

高等学校教育「情報Ⅱ」

# 高等学校における「情報Ⅱ」のための データサイエンス・データ解析入門



監修：総務省統計研究研修所

制作：株式会社Rejoui

## 第1章

データサイエンス(機械学習のアルゴリズム)によるデータ解析が社会にもたらす変化 P.1~

はじめに(社会における変化)

### 機械学習とデータ×AIの関係性

- ・ データサイエンスとは
- ・ 機械学習の基本概念
- ・ ビッグデータ
- ・ IoT
- ・ AI
- ・ ロボティクス
- ・ 量子コンピュータ

### データとは何か

- ・ 構造化データ、非構造化データの種類
- ・ 構造化データとは
- ・ 非構造化データとは
- ・ テキスト
- ・ 画像／動画
- ・ 音声／音楽

### AIの根幹となる技術・機械学習とは

- ・ 機械学習とは
- ・ 教師あり学習
- ・ 教師なし学習

### AI最新技術の活用例

- ・ 製造業
- ・ 医療
- ・ インフラ
- ・ 金融業
- ・ 流通業

## 第2章

### 機械学習を用いたデータサイエンスのプロセス

P.24～

データ収集

データの可視化と特徴量の選択

モデリング

評価

業務への適用

## 第3章

### 機械学習(教師あり学習)

P.37～

教師あり学習とは

予測モデルおよび予測結果の評価(教師データと評価データに分割して検証する交差検証法)

交差検証の方法:K分割法

データ分割方法:ホールドアウト法

教師あり学習の手法と活用例

- ・線形回帰分析
- ・
- ・数量化理論(I類、II類)
- ・サポートベクターマシン
- ・決定木
- ・ランダムフォレスト
- ・ニューラルネット
- ・リカレントネットワーク(RNN)・ディープラーニング

・ナイーブベイズ法

・K 近傍法

アノテーション

過学習と次元の呪い(及びその対処法)

## 第4章

機械学習(教師なし学習)

P.88~

教師なし学習とは

教師なし学習の手法と活用例(具体的ケースの提示・手法の説明)

・クラスター分析

・主成分分析

・因子分析

・アソシエーション分析

・数量化理論(III 類、IV 類)

## 第5章

構造化データ処理の基本

P.124~

構造化データ処理の基本

・並べ替え

・抽出

・データ型変換

・集計

・重複・欠損・異常値の処理

・データの結合

## 第6章

### 非構造化データ処理の基本

P.128~

非構造化データ処理の種類／R/PYTHON を用いた非構造化データ処理の実行

- ・言語処理の基本とその実行方法
- ・画像処理の基本とその実行方法

## 第7章

### プログラミングの基本

P.132~

R/PYTHON の基本構文

- ・四則演算
- ・代入
- ・抽出
- ・代表値の算出
- ・集計
- ・可視化およびグラフ作成

### 参考

P.156~

#### 【参考】環境構築

R の環境構築方法

PYTHON の環境構築方法

#### 【参考】ライブラリの紹介

R の代表的なライブラリ

PYTHON の代表的なライブラリ

#### 【参考】参考文献・推薦図書・データセット