

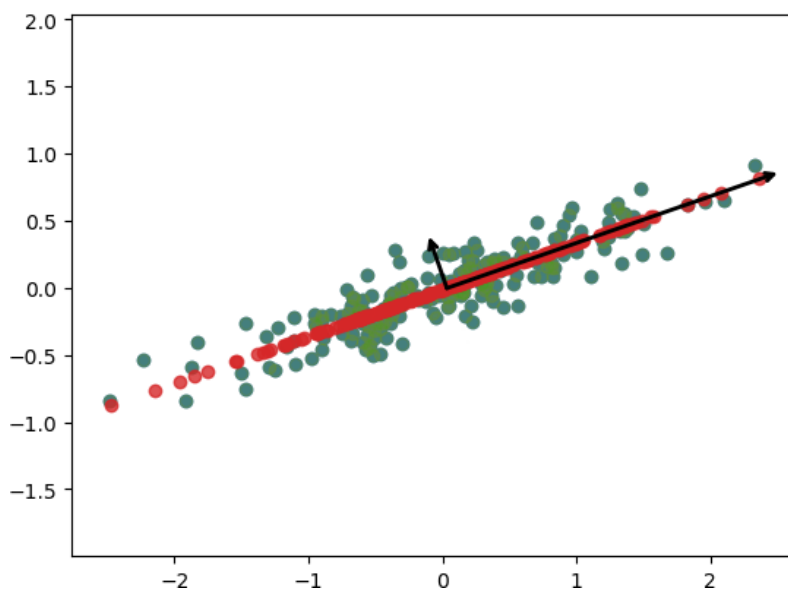
Laboratorium 3 — Transformacje ortogonalne

Celem ćwiczenia jest implementacja algorytmu Analizy Głównych Składowych PCA dla danych wielowymiarowych. Algorytm ten umożliwia redukcję wymiarowości danych wejściowych.

Zadanie 1

Wykonaj własną implementację algorytmu PCA. Funkcja `PCA_fit_transform`, powinna przyjmować na wejście *zbiór danych w formie macierzy numpy*, oraz *liczbę wymiarów po redukcji*. Jeśli liczba wymiarów jest większa niż rozmiar danych wejściowych funkcja powinna zgłosić odpowiedni wyjątek. Jeśli dane wejściowe będą poprawne funkcja powinna zwrócić *macierz w zredukowanej przestrzeni*, *macierz wektorów własnych* oraz *wektor liczb własnych*. Następnie:

1. Wygeneruj w sposób losowy zbiór 200 obiektów dwuwymiarowych za pomocą funkcji `dot` i `rand` lub `randn`;
2. Zwizualizuj obiekty na pomocą funkcji pakietu `matplotlib`, `scatter`;
3. Napisaną własnoręcznie funkcję należy porównać z implementacją dostępną w pakiecie `scikit-learn` (poznana na poprzednich zajęciach). W tym celu, zastosuj obie implementacje na wygenerowanych danych ustawiając liczbę komponentów na 2, porównaj wektory własne, wartości własne oraz wykresy po transformacji;
4. Dokonaj redukcji wymiarowości do jednego wymiaru za pomocą utworzonej funkcji;
5. Zwizualizuj rzut wygenerowanych danych na pierwszą składową oraz zaznacz wektory własne, tak jak na rysunku 1. W tym celu niezbędne będzie wykonanie transformacji odwrotnej.

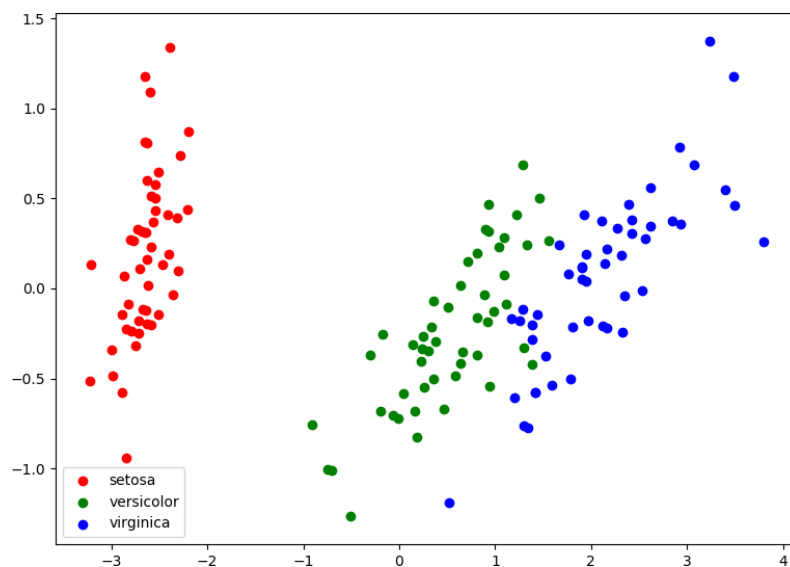


Rysunek 1: Przykładowa wizualizacja przestrzeni cech i wektorów własnych

Zadanie 2

Zaimportuj moduł datasets (`from sklearn import datasets`) i załaduj zbiór iris (`iris = datasets.load_iris()`), a następnie:

1. Dokonaj redukcji wymiarowości wszystkich obiektów w zbiorze do 2 najbardziej znaczących wymiarów za pomocą opracowanej funkcji;
2. Zwizualizuj elementy zbioru w przestrzeni cech z oznaczonymi klasami, np. za pomocą kolorów, etykiet lub symboli.

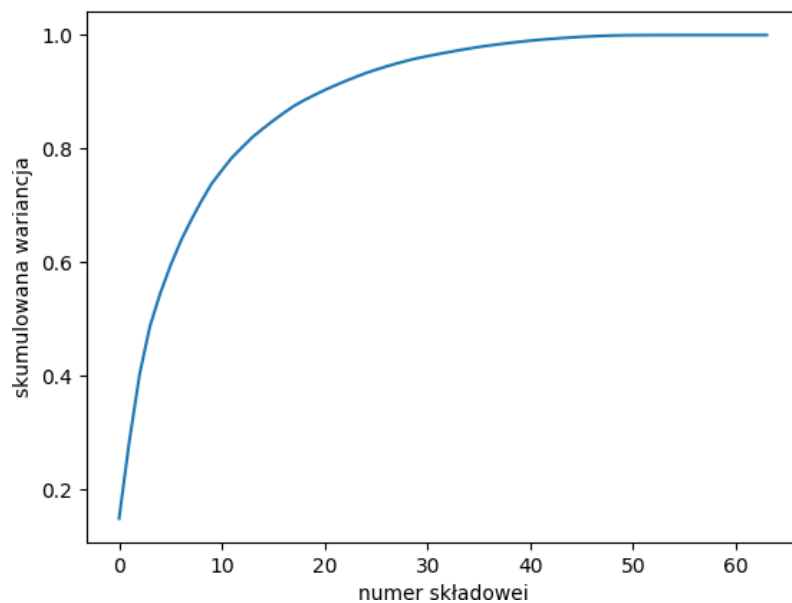


Rysunek 2: Przykładowa wizualizacja obiektów z bazy iris

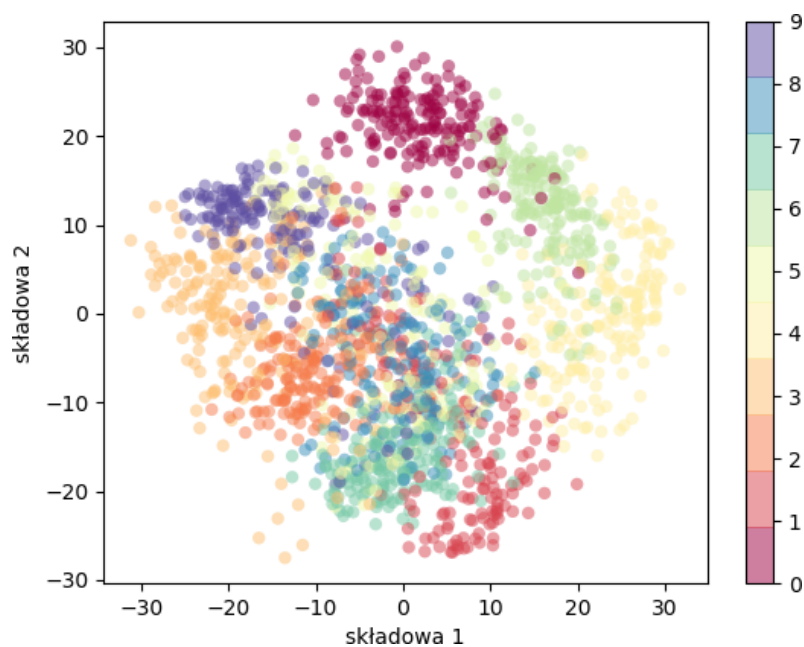
Zadanie 3

Załaduj zbiór digits, a następnie:

1. Dokonaj redukcji wymiarowości wszystkich obiektów w zbiorze do 2 najbardziej znaczących wymiarów za pomocą opracowanej funkcji;
2. Pokaż krzywą wariancji dla rosnącej liczby składowych głównych (tak jak na Rysunku 3)
3. Zwizualizuj elementy zbioru w przestrzeni cech z oznaczonymi klasami (podobnie do Rysunku 4)



Rysunek 3: Przykładowa wizualizacja wariancji składowych głównych dla bazy digits



Rysunek 4: Przykładowa wizualizacja obiektów z bazy digits