

Laboratorium 1

Autorzy: Marcin Aman, Bartłomiej Tonia, Miłosz Janowski

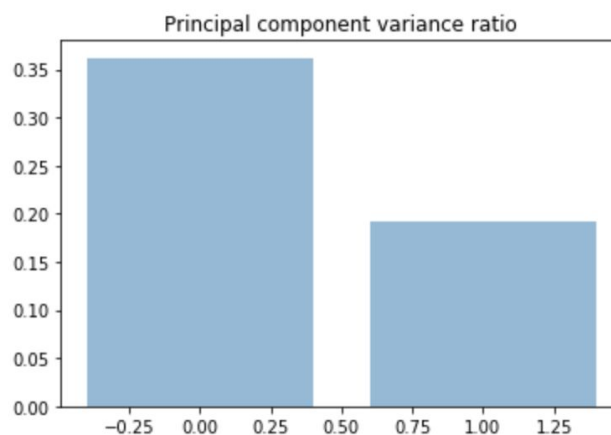
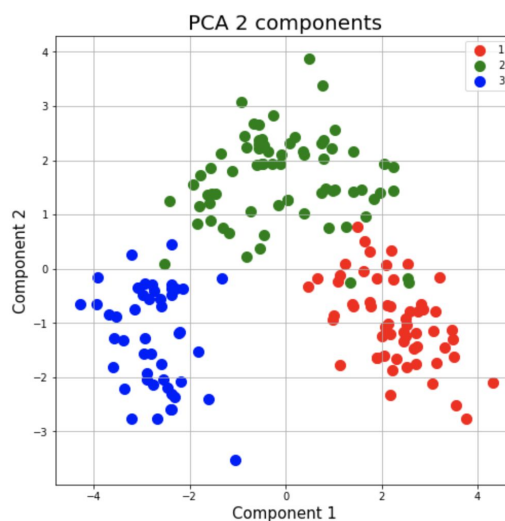
1.

Opis zbioru wine:

Cechy(**13**): Alcohol, Malic acid, Ash, Alcalinity of ash, Magnesium, Total phenols, Flavanoids, Nonflavanoid phenols, Proanthocyanins, Color intensity, Hue, OD280/OD315 of diluted wines, Proline

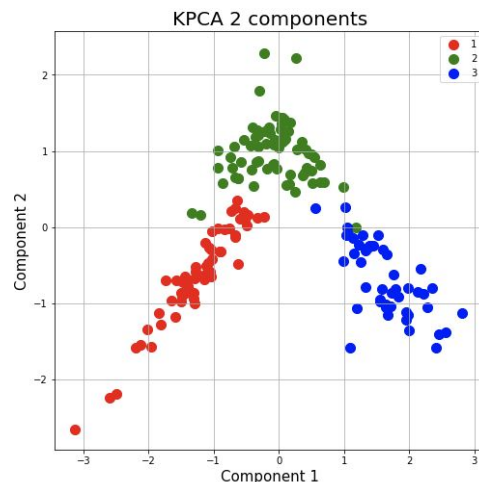
Elementy(**178** w **3** klasach): klasa 1 - 59, klasa 2 - 71, klasa 3 - 48

Charakterystyka zbioru:



Przy użyciu PCA można zauważyć, iż zbiór ma bardzo zauważalną granicę między kolorem zielonym a niebieskim, jednak granica zielony-czerwony jest mniej widoczna. Część elementów ze zbiorów miesza się ze sobą przez co ciężko jest jednoznacznie oddzielić poszczególne części.

Ciekawe są natomiast wyniki KPCA z kernelem poly:



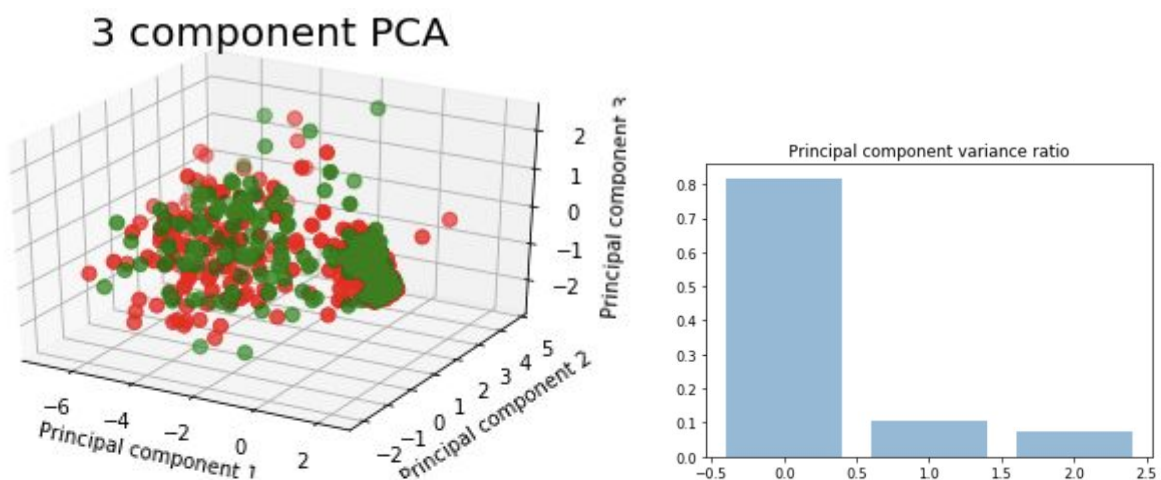
Można tutaj zauważyć bardzo dobre zgrupowanie danych oraz ich jasne oddzielenie w poszczególnych grupach. Efekt jest znacznie lepszy niż przy wizualizacji przez PCA lub inne kernele.

2.

Kolejnym zbiorem jaki należało zwizualizować przy użyciu tej metody jest breast-cancer. W tym przypadku mamy do dyspozycji 9 cech takich jak: 'Clump Thickness', 'Uniformity of Cell Size', 'Uniformity of Cell Shape', 'Marginal Adhesion', 'Single Epithelial Cell Size', 'Bare Nuclei', 'Bland Chromatin', 'Normal Nucleoli', 'Mitoses'.

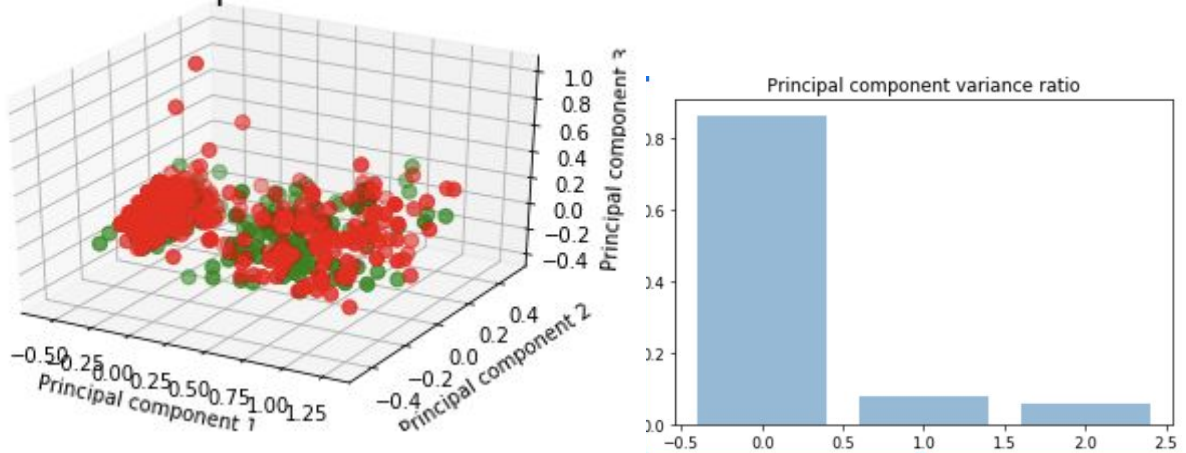
Składa się on z 699 elementów w 2 klasach: klasa 2 - 414, klasa 4 - 235.

W przypadku tego zbioru użyliśmy Kernel PCA a tytułowym kernelem był w pierwszej kolejności linear:



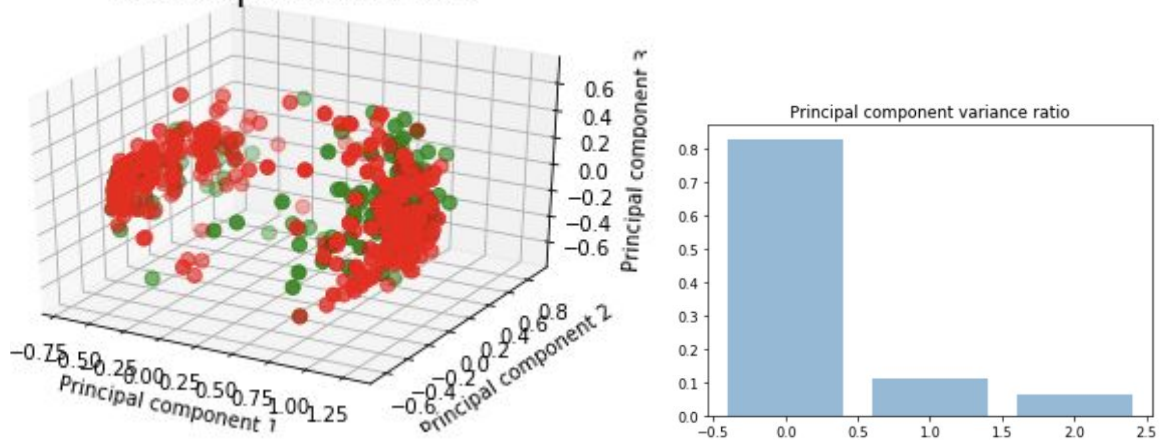
Sama wariancja była jedną z najwyższych, podobne wartości uzyskiwał jedynie kernel sigmoid:

3 component PCA



Kernel cosine też dawał nie najgorsze rezultaty:

3 component PCA



3.

Kolejnym zadaniem była analiza zbioru zawierającego twarze.

Zbiór składa się ze zdjęć 5000 twarzy, każda opisana 1024 pikselami w skali szarości.

Dokonałymi PCA do dwóch wymiarów i przedstawiliśmy 100 losowo wybranych twarzy na wykresie.



Można zauważyć na przykład, że po prawej stronie są zdjęcia w ciemniejszej tonacji, po lewej zaś jaśniejsze. Twarze często grupują się pod względem pozy.

Wnioski:

- W KPCA należy eksperymentować z różnymi kernelami, wtedy daje ono bardzo różniące się rezultaty
- Im większe współczynniki wariancji dla głównych składowych, tym lepsze wyniki daje PCA
- Jeżeli separacja nie jest wystarczająca w 2D warto spróbować PCA w 3D