

Zadanie: PCA – Zbiór Wine

1 Cel

Zastosuj PCA do zbioru Wine w celu redukcji wymiarowości i analizy struktury danych.

2 Dane

- Źródło: `sklearn.datasets.load_wine()`
- Próbki: 178, Cechy: 13, Klasy: 3

Nr	Cecha (ang.)	Cecha (pol.)	Opis
1	Alcohol	Alkohol	Zawartość alkoholu (%)
2	Malic acid	Kwas jabłkowy	Stężenie kwasu
3	Ash	Popiół	Zawartość popiołu
4	Alcalinity of ash	Zasadowość popiołu	pH popiołu
5	Magnesium	Magnez	Zawartość magnezu
6	Total phenols	Fenole całkowite	Związki fenolowe
7	Flavanoids	Flawonoidy	Rodzaj fenoli
8	Nonflavanoid phenols	Fenole nieflaw.	Inne fenole
9	Proanthocyanins	Proantocyjaniny	Barwniki
10	Color intensity	Intensywność barwy	Głębokość koloru
11	Hue	Odcień	Ton barwy
12	OD280/OD315	OD280/OD315	Stosunek absorbancji
13	Proline	Prolina	Aminokwas

3 Polecenia

3.1 Część 1: Przygotowanie danych

1. Załaduj zbiór danych Wine.
2. Wyświetl: liczbę próbek i cech, nazwy cech, rozkład klas.
3. Przeprowadź standaryzację. **Wyjaśnij**, dlaczego jest konieczna przed PCA.

3.2 Część 2: Wykonanie PCA

1. Wykonaj PCA (wszystkie składowe).
2. Wyświetl: wartości własne, % wyjaśnionej wariancji, skumulowaną wariancję.
3. Ile składowych potrzeba dla $\geq 80\%$ wariancji?

3.3 Część 3: Wizualizacja

- 3.1. Scree Plot – zaznacz “łokieć”.
- 3.2. Projekcja 2D (PC1 vs PC2) z kolorami klas.
- 3.3. Biplot – punkty danych + wektory cech.

3.4 Część 4: Interpretacja

- 4.1. Wyświetl ładunki (loadings) dla PC1 i PC2.
- 4.2. Odpowiedz:
 - Które cechy dominują w PC1? Co reprezentuje ta składowa?
 - Które cechy dominują w PC2?
 - Czy klasy są dobrze separowalne w 2D?

Wyjaśnij różnice – która cecha zdominowała analizę bez standaryzacji?

4 Wskazówki

1. Standaryzacja wyrównuje wpływ cech o różnych skalach.
2. Szukaj “łokcia” na Scree Plot lub progu 80–95% wariancji.
3. W biplotie przeskaluj wektory cech dla widoczności.
4. Cechy w tym samym kierunku są skorelowane.