熊本高等専門学校	開講年度	令和03年度 (2021年度)	授業科目	データサイエンス

科目基礎情報

科目番号	AE1111	科目区分	専門 / 選択
授業形態	授業	単位の種別と単位数	学修単位: 2
開設学科	電子情報システム工学専攻	対象学年	専1
開設期	後期	週時間数	2
教科書/教材	（教科書）金 明哲 「Rによるデータサイエンス（第２版）」 森北出版 ／（参考書）村上 純・日野 満司・山本 直樹・石田 明男 「統計ソフトRによる多次元データ処理入門」および「統計ソフトRによるデータ活用入門」 日新出版		
担当教員	山本 直樹		

到達目標

① データの解析・マイニングについて、これらの各手法が説明でき、これらを適用してデータの関係性や傾向の把握、データの予測・判別、データのグルーピング、データからのパターン発見などができる。
② データの視覚化について、一般的な2軸チャートによる視覚化、散布図行列などによる多次元の視覚化、ツリーやネットワークなどにより関係性の視覚化などができる。
③ 非構造化データ処理について、テンソル分解とそれを応用した画像および動画像処理について説明でき、データの分解処理と復元処理に適用できる。

ルーブリック

	理想的な到達レベルの目安	標準的な到達レベルの目安	未到達レベルの目安
データの解析・マイニング	これらの各手法がすべて説明でき、これらを適用してデータの関係性や傾向の把握、データの予測・判別、データのグルーピング、データからのパターン発見などが的確にできる。	これらの各手法が説明でき、これらを適用してデータの関係性や傾向の把握、データの予測・判別、データのグルーピング、データからのパターン発見などができる。	これらの各手法の一部しか説明できず、これらを適用してデータの関係性や傾向の把握、データの予測・判別、データのグルーピング、データからのパターン発見などができない。
データの視覚化	一般的な2軸チャートによる視覚化、散布図行列などによる多次元の視覚化、ツリーやネットワークによる関係性の視覚化などが的確にできる。	一般的な2軸チャートによる視覚化、散布図行列などによる多次元の視覚化、ツリーやネットワークによる関係性の視覚化などができる。	一般的な2軸チャートによる視覚化、散布図行列などによる多次元の視覚化、ツリーやネットワークによる関係性の視覚化などができない。
非構造化データ処理	テンソル分解とそれを応用した画像および動画像処理についてすべて説明でき、データの分解処理と復元処理に的確に適用できる。	テンソル分解とそれを応用した画像および動画像処理について説明でき、データの分解処理と復元処理に適用できる。	テンソル分解とそれを応用した画像および動画像処理について一部しか説明できず、データの分解処理と復元処理に適用できない。

学科の到達目標項目との関係

教育方法等

概要	現代社会においては、諸領域でたえず蓄積され続けている大規模なデータ（ビッグデータ）から知見を得て、新たな付加価値を見出すことが求められてきている。このようなデータを取り扱う分野はデータサイエンスと呼ばれ、この分野の実施者・研究者はデータサイエンティスト（DS）と呼ばれる。本授業では、データサイエンスにおけるデータの解析・マイニングに焦点を当て、それら各手法の基礎について説明し、専用のツールを活用してデータ解析を体験しながら各手法の理解を深めることにより、DSが身につけるべき基礎的なスキルを学習できる。
授業の進め方・方法	講義の形式で進める。授業内容としては、データマイニングの基礎、各データマイニング手法の基礎および各手法のケーススタディ、データの視覚化、非構造化データ処理などについて説明していく。さらに、複数回の小テストやRを活用した演習を行うことにより、本授業で説明された授業内容の理解をさらに深めさせる。
注意点	小テスト、レポートの成果物の内容、レポートの提出状況、授業の取り組み状況を考慮して評価する。 規定授業時数は30時間である。本科目はレポート課題作成等のため放課後・家庭で60時間の自学自習が求められる。

授業の属性・履修上の区分

☐ アクティブラーニング	☐ ICT 利用	☐ 遠隔授業対応	☐ 実務経験のある教員による授業

授業計画

		週	授業内容	週ごとの到達目標
後期	3rdQ	1週	ガイダンス・データマイニングとツール	データマイニングの概要について説明できる。解析ツールの基本操作ができる。
		2週	データの演算と行列演算・基本統計量	データの演算、行列演算などが説明でき、演算ができる。基本統計量について説明でき、計算できる。
		3週	データの視覚化	データを視覚化するための基本的なグラフが作成できる。
		4週	主成分分析	主成分分析（PCA）について説明できる。PCAの計算ができ、縮約されたデータから元のデータ構造の再現性が説明できる。
		5週	因子分析	因子分析（FA）について説明できる。FAの計算ができ、データの関係性の把握、因子の解釈などができる。
		6週	クラスター分析	クラスター分析について説明できる。クラスター分析の計算ができ、データのグルーピング、グループの特徴などについて分析できる。
		7週	線形回帰分析	単回帰分析、重回帰分析について説明できる。これらの回帰分析の計算ができ、データの予測ができる。
		8週	非線形回帰分析・自己組織化マップ	ロジスティック回帰、多項式回帰について説明できる。これら回帰分析の計算ができ、データの予測ができる。ニューラルネットワークの１つである自己組織化マップ（SOM）について説明できる。SOMの計算ができ、パターンの分類およびパターンの特徴を分析できる。
	4thQ	9週	決定木・回帰木	決定木、回帰木について説明できる。これらモデルの計算ができ、データの判別・予測ができる。
		10週	高階テンソル	高階テンソルについて説明できる。n-モード行列展開、畳み込み、n-モード積の計算ができる。

		11週	テンソル分解・非構造化データ処理	高次特異値分解（HOSVD）について説明でき、HOSVDの計算ができる。HOSVDを利用した画像および動画像処理について説明できる。HOSVDにより画像および動画像データを分解することができ、復元精度を調整して復元できる。
		12週	判別分析	判別分析について説明できる。線形および非線形判別分析の計算ができ、データの各個体がどのグループに属するかを判別できる。交差確認法を用いて判別率を求めることができる。
		13週	サポートベクターマシン・ランダムフォレスト	サポートベクターマシン（SVM）、ランダムフォレスト（RF）について説明できる。SVM、RFの計算ができ、データの判別・予測ができる。
		14週	多次元主成分分析	多次元主成分分析（MPCA）について説明できる。MPCAの計算ができ、データの特徴抽出および分類ができる。
		15週	確認テスト	各種手法について説明でき、それらのアルゴリズムが実装できることを確認する。
		16週	テスト返却、学習のまとめ	テストの結果から理解していなかったところを把握し、適切な解答を理解できる。

モデルコアカリキュラムの学習内容と到達目標

分類	分野	学習内容	学習内容の到達目標	到達レベル	授業週

評価割合

	小テスト・レポート						合計
総合評価割合	100	0	0	0	0	0	100
基礎的能力	20	0	0	0	0	0	20
専門的能力	80	0	0	0	0	0	80
分野横断的能力	0	0	0	0	0	0	0