

**「データ解析」のおわりに
～まとめ～**

授業計画

Prologue: データを読み解くとは何なのか

(多変量解析とデータサイエンスと統計学とパターン認識と機械学習とデータマイニング)

DAY-1 6/16 (01)(02) 単回帰: 点群への直線当てはめを“真剣に”考える

(見えない世界へようこそ)

DAY-2 6/23 (03)(04) 重回帰と線形代数: 回帰の行列計算とその意味

(データの計算とデータの解釈)

DAY-3 6/30 (05)(06) 重回帰と確率統計: なぜ回帰に確率が必要?

(推測統計入門: データの向こう側について語るための代償)

DAY-4 7/07 (07)(08) 多変量正規分布: 多次元の正規分布と線形代数

(ゼロから理解する正規分布)

DAY-5 7/14 (09)(10) マハラノビス距離と判別分析: 線形代数を使う1

(最適な判別とは)

DAY-6 7/28 (11)(12) 固有値分解と主成分分析: 線形代数を使う2

(高次元データがかかえる大問題)

DAY-7 8/04 (13)(14) 特異値分解と数量化: 線形代数を使う3

(数値じゃない対象に統計を効かすには)

Epilogue: 基礎の上に在る世界(話したことと話さなかったこと)

多変量解析の主題

回帰分析 ⇒ 4,5章

数量化I類

⇒ 6章

判別分析 ⇒ 7章

数量化II類

⇒ 8章

主成分分析 ⇒ 9章

数量化III類

⇒ 10章

多次元尺度構成法(12章)、クラスター分析(11章)、
因子分析・パス分析・グラフィカルモデル(13章)、
正準相関分析(13章)、多段層別分析(13章)

線形代数の技術

射影行列

線形写像の像と核

直交直和分解

(線形代数学の基本定理)

2次形式

基底変換

固有値・固有ベクトル

直交行列による対角化



多変量正規分布

標準化、マハラノビス距離

主な多変量解析手法の分類

外的基準(目的変数)あり “教師あり”

目的変数=量的 説明変数=量的 回帰分析

目的変数=質的 説明変数=量的 判別分析

ロジスティック回帰分析

目的変数=量的 説明変数=質的 数量化I類

目的変数=質的 説明変数=質的 数量化II類

外的基準(目的変数)なし “教師なし”

変数=量的 主成分分析, クラスタ分析, 多次元尺度構成

変数=質的 数量化III類, IV類~VI類, クロス表(分割表)分析

この先は？(講義で話していないこと)

- 分散分析・多重比較・実験計画法
- クラスタ分析 (クラスタリング)
- 決定木・回帰木モデル
- ロジスティック回帰・一般化線形モデル
- カテゴリカルデータ分析 (クロス集計)
- 多次元尺度構成法 (MDS)
- 因子分析・パス解析・構造方程式モデリング・共分散分析
- 正準相関分析・独立成分分析
- 非線形回帰分析・ノンパラメトリック回帰
- ベイズ推論、階層ベイズ法、ベイジアンネットワーク
- 隠れ変数モデリング、確率モデル、生成モデル
- 機械学習 (教師付き学習, 教師無し学習, 強化学習)

先週配布したレポートを提出すること！

提出先: 8/4の授業日 or 6-16瀧川居室ドアポスト or takigawa@ist.hokudai.ac.jpへメール提出

〆切: 8/5(金) 17:00まで (難しい場合は事前に相談してください！)

注意: レポートには必ず名前と学生番号を記載すること！！

【講義Webサイト】

いままでの講義スライドや下の表1のデータファイルを置いてあります。

注: httpじゃなくhttpsでないと資料にアクセスできません。(たぶん)

<https://art.ist.hokudai.ac.jp/~takigawa/course/da2016>

(閲覧ID da2016 パスワード pokemongo)