

ここをクリックして cs3-end1-p1-ij.zip をダウンロードし、展開・解凍すると、cs3-end1-p1 フォルダが生成し、その中に解答に必要なファイルが置かれる。必要に応じて、cs3-end1-p1-sample.ipynb を使用して良い。

「小数第N位まで」で答える場合は、第(N+1)位を四捨五入すること。たとえば「13.79584...」を「小数第2位まで」で答える場合は「13.80」となる。

問1. 以下の問いに答えよ。

1-1. ロジスティック回帰に関する以下の問いに答えよ。(各1点)

1-1-1 目的変数はどのようなものか、もっとも適当なものを選択せよ。

(未選択)

1-1-2 数値モデルとして実質的に同等なものはどれか、もっとも適当なものを選択せよ。

(未選択)

1-1-3 性能評価指標の1つに、目的変数の予測値が、実際の目的変数と一致した割合がある。これをなんと呼ぶか、もっとも適当なものを選択せよ。

(未選択)

1-2. end1-p1.csv を読み込んで、以下の手順で解析を行い、その結果を報告せよ。なお、解答の根拠となる ipynb/html ファイルが両方とも提出されていない場合は、解答を無効とする。(各3点)

- 説明変数を x_1, x_2, x_4 、目的変数を y とする。
- まず、標準化なしで、線形重回帰分析を行う。自由度調整済み決定係数を小数第3位までで答えよ。
- 次に、全説明変数と目的変数を標準化して、線形重回帰分析を行う。自由度調整済み決定係数を小数第3位までで答えよ。
- 標準化なしで得られたモデルに、データ $x_1 = [1, 1, 1]$ および $x_2 = [2, 2, 2]$ をそれぞれ入力したときの予測値 y_1, y_2 を小数第3位までで答えよ。
- 説明変数を、目的変数に対する影響の大きい順に並べよ。解答は、括弧や引用符などは付けずに、説明変数名だけをコンマで区切って並べること。
- その説明変数の値が増加したときに目的変数の値も増加する傾向にある説明変数をすべて挙げよ。解答方式は5.と同じ。
- その説明変数の値が1増えて、残りの説明変数の値が変わらないときに、目的変数の値がもっとも増える説明変数名と、目的変数の増分(小数第3位まで)をそれぞれ答えよ。

2.	<input type="text"/>
3.	<input type="text"/>
4. y_1	<input type="text"/>
4. y_2	<input type="text"/>
5.	<input type="text"/>
6.	<input type="text"/>
7. 説明変数名	<input type="text"/>
7. 目的変数の増分	<input type="text"/>

ipynbファイルのアップロード:

① 未提出

htmlファイルのアップロード:

① 未提出

ここをクリックして cs3-end1-p2-nj.zip をダウンロードし、展開・解凍すると、cs3-end1-p2 フォルダが生成し、その中に解答に必要なファイルが置かれる。必要に応じて、cs3-end1-p2-sample.ipynb を使用して良い。

「小数第N位まで」を答える場合は、第(N+1)位を四捨五入すること。たとえば「13.79584...」を「小数第2位まで」で答える場合は「13.80」となる。

問2. 以下の問いに答えよ。

2-1. 決定木に関する以下の問いに答えよ。(各1点)

2-1-1 データを「質問」により2群に分ける際、データの偏りができるだけ (1) となるような質問から順に適用していく。空欄に当てはまる語を選択せよ。

(1)

(未選択)

2-1-2 以下の選択肢から、目的変数がカテゴリー変数の場合のデータの偏りの指標としてもっとも適当なものを選択せよ (1)。また偏りがより大きいときに、その指標は値が大きくなるか小さくなるかを答えよ (2)。

(1)

(未選択)

(2)

(未選択)

2-1-3 以下の選択肢から、目的変数が量的変数の場合のデータの偏りの指標としてもっとも適当なものを選択せよ (1)。また偏りがより大きいときに、その指標は値が大きくなるか小さくなるかを答えよ (2)。

(1)

(未選択)

(2)

(未選択)

2-2. end1-p2.csv を読み込んで、以下の手順で解析を行い、その結果を報告せよ。なお、解答の根拠となるipynb/htmlファイルが両方とも提出されていない場合は、解答を無効とする。(各3点)

- 1. 説明変数を e1, e2, e3, e4、目的変数を y とする。
- 2. e1 列の値は、数が小さい方から 1, 2, 3, ... と 1から始まる整数値を順に割り当てる。
- 3. e2 列の値はダミー変数化する。その際、アルファベット順で先頭の値に対するダミー変数は作成しないようにする。
- 4. train_test_split関数を用いて、訓練データとテストデータに分ける。random_state=28を引数に含め、テストデータの割合を0.33 とする。テストデータ数を答えよ。
- 5. ランダムフォレスト回帰モデルを n_estimators=300, max_depth=None, random_state=43 のオプション引数をつけて作成し、訓練データを用いて学習させる。
- 6. 学習済みモデルに訓練データを入力したときの平均2乗誤差の平方根を求め、小数第3位まで答えよ。
- 7. 学習済みモデルにテストデータを入力したときの平均2乗誤差の平方根を求め、小数第3位まで答えよ。
- 8. 各説明変数の重要度を求め、重要度がもっとも大きい説明変数名と、その重要度(小数第3位まで) を答えよ。

(以下はオプション問題、5点)

9. 学習済みモデルにend1-p2-pred.csv のデータを入力したときの出力を求め、出力の最大値を小数第3位までで答えよ。なおこのプログラミングは、8.までに用いたものと同じipynbファイル上で、8.の続きとして実行し、別のipynbファイルを作成しないこと。

4.

6.

7.

8. 変数名

8. 重要度

(オプション問題)

9.

ipynbファイルのアップロード:

📎 ファイルをアップロード

🕒 未提出

htmlファイルのアップロード:

📎 ファイルをアップロード

🕒 未提出

「小数第N位まで」を答える場合は、第(N+1)位を四捨五入すること。たとえば「13.79584...」を「小数第2位まで」で答える場合は「13.80」となる。

問3. 以下の問いに答えよ。(各2点)

3-1. 以下のプログラムで、あるRGBカラー画像ファイル image.jpg を操作した。なお、左端の番号列は行番号であり、プログラムの一部ではない。

```
1 import numpy as np
2 import matplotlib.pyplot as plt
3 from PIL import Image
4 img = Image.open('image.jpg')
5 print(img.getpixel(100,100))
6 img2 = img.__(1)___
7 img3 = img.__(2)___
8 img4 = img.__(3)___
9 img5 = img.__(4)___
```

3-1-1 行番号5の出力が、(0, 64, 0)であったとする。このピクセルの色としてもっとも適当なものを以下から選択せよ。

(未選択)

3-1-2 img2 が img を 時計回り に35度回転させたものになるように、(1)を埋めよ。

(1)

3-1-3 img3 が img を 横 40 ピクセル、縦 80ピクセル にサイズ変更したものになるように、(2)を埋めよ。

(2)

3-1-4 img4 が img を 回転させずに、右に 100 ピクセル、上に 150 ピクセル 平行移動したものになるように、(3)を埋めよ。

(3)

3-1-5 img5 が img をグレースケール画像に変換したものになるように、(4)を埋めよ。

(4)

3-2. 以下のプログラムは、カレントディレクトリにある、拡張子が「.txt」であるすべてのファイルの中身を、余分な改行なしに表示するプログラムである。各空欄を、それぞれ1行のプログラムで埋めよ。

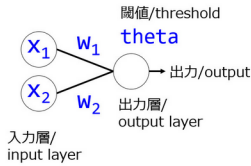
```
for file1 in glob.glob('*.txt'):
    with __ (1) __ as f:
        for __ (2) __:
            print(one_line, end='')
```

(1)

(2)

問4. パーセプトロンに関する以下の問いに答えよ。

4-1. 下図のような2入力1出力の単純パーセプトロンについて、下の文章の空欄を埋め、また問いに答えよ。
(3)(4) 各3点, (5) 4点, 他は各1点)



機械学習法による分類予測とは、与えられた訓練データを、教えられた分類(教師信号)にしたがってうまく分けられるような境界を求め(この操作を学習と呼ぶ)、その境界を用いて未知データに対する予測を行う。この境界を (1) と呼ぶ。一般に、予測は学習に比べて計算時間が (2) 。

(1)

(2) ☐ かかる ☐ かからない

出力層のニューロンの活性化関数が階段関数 $f(z)$ であり、重み $(w1, w2) = (-1.5, 2.5)$ 、閾値 $theta = 2.0$ であるとする。このパーセプトロンへの入力 $(x1, x2) = (2.0, -4.0)$ のとき、 $z =$ (3) であり、このパーセプトロンからの出力は (4) である。

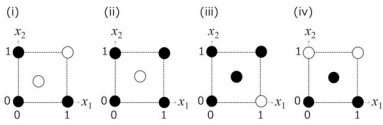
※(3)(4)に当てはまる数値のみを答えよ。

(3)

(4)

以下の図(i)から(iv)のデータのうち、この単純パーセプトロンで学習可能なものをすべて選択せよ。(5)。

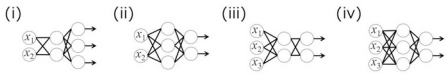
- ☐ 教師信号/teaching signal 1
☒ 教師信号/teaching signal 0



(5) ☐ (i) ☐ (ii) ☐ (iii) ☐ (iv)

4-2. TensorFlowを用いて、以下のプログラムで、パーセプトロンを構築した。なお、プログラムは主要部分のみを示している。このパーセプトロンの構造を以下の図から選択せよ。(各1点)

```
model = Sequential()  
model.add(Dense(3, input_shape=(2,), activation='relu'))  
model.add(Dense(2, activation='sigmoid'))
```



4-2 ☐ (i) ☐ (ii) ☐ (iii) ☐ (iv)

問5. CNNに関する以下の問いに答えよ。

5-1. 1次元配列 $A = [3, 1, -2, -1, 5]$ に1次元のフィルタ $f = [2, -1, 2]$ を用いて、パディングなしの畳み込み演算を行った結果を答えよ。得られた数値を、例のようにコンマで区切って並べる。全体を囲む括弧などは不要。例: 1, 2, 3, 4, 5 (3点)

5-1

5-2. TensorFlowを用いて、以下のプログラムで、画像に対する分類予測を行うCNNを構築した。なお、プログラムは主要部分のみ示している。また左端の番号列は行番号であり、プログラムの一部ではない。以下の文章の空欄を埋めよ。(各1点)

訓練用の画像データ10000枚と対応する教師データがそれぞれ X_{train} 、 y_{train} に、またテスト用の画像データ2000枚と対応する教師データがそれぞれ X_{test} 、 y_{test} に格納されているとする。

```
1 model = Sequential()
2 model.add(Conv2D(16, (4,4), activation='relu', padding='same', input_shape=(32, 24, 3)))
3 model.add(MaxPooling2D(pool_size=(2,2))
4 model.add(Dropout(0.45))
5 model.add(Flatten())
6 model.add(Dense(32, activation='relu'))
7 model.add(Dense(4, activation='softmax'))
8 model.compile(loss='categorical_crossentropy', optimizer='rmsprop', metrics=['accuracy'])
9 es=EarlyStopping(monitor='val_loss', patience=14, verbose=1)
10 fit_log = model.fit(X_train, y_train, epochs=100, validation_split=0.25, callbacks=[es])
11 score = model.evaluate(X_test, y_test)
```

このCNNの入力は、縦 (1) ピクセル、横 (2) ピクセルの (3) 画像である。

(1)
(2)

(3) ☐ カラー ☐ グレースケール

行番号2は (4) 層で、縦横ともに (5) ピクセルの重みフィルタ (6) 枚を入力画像に適用している。

(4)
(未選択)

(5)
(6)

行番号3は、(7) 層 (縦横ともに (8) ピクセル) で、ここから出力されるのは、縦 (9) ピクセル、横 (10) ピクセルの (11) 枚の縮小された特徴マップである。

(7)
(未選択)

(8)
(9)
(10)
(11)

行番号8は学習方法の設定である。(12) という最適化方法で、(13) 関数で計算した「出力と教師信号の間の誤差」をできるだけ小さくするように学習する。

※ (12) (13) はプログラムから抜き出して答えること。

(12)
(13)

このプログラムでは、訓練データ(X_{train})とその教師データ(y_{train})のうち、(14) % は「評価データ」として学習に使わずにとっておき、残りを使って学習するように設定している。これは誤差を減らしすぎること、かえって未知データに対する予測精度が落ちてしまう (15) と呼ばれる現象を検知し、学習を打ち切る防止策をとるためである。accuracyの値が (16) 回を超えて改善しなければ学習を打ち切る設定になっている。

(14)
(15)
(未選択)

(16)

このCNNは、(17) 個のクラス(カテゴリ)进行分类する。教師信号は、クラス数と同数の0を並べ、各クラスに対応する位置だけを1にしたものが用いられる。この方法を (18) encoding と呼ぶ。

(17)
(18)
(未選択)

ランダムに選んだ一部のニューロンを、0しか出力できないように強制して、学習がよりうまくいくように設定しているのは、行番号 (19) であり、(20) % のニューロンを選ぶ設定になっている。

(19)
(20)

「斜め線の有無」のような画像の局所部分の特徴を抽出する層を設定しているのは、行番号 (21) である。

(21)

未知データに対する予測精度(これを (22) と呼ぶ)を求めているのは、行番号 (23) である。

(22)
(未選択)

(23)

* 回答は自動的に記録されますが、最後に「提出」ボタンをクリックし提出を確認してください。