○			
	評価割合 (%)	100%			

成績評価に関する補足情報	質疑応答(20%), レポート課題(80%)により評価。各講義に対して、講師に100文字以内で質問をすることがレポート課題になる。その質問が本質をついているどうか採点基準になる。
合理的配慮	・ 本授業を受けるにあたり、障がい（難病・慢性疾患等を含む）に起因して合理的配慮を要する場合は、所属学部／研究科の障がい学生支援担当窓口（教務係／学務係／学生支援係等）やキャンパスライフ健康支援・相談センターアクセシビリティ支援室に相談してください。 ・ 詳細はこちらを参照してください。 キャンパスライフ健康支援・相談センターアクセシビリティ支援室 Website : https://acs.hacc.osaka-u.ac.jp Tel : 06-6850-6107 E-mail : campuslifekenkou-ac@office.osaka-u.ac.jp
特記事項	オンラインですので吹田の方も受講できます。
オフィスアワー	集中講義のため、オフィスアワーはなし
実務経験のある教員による授業科目	

授業担当教員

教員氏名	ふりがな	所属・職名・講座名	居室	内線	e-mail
内田 雅之	うちだ まさゆき	基礎工学研究科・社会システム数理領域・教授			uchida@sigmath.es.osaka-u.ac.jp
鈴木 譲	すずき じょう	基礎工学研究科・数理科学領域・教授			j-suzuki@sigmath.es.osaka-u.ac.jp

学生への注意書き