

問4. 以下の問いに答えよ。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。(20点)

CSV ファイル「car-data.csv」には、さまざまな車の重量、CO2 排出量などに関する情報が含まれている。car-data.csv を読み込んで一連の処理を行うノートブックの __ (1) __ などの空欄を埋め、また問いに答えよ。空欄を埋めるプログラムは1行で答えること。なお、提出するipynb/htmlファイルには、以下に表示されていないプログラム行が含まれていても構わない。

```
import os
os.environ['OMP_NUM_THREADS'] = '1'

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn.preprocessing import scale
```

car-data.csv の全データをデータフレーム df に読み込んであるとする。df の「Weight」列と「CO2」列をデータフレーム dfX に読み込み (1)。

```
dfX = __ (1) __
```

(1)

dfX を標準化し、結果を変数 X_scaled に代入 (2)。

```
X_scaled = __ (2) __
```

(2)

X_scaled をデータフレーム型に変換し、emit_df に代入。列名はdfX と同じにする (3)。

```
emit_df = __ (3) __
```

(3)

emit_df に対して、クラス数4、n_init=6、random_state=12 の KMeans クラスタリングを実行するためのオブジェクト変数 km を作成 (4)(5)。クラスタリングを実行し、結果を clstr に代入 (6)。

```
km = __ (4) __ (n_clusters=__ (5) __, n_init=6, random_state=12)
clstr = km.__ (6) __ (emit_df)
```

(4)

(5)

(6)

このクラスタリングの「distortion(慣性)」を表示する (7)。

```
print(__ (7) __)
```

(7)

クラスタリング結果を emit_df に、「cluster_no」というラベルを持つ新しい列として追加 (8)。

```
__ (8) __
display(emit_df.head())
# value: (9)
```

(8)

クラスタ番号 2 のメンバー数を答えよ (9)。

(9)