Laboratorium: Redukcja wymiarów

1 Cel/Zakres

- Redukcja liczby wymiarów.
- Ocena efektów redukcji wymiarów.

2 Przygotowanie danych

Dane są dwa poniższe wielowymiarowe zbiory danych.

```
from sklearn import datasets
data_breast_cancer = datasets.load_breast_cancer()
```

```
from sklearn.datasets import load_iris
data_iris = load_iris()
```

3 Ćwiczenie

1. Przeprowadź analizę PCA, tak aby tak zredukować liczbę wymiarów dla każdego z w/w zbiorów. Nowa przestrzeń ma pokrywać przynajmniej 90% różnorodności (zmienności) danych i ma mieć jak najmniej wymiarów.
2. Ćwiczenia przeprowadź najpierw na oryginalnych danych, a później na danych przeskalowanych. Porównaj wyniki.

   W podanych zbiorach są istotnie różne zakresy dla poszczególnych cech. Aby je przeskalować aby były porównywalne użyj `StandardScaler()`. Klasa `PCA()` centruje dane automatycznie, ale ich nie skaluje!
3. Utwórz listę z współczynnikami zmienności nowych wymiarów (dla danych przeskalowanych). W przypadku `data_breast_cancer` listę zapisz w pliku Pickle o nazwie `pca_bc.pkl`

   3 pkt.

   W przypadku `data_iris` listę zapisz w pliku Pickle o nazwie `pca_ir.pkl`

   3 pkt.
4. Utwórz naiwną listę indeksów cech (oryginalnych wymiarów), które mają największy udział w nowych cechach (wymiarach), po redukcji (dla danych przeskalowanych). Podpowiedź: zob. atrybut `components_` klasy `PCA`, znajdź indeks z największą wartością dla każdej składowej

(wartości mogą być ujemne!). Uwaga, jest to metoda naiwna, która nie zawsze będzie dawać pożądane efekty np. wtedy gdy różnica pomiędzy największą wartością, a następną będzie niewielka, co oznacza, że więcej niż jedna stara cecha istotnie wpływa na nowy wymiar.

W przypadku `data_breast_cancer` listę zapisz w pliku Pickle o nazwie `idx_bc.pkl`

3 pkt.

W przypadku `data_iris` listę zapisz w pliku Pickle o nazwie `idx_ir.pkl`

3 pkt.

4 Prześlij raport

Prześlij plik o nazwie `lab08/lab08.py` realizujący ww. ćwiczenia.

Sprawdzane będzie, czy skrypt Pythona tworzy wszystkie wymagane pliki oraz czy ich zawartość jest poprawna.