

## シラバス参照



科目名	データマイニング
科目名(英字)	Data Mining
ナンバリング	1GBL35
年次	2年次
単位数	2
期間	後期
担当者	坂平 文博(サカヒラ フミヒロ)

授業のねらい・概要	データマイニングとは機械学習、パターン認識の技術を用いて、多くのデータから知識を抽出する技術である。本講義ではデータマイニングの目的や意義と具体的な手法について解説する。さらに、データマイニングで使用されるツールやライブラリを用いて、データセットの分析を行う。			
CSコース				
スパイラル型教育				
授業計画		テーマ	内容・方法等	予習／復習
	第1回	ガイダンスとデータマイニングとは	本講義の概略とその意義について説明する。その後、データマイニングの概念やその適用事例について解説する。	教科書第1章を読んで専門用語を整理しておくこと(1時間)/専門用語について自分の言葉で説明できるように確認しておくこと(1時間)
	第2回	Rの基本操作とデータマイニング	統計分析ソフトRの基本的な使い方と文法、データ構造及び制御構造について解説する。	教科書第2章p.35-54までを読んで分析手法を整理しておくこと(1時間)/分析手法と操作方法を確認しておくこと(2時間)
	第3回	確認テスト データの 特徴を捉える	到達度を確認するためのテストを実施する。また、データの特徴を捉える方法として、要約統計量と相関分析について解説する。	教科書第3章p.84-114までを読んで分析手法を整理しておくこと(1時間)/分析手法と操作方法を確認しておくこと(2時間)
	第4回	回帰モデルを作るその1	データからモデルを作る方法として、単回帰モデルについて解説する。	教科書第3章p.115-134までを読んで分析手法を整理しておくこと(1時間)/分析手法と操作方法を確認し、演習課題に取り組む(4時間)
	第5回	回帰モデルを作るその2	データからモデルを作る方法として、重回帰モデルについて解説する。	教科書第3章p.134-153までを読んで分析手法を整理しておくこと(1時間)/分析手法と操作方法を確認し、演習課題に取り組む(4時間)
	第6回	モデルを評価するその1	モデルを評価する方法として、モデルの有意差検定と決定係数などについて解説する。	教科書第3章p.154-173までを読んで分析手法を整理しておくこと(1時間)/分析手法と操作方法を確認しておくこと(2時間)
	第7回	モデルを評価するその2	モデルを評価する方法として、予測精度と多重共線性などについて解説する。	教科書第3章p.173-198までを読んで分析手法を整理しておくこと(1時間)/分析手法と操作方法を確認しておくこと(2時間)
	データを	実践的なデータ分析の準備として、デー	教科書第4章p.200-208まで、p.231-242までを	

	第8回	加工するその1	データのクレンジングについて解説する。	読んで分析手法を整理しておくこと(1時間)/分析手法と操作方法を確認しておくこと(3時間)
	第9回	データを加工するその2	実践的なデータ分析の準備として、変数と値の加工について解説する。	教科書第4章p.209-230までを読んで分析手法を整理しておくこと(1時間)/分析手法と操作方法を確認しておくこと(2時間)
	第10回	クラスタリング	データを分類する方法として、クラスタリングについて解説する。	教科書第4章p.243-254までを読んで分析手法を整理しておくこと(1時間)/分析手法と操作方法を確認し、演習課題に取り組む(4時間)
	第11回	因子分析と主成分分析	データに対する指標を集約する方法として、因子分析と主成分分析について解説する。	教科書第4章p.254-279までを読んで分析手法を整理しておくこと(1時間)/分析手法と操作方法を確認し、演習課題に取り組む(4時間)
	第12回	決定木	結果に対する要因を抽出する方法として、決定木について解説する。	教科書第4章p.298-308までを読んで分析手法を整理しておくこと(1時間)/分析手法と操作方法を確認し、演習課題に取り組む(4時間)
	第13回	因果推論	データから因果関係を明らかにする方法として、統計的因果推論について解説する。	教科書第4章p.309-347までを読んで分析手法を整理しておくこと(1時間)/分析手法と操作方法を確認しておくこと(3時間)
	第14回	講義のまとめ	講義のまとめとレポート課題の取り組み方の解説を行う。	講義全般で得た知識を確認しておくこと(5時間)/レポートを作成すること(5時間)
到達目標	(1) データマイニングの概念とその意義を説明することができる。 (2) データマイニングに用いられるツールおよびライブラリの基本的な利用方法を説明することができる。 (3) 問題の性質に応じて有効なツールを選択・組み合わせることができる。 (4) データセットに適切なツールを適用し、複数のアルゴリズムを用いてデータの解析ができる。			
評価方法	定期試験は実施しない。確認テストと演習課題、レポート等をあわせて評価する。到達目標(1)は、第3回の確認テストで評価する。到達目標(2)と(3)は第4回～第5回及び第10回～第12回の講義で示す演習課題で評価する。到達目標(4)は、第14回の講義で示すレポートで評価する(確認テスト20%、課題40%、レポート40%)。			
成績評価基準	A:(1)と(2)を達成しており、(3)と(4)を90%以上達成している。 B:(1)と(2)を達成しており、(3)と(4)を80%以上達成している。 C:(1)と(2)を達成しており、(3)と(4)を70%以上達成している。 D:(1)と(2)を達成しており、(3)と(4)を60%以上達成している。 F:上記以外。			
教科書	書名	著者名	出版社名	
	1. RとPythonで学ぶ実践的データサイエンス&機械学習[増補改訂版]	有賀友紀・大橋俊介	技術評論社	
	2. <ノートPC必携>			
参考書	書名	著者名	出版社名	
	1. まずはこの一冊から 意味がわかる統計学	石井俊全	ベレ出版	
	2. まずはこの一冊から意味がわかる多変量解析	石井俊全	ベレ出版	
	3. 多変量解析入門	小西貞則	岩波書店	
受講心得	教科書の該当範囲を事前に読んで専門用語の整理をしておく必要があります。Rプログラミングのトラブルについては各自で調べて対応することが重要です。演習課題とレポートは、授業を受けたうえで課題となるデータセットに対する分析を行い、結果を提出しなくてはなりません。またフィードバックとして、確認テスト及び演習課題において理解が十分でなかった点は次回の講義内で解説するので理解を深めること。			
オフィスアワー	水曜日3限 1号館 4F 415研究室			
実践的教育	データマイニングを用いたマーケティングリサーチの実務経験を持つ教員がその経験を生かしてデータマイニングの基本的な手法について講義する。			

