



Fundusze
Europejskie
Polska Cyfrowa



Rzeczpospolita
Polska

Unia Europejska
Europejski Fundusz
Rozwoju Regionalnego



AKADEMIA INNOWACYJNYCH ZASTOSOWAŃ TECHNOLOGII CYFROWYCH (AI TECH)

„Uczenie maszynowe” - laboratorium

Laboratorium 0

wprowadzenie do *python*

data aktualizacji: 20.02.2021

Cel ćwiczenia

Celem ćwiczenia laboratoryjnego jest uruchomienie (wraz z instalacją) środowiska programistycznego języka python oraz narzędzi potrzebnych do realizacji zadań następnych list. W trakcie realizacji zadania wczytane zostaną standardowe zbiory danych, które będą podstawą dokładniejszej analizy. Użyty zostanie algorytm PCA i biblioteki wizualizacji danych.

Dostępność materiałów i narzędzi

Narzędzia oraz ich dokumentacja jest ogólnodostępna w sieci Internet na licencji *opensource*.

Sugerowane narzędzia

- Python w wersji 2.x i/lub 3.x – jako język i środowisko oprogramowania – <https://www.python.org/>
- Jupyter (notebook) – środowisko programowania/generowania dokumentacji – <https://jupyter.org/>

- sci learn – biblioteka python modeli do uczenia maszynowego –
<https://scikit-learn.org/stable/>
- scipy – zbiór bibliotek python do operacji na danych –
<https://www.scipy.org/> Szczególnie przydatne:
 - pandas – struktury danych i analizy – <https://pandas.pydata.org/>
 - numpy – przydatna biblioteka do obliczeń w python –
<https://numpy.org/>
 - matplotlib – biblioteka do wizualizacji (wykresy) w python –
<https://matplotlib.org/stable/>

Mogą być też przydatne:

- seaborn – zaawansowana biblioteka wizualizacji danych –
<https://seaborn.pydata.org/>

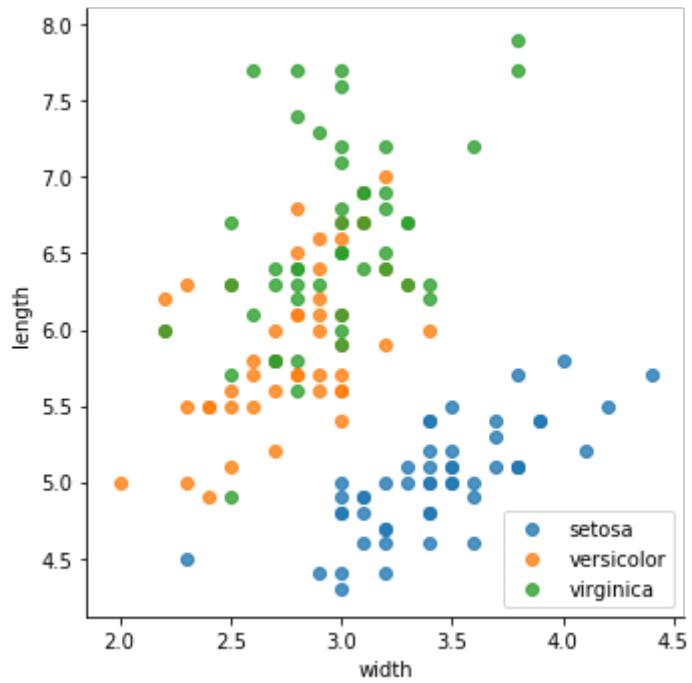
Użyte zbiory danych

W ćwiczeniu użyte będą powszechnie używane zbiory:

- IRIS – <https://archive.ics.uci.edu/ml/datasets/iris>
- GLASS – <https://archive.ics.uci.edu/ml/datasets/glass+identification>
- WINE - <https://archive.ics.uci.edu/ml/datasets/wine>

Przebieg ćwiczenia

1. Instalacja python wraz z niezbędnymi bibliotekami, sugerowane: sci learn, numpy, pandas itp.
2. Instalacja środowiska programistycznego (np. jupyter)
3. Wczytanie zbioru IRIS, WINE, GLASS
4. Analiza zbiorów IRIS, WINE, GLASS: klasy (liczba, interpretacja), instancje, atrybuty, dystrybucja klas w zbiorze.
5. Wyrysowanie wykresu zależności długości/szerokości płatków IRIS a klasą (z kolorem)

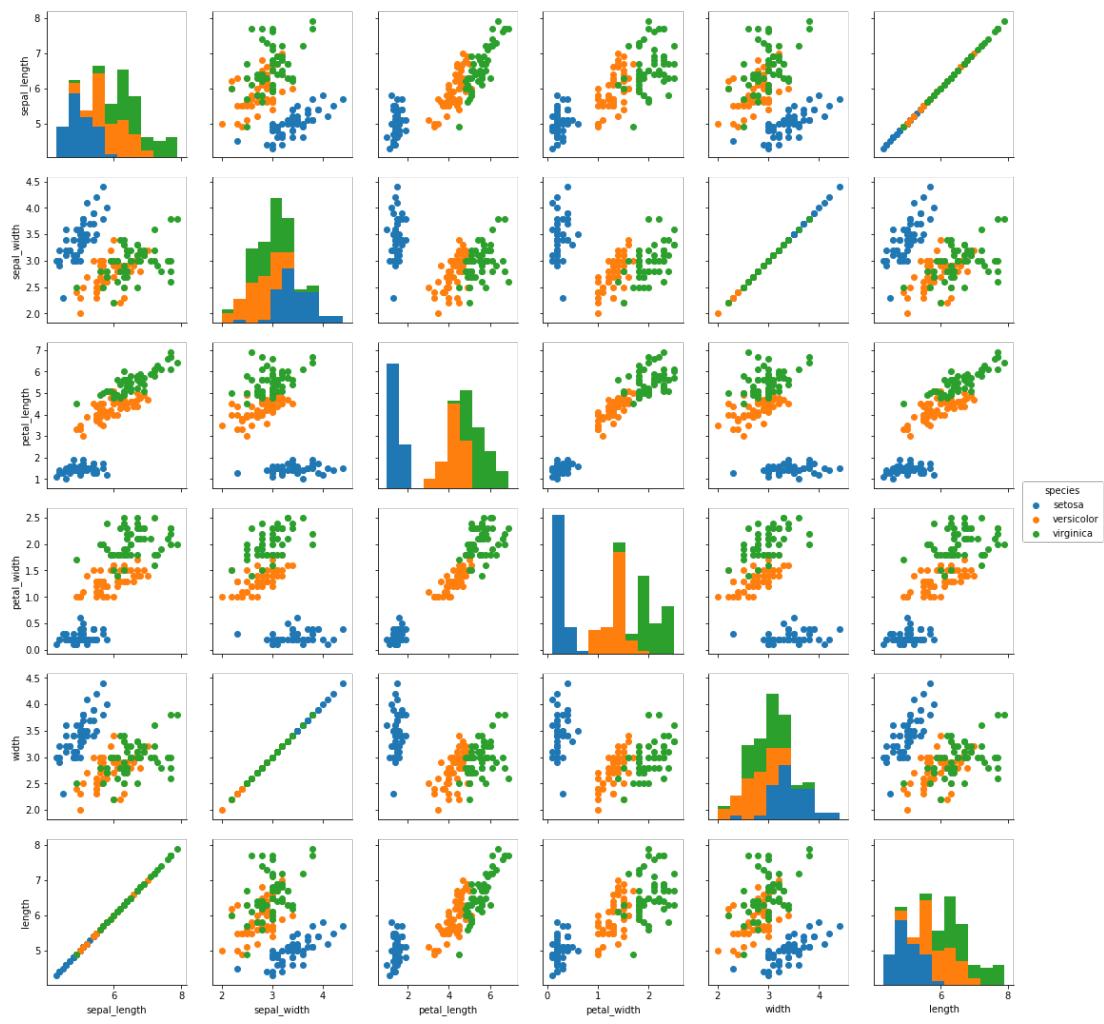


Wykres 1. Zależność długość i szerokości płatka dla IRIS (użyto: seaborn
lmplot

6. Użycie PCA i narysowanie wykresu na wyniku działania PCA

Algorytm PCA (ang. *principal component analysis*) tj. wyznaczania głównych składowych analizowanego zbioru. PCA stosuje się do zmniejszenia wymiarowości zbioru.

Tutorial PCA: <https://towardsdatascience.com/pca-using-python-scikit-learn-e653f8989e60>



Wykres 2. Zależność długość i szerokości płatka dla IRIS (użyto: seaborn PairGrid)

Punktacja

Przy realizacji zadania student może otrzymać **max 5 punktów** wedle poniższej punktacji.

1	Instalacja środowiska z niezbędnymi bibliotekami
1	Wczytanie zbioru IRIS, wyrysowanie wykresu zależności długości/szerokości płatków (jak Wykres 1), Analiza zbioru i wizualizacja rozkładu danych (jak Wykres 2)
1	Wczytanie zbioru GLASS, wyrysowanie wykresu zależności wybranych atrybutów (jak Wykres 1), Analiza zbioru i wizualizacja rozkładu danych (jak Wykres 2)

1	Wczytanie zbioru WINE, wyrysowanie wykresu zależności wybranych atrybutów (jak Wykres 1), Analiza zbioru i wizualizacja rozkładu danych (jak Wykres 2)
1	Użycie PCA i narysowanie wykresu wynikowego dla trzech zbiorów

Przy realizacji tego zadania wystarczy prosty raport PDF utworzony przy użyciu Jupyter.

Pytania pomocnicze

1. Czym się różnią zbiory danych analizowane w treści zadania? Na czym może polegać „trudność” analizy. Który z nich wydaje się być łatwiejszy/trudniejszy?
2. Czy nierównomierny rozkład klas w zbiorze może stanowić problem dla analizy i dalszej budowy modelu danych?
3. Jak działa PCA i kiedy warto go stosować?

Literatura

1. Materiały do wykładu
2. Cichosz P. "Systemy uczące się", WNT Warszawa
3. Zasoby Internetu, słowa kluczowe: uczenie maszynowe (machine learning), data mining, pca

Kontakt

W przypadku pytań/uwag proszę o kontakt: pawel.myszkowski@pwr.edu.pl