

# データサイエンス 基礎

## 第1回 効果測定 問題

### 問題1 (1):2点、(2)～(7):各3点(合計20点)

データサイエンスに関する以下の説明に当てはまる用語を解答群から選択して記号を答えなさい。

\_\_\_\_(1)\_\_\_\_学習とは、記憶したデータから特定のパターンを見つけ出すなどの、人が自然に行っている学習能力をコンピュータにもたせるための技術の総称である。

\_\_\_\_(1)\_\_\_\_は、大きく、教師あり学習と、教師なし学習、強化学習に分けられる。

教師あり学習とは、\_\_\_\_(2)\_\_\_\_から\_\_\_\_(3)\_\_\_\_を予想するモデルを求める手法である。

例えば、メールの内容やタイトルから、そのメールがスパムメールか否かを判断する場合、メールのタイトルや内容は\_\_\_\_(2)\_\_\_\_、スパムメールかどうかの予想結果は\_\_\_\_(3)\_\_\_\_である。教師あり学習には、\_\_\_\_(4)\_\_\_\_と\_\_\_\_(5)\_\_\_\_の手法が存在する。

教師なし学習は、データのパターンなどを解析し与えられたデータの中からグループ化を行う\_\_\_\_(6)\_\_\_\_や、大量のデータを少ないデータ数で表す手法である\_\_\_\_(7)\_\_\_\_などのタスクが存在する。

### 解答群

ア AI	イ 強化	ウ 機械
エ 説明変数・特徴量	オ 目的処理	カ 可視化
キ 目的変数	ク 前処理	ケ クラスタリング
コ 分類	サ 回帰	シ 次元削減
ス 数学	セ 分析	ソ 同一化

# データサイエンス 基礎

## 第1回 効果測定 問題

### 問2 合計 40 点(①～⑥:各 5 点、⑦:10 点)

「ds\_exam1\_2.ipynb」のテキストやコメントの指示どおりにデータの前処理を行い、実行可能な ipynb ファイルを提出してください。

提出するファイル名は、「ds\_exam1\_2\_氏名.ipynb」とすること。

#### 【処理内容】

- ① 1 以上、1000 未満の一樣分布の乱数を10個発生させて、2行5列の行列を生成する
- ② 100 以上、2000 未満の一樣分布の乱数を15個発生させて、3行5列の行列を生成する
- ③ ①②で生成した行列をそれぞれ表示する
- ④ ①②で生成した行列を結合して、5行5列の行列を生成する
- ⑤ ④で生成した行列を表示する
- ⑥ ④で作成した各列ごとの最大値、最小値、合計、平均を表示
- ⑦ ④で作成した5行5列の行列を下図のように、赤、青、緑の3つの部分に分割してそれぞれを表示する

685	560	630	193	836
764	708	360	10	724
377	1878	1928	699	1194
1596	700	1520	414	805
1610	651	187	274	1724

# データサイエンス 基礎

## 第1回 効果測定 問題

問題3 (1):各3点、(2),(3):各2点、合計:40点

(1)

下記の実行結果となるように、「ds\_exam1\_3.ipynb」のテキストやコメントの指示どおりにデータの処理を行い、実行可能な ipynb ファイルを提出してください。

提出するファイル名は、「ds\_exam1\_3.ipynb\_氏名.ipynb」とすること。

プログラム内で使用している「3-運動能力.csv」「3-名前.csv」は配付されたデータを使用すること。

【処理内容】

- ① 日本語を含む csv データを読み込む
- ② DataFrame の先頭 5 行を表示する
- ③ データの 1 行目を DataFrame から削除する
- ④ 欠損値があるか確認(列ごとに欠損値の個数をカウント)する
- ⑤ 列名「腹囲」のみを出力して内容を確認する
- ⑥ 腹囲列の欠損値を、一つ前のデータで補完する
- ⑦ 心拍数列の欠損値を、中央値で補完する
- ⑧ Matplotlib を用いて、体重、腹囲の箱ひげ図を描画する
  - ✓ x 軸を体重、y 軸を腹囲とする
  - ✓ y 軸の範囲を 0～130 とする
  - ✓ グラフのグリッドを表示する
- ⑨ 各列の相関を調べるために、ID 列を除いたデータを使って相関マトリックスを作成し表示する
- ⑩ 氏名、出身地が格納された別の CSV データを読み込み、①で読み込んだ DataFrame に内部結合させ、内容を表示する
- ⑪ 体重が 90 以上の行のみを抽出して表示する
  - ✓ 以下の表示例のように DataFrame の形式で表示すること

	ID	逆手懸垂回数	上体起こし回数	ジャンプ	体重	腹囲	心拍数	name	birthplace
6	1000007	8	101.0	38	95.69	96.52	56.0	Lalihi	Jaipur
13	1000014	1	50.0	50	112.00	130.80	50.0	Ray	Mumbai
15	1000016	12	NaN	120	91.60	93.98	62.0	Ghosh	Chennai

- ⑫ ⑪で表示したデータを体重の降順に並び替えて表示する

(2)

(1)–⑧で作成した体重の箱ひげ図を確認し、腹囲列に外れ値がある場合は、解答欄に「あり」と、外れ値がない場合は「なし」と答えなさい。

(3)

(1)–⑨で作成した相関マトリックスより、体重列と最も相関関係が高い列を答えなさい。回答欄には列名を記載してください。