

統計研修講義計画書

研修課程	特別コース データサイエンスセミ応用	講義 科目	データサイエンス実践	講義時間	3コマ
講師	国立大学法人滋賀大学 データサイエンス教育研究センター 教授 佐藤健一				

※1コマ70分

講義のねらい
<p>複数の項目間の類似性を定量化し、分類する方法を学ぶ。そして、与えられた分類に対して、関連する要因から予測ルールを作成し、これを評価する方法を学ぶ。</p>

指導項目と内容	
指導項目	内 容
クラスター分析 階層的クラスタリング デンドログラム k-means 法	<p>多変量データに対する相関係数と距離の関係を述べ、距離行列を求める。距離行列から項目間の結合方法を与えることでボトムアップのクラスタリング手法として層別クラスタリングを紹介し、その視覚化としてデンドログラムを学ぶ。一方で、トップダウン方式のクラスタリングとして k-means 法を紹介し、それぞれの長所と短所を指摘する。</p>
判別分析とロジスティック回帰 線形判別 決定木による判別 ロジスティック回帰	<p>関連要因をもつ2群の分類として、線形判別分析を考える。また、線形判別とマハラノビス距離の関係を示す。さらに、平面上に分類ラベルが図示されている場合に、ラベルの平面分割として決定木を紹介する。また、2値データの判別に有用なロジスティック判別と、ROC 曲線による判別結果の評価方法を示す。</p>
ロジスティック回帰モデル 係数の検定と信頼区間 変数選択 パソコンを使った演習	<p>2値データに対するオッズ比を推定するための回帰分析としてロジスティック回帰を学ぶ。2 × 2分割表や用量反応曲線としての解釈を紹介し、推定結果の読み方についても学ぶ。また、複数の説明変数がある場合に有用な AIC（赤池情報量基準）を用いた変数選択を示す。解析例については、統計ソフト R を用いた実演を行い、解析手法の理解を深める。</p>

講義形態 指導方法	PC プロジェクタを用いる。パワーポイントによるスライドを使った解説や、統計ソフト R を用いながら解析例を実演により紹介する
受講に必要な 基礎知識等	高校数学（数 III を除く）