

PCA

Materiały dodatkowe

[Practical Guide to Principal Component Analysis \(PCA\) in R & Python](#)

[Principal Component Analysis in 3 steps](#)

[Principal Component Analysis](#)

[3Blue1Brown - Essence of linear algebra Chapter #9 - Dot products and duality](#)

[3Blue1Brown - Essence of linear algebra Chapter #14 - Eigenvectors and eigenvalues](#)

Tasks

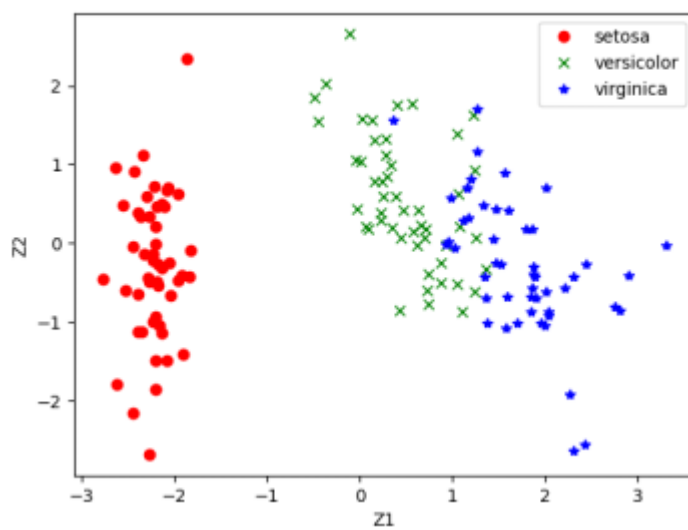
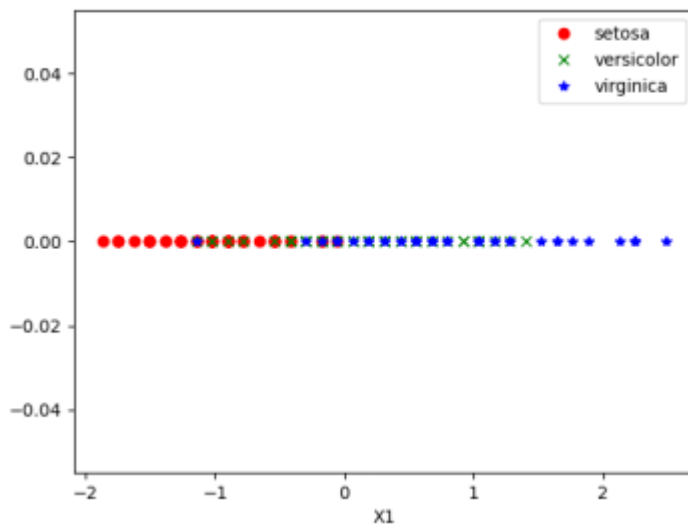
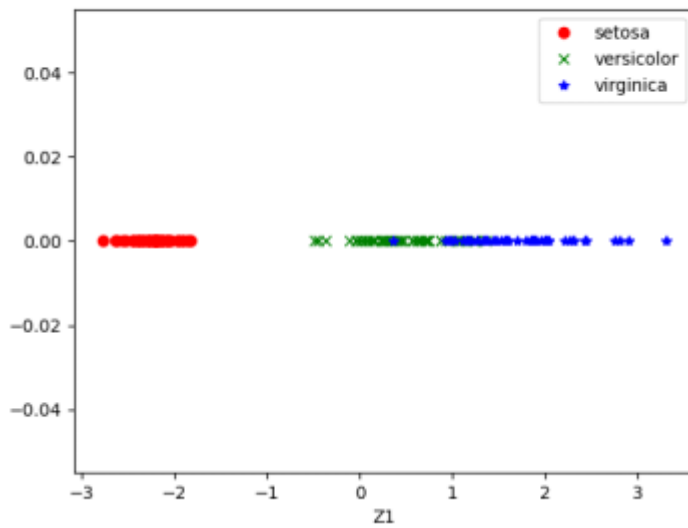
Zadanie 1

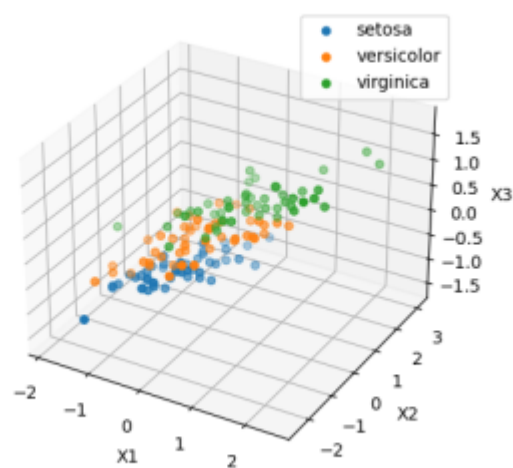
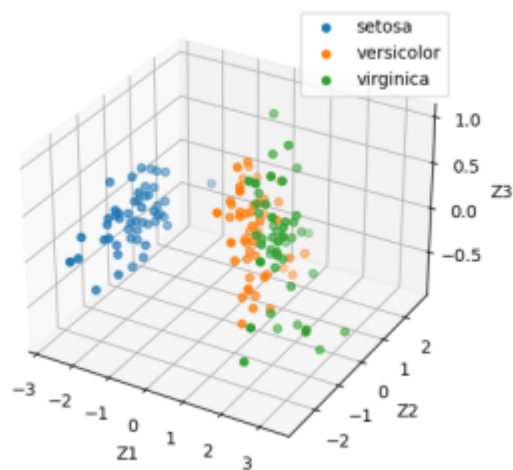
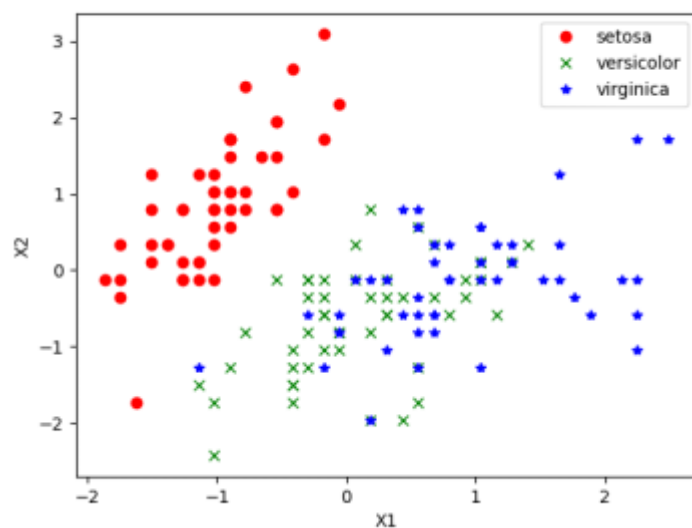
Celem zadania jest wyznaczenie głównych składowych z wykorzystaniem algorytmu PCA.

W celu wizualizacji działania, proszę wybrać jedną z dostępnych [baz](#) w pakiecie scikit-learn (np. bazę [iris](#))

Proszę wykonać następujące kroki zgodnie z procedurą algorytmu PCA:

1. Standaryzacja danych (odjęcie średniej, podzielenie przez odchylenie standardowe - względem cechy).
2. Obliczenie macierzy kowariancji (`numpy.cov()` albo $X.T@X/m$, gdzie X to macierz z danymi postaci $m \times n$: m - ilość przykładów; n - ilość cech).
3. Obliczenie wartości i wektorów własnych (np. z wykorzystaniem `numpy.linalg.eig()`). Wartości własne wskazują "ważność" nowej cechy Z_x (jak wiele informacji zawiera). Wektory własne pozwalają przekształcić dany wektor cech x_i w wektor z_i poprzez kombinację liniową tego wektora x_i z wektorami własnymi macierzy kowariancji X_{cov} . Kolejne elementy danego j -tego wektora własnego wskazują ile danej cechy z wektora x_i jest w j -tej cenie wektora z_i .
4. W celu przekształcenia danych z X na Z wystarczy dokonać działania $Z = X@w$, gdzie w to macierz wektorów własnych ($n \times n$; wiersze przedstawiają kolejne elementy wektorów, natomiast kolumny - kolejne wektory). W szczególności, możemy dokonać kompresji danych eliminując składowe, dla których wartości własne są najmniejsze: w tym celu wystarczy wybrać odpowiednią ilość wektorów własnych k i wykonać mnożenie: $X@w[:, :k]$ - w ten sposób otrzymamy macierz Z o wymiarach $m \times k$, gdzie k to nowy wymiar przestrzeni cech po przekształceniu liniowym i kompresji.
5. Proszę narysować wykres przedstawiający przykłady po przekształceniu i redukcji wymiarowości (1 wymiar, 2 wymiary, 3 wymiary). Cechy wybrać względem największych wartości własnych.
6. Przykładowe wykresy dla danych z bazy *iris* mogą wyglądać jak poniżej (razem z porównaniem z oryginalnymi cechami obciętyymi do wymaganej ilości poprzez odrzucenie tych na końcu):





Zadanie 2

Proszę zapoznać się z instrukcją

PCA

i wykonać zawarte tam zadania.

Do zadań potrzebne będą dwa zbiory danych:

1)

USArrests

2)

NCI60

Etykiety

dla bazy NCI60

W przypadku użycia nowych wersji biblioteki numpy może być konieczność użycia funkcji `df.to_numpy()` zamiast `as_matrix()`

From:

<https://home.agh.edu.pl/~mdig/dokuwiki/> - **MVG Group**

Permanent link:

https://home.agh.edu.pl/~mdig/dokuwiki/doku.php?id=teaching:data_science:ml_pl:topics:pca



Last update: **2023/05/24 06:07**