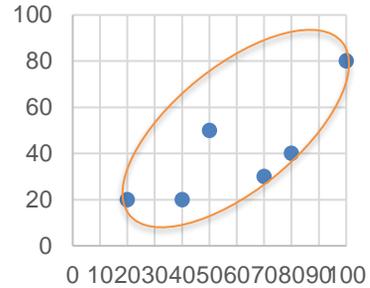


DAY1 データサイエンス入門

1. データサイエンスとは

- ・データとは
- ・分布の特定の位置を示す統計量
代表値（平均値、中央値、最頻値）
- ・分布のばらつきを示す統計量
散布度（分散、標準偏差、四分位範囲）
- ・データの可視化（グラフ）
ヒストグラム、箱ひげ図、散布図、棒グラフ、円グラフ
- ・オープンデータ（e-Stat）

散布図



2. データサイエンスの手法

- ・統計学の歴史
- ・機械学習と多変量解析
- ・教師あり学習 回帰
回帰と相関、単回帰、重回帰
- ・教師あり学習 分類
k近傍法、決定木、ランダムフォレスト、勾配ブースティング木
- ・教師なし学習 次元削減
主成分分析
- ・教師なし学習 クラスタリング
階層クラスター分析、非階層クラスター分析
- ・ニューラルネットワーク

3. ベイズ推定

- ・確率の公理、条件付き確率、全確率の定理、ベイズの定理

4. Kaggle タイタニック生存予測問題（分類）

- ・分析の流れ、注意点

5. 演習（Excel, Python）

- （Excel）
- ・ピボットテーブル
- ・マンションの家賃予測（回帰）
- （Python）
- ・得点データの主成分分析（次元削減）
- ・クラスター分析（クラスタリング）

DAY2 Python実習 深層学習

1. 開発／実行環境（Google Colaboratoryを利用）の構築と利用
 - ・Anacondaを用いたオンプレ、Anaconda Cloudについての概略説明
 - ・Google Colaboratory（以降Colab）についての概略説明
 - ・利用設定の実施（各自、Googleアカウントは事前作成済みとします）
2. 畳み込みニューラルネットワーク（Convolutional Neural Networks,CNN）の理解と実行（Colab利用）
 - ・画像分類問題
 - ・CNNの説明
 - ・CNNサンプル（PyTorch版、Notebook形式）をColabにアップロード（各自）
 - ・CNNサンプルコードの説明と段階的な実行（各自）
3. NNの調整方法についての実践（CNNサンプルを基に、正答率、学習速度の向上を目指す）
 - ・ハイパーパラメータとその調整方法の説明
 - ・ハイパーパラメータ調整結果の紹介
 - ・ハイパーパラメータ調整の実行（各自、Colab利用）

DAY3 統計データ分析（記述統計・推測総計・確率）

- ・統計学とは
 - ・1次元データ
 - ・2次元データ
 - ・回帰分析
 - ・確率
 - ・標本分布
 - ・推測統計学
 - ・演習
- 1.記述統計学と推測統計学、2.質的データと量的データ、3.代表値、4.散布度、5.偏差値、6.歪度と尖度、7.共分散、8.相関係数、9.組合せ、10.条件付き確率、11.確率、12.標準正規分布、13.期待値、14.標本抽出法、15.母平均の区間推定、16.母比率の区間推定（視聴率）、17.母比率の区間推定（支持率）、18.標本の大きさ、19.平均値に関する検定、20.t検定、21.平均値の差の検定等

学習項目（データサイエンス人材育成講座）

DAY4 Python実習 機械学習

- 1.はじめに
- 2.Google Colaboratoryの利用
- 3.機械学習サンプルプログラムの理解と実行
 - ・機械学習サンプルプログラムの概要
 - ・機械学習サンプルプログラムのソースコードの概要
 - ・Pythonプログラムの実行
 - ・機械学習サンプルプログラムのソースコードの説明
(線形サポートベクターマシン、カーネルサポートベクターマシン、ロジスティック回帰、k-近傍法、決定木、ランダムフォレスト、勾配ブースティング回帰木)
 - ・考察（各自）
 - ・Google Colaboratoryの編集機能の説明
- 4.まとめ

DAY5 kaggle（データコンペティション）演習

- 0.kaggleとは
- 1.分析コンペ課題の理解
- 2.データの読み込み
- 3.データの理解
- 4.データの前処理
- 5.特徴量
- 6.モデリング
- 7.評価
- 8.提出
- 9.発表