

2022年度秋学期
データサイエンス特論 第7回
レポート課題 #2

立命館大学
情報理工学部
前田忠彦

予測：静的対象 動的対象

- 過去のデータで完結
- 分析結果が未来の事象に影響を与えないもの
- 分析結果が共有されることで未来の事象に影響を与えるもの
- それぞれの例は？

判別問題

■ 単判別問題

2 グループの判別問題

合格 不合格 共和党 民主党

- y を x の何かしらの関数 $y = f(x)$ として表現できるか？

データ集積手法の適正性

■ データの調査方法 信頼性

- *What Is Public Opinion Polling and Why Is It Important? — Whidbey Island Center for the Arts (wicaonline.org)*

■ アメリカ選挙予測

- *Public opinion polls*
- *Polling for United States presidential elections - Wikipedia*

予備選挙と本選挙の関係

- [Election and voting information | FEC](https://www.fec.gov/introduction-campaign-finance/election-and-voting-information/)
- <https://www.fec.gov/introduction-campaign-finance/election-and-voting-information/>
- 多角的に分析
 - 州による差異・特徴
 - 上院と下院による差異・特徴
 - これらの年度による変化

Public Opinion Polls の意義

- 世の中の把握
- 世の中の未来予測
- 世の中の改革？
- Mass media
- Social media

レポート課題 # 2

- <https://www.fec.gov/introduction-campaign-finance/election-and-voting-information/>
- 上記のデータを分析する
 - 上記のページから年度ごとのPDFとExcelデータにアクセス可能 2010から2018年までが使いやすいが.....
- 上記に加えて独自に調査して2022年11月の中間選挙の結果をデータに加える
- 期限
 - 12/5(月) 9:00 AM

レポート課題 # 2

- 例えば前述のデータを判別分析 機械学習 SVM 主成分分析 クラスタリング等を用いて分析・予測する
 - *Senate, House, State, and Year* など分析要素
- 提出物
 - 分析方法・結果・分析を含む論文形式レポート (PDF)
 - ソースコードは付録
 - *Manaba+R* より提出すること
- 期限
 - 12/5 (月) 9:00 AM

レポートに含むべき事項

- 生データのどの部分をどういう方法で取り出し、何を分析しようとしているか。
- 実際にどのように分析をしたか
- 結果：表 グラフや数値など。
- 分析：上記から何が読み取れるか。何が分かったか。妥当性の検証。
- 今回の達成成果を予測に使えるか？

補足

- 論文形式で作成する
 - 冒頭のあらましで本人のオリジナリティ 達成成果がわかること
 - 参考文献 付録（プログラムリスト） など
 - レポートは下記のフォーマットを使用
https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fwww.ieice.org%2Fjpn%2Fshiori%2Fjb%2Fcs_papertemplate.docx&wdOrigin=BROWSELINK
- 本文ページ数は6程度を推奨(最大18ページ) 付録はこれに含まない PDFファイルの容量1MB以下推奨 2.5MB以上は減点 写真は避けたい
- 計算しました。終わり。では不十分
 - 選択したモデルの妥当性は？
 - データから何が分析できたか？
 - 結果の妥当性 結果の一般性 予測の妥当性
- 言語の種別は問わない。
 - *R, Python, MATLAB, C++, Excel*どれもOK