

PCA visualization and PCA-kernel

Piotr Rząsa, Paweł Banach, Michał Kawalek

PCA - czym jest? do czego służy?

Principal Component Analysis - Analiza Składowych Głównych

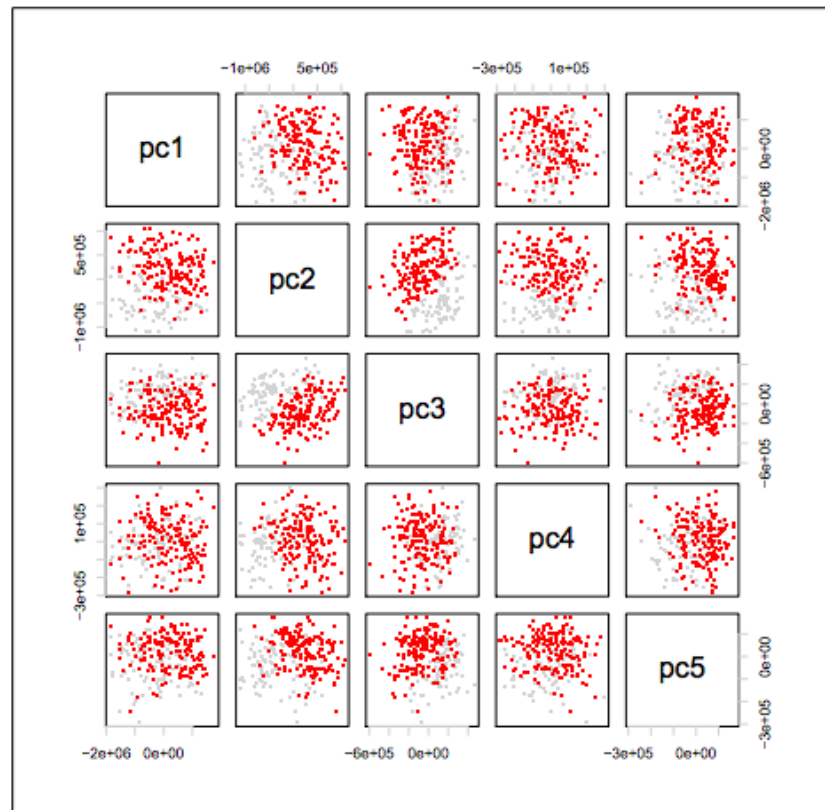
Zastosowania:

- Wizualizacja wielowymiarowych danych w np. 2-, 3-wymiarowej przestrzeni.
- Redukcja wymiarów danych wielowymiarowych
- Takie rzutowanie danych aby osiągnąć jak największą wariancję głównych składowych
- oparta na macierzy kowariancji

Przykładowe zastosowanie

- Rak jajnika

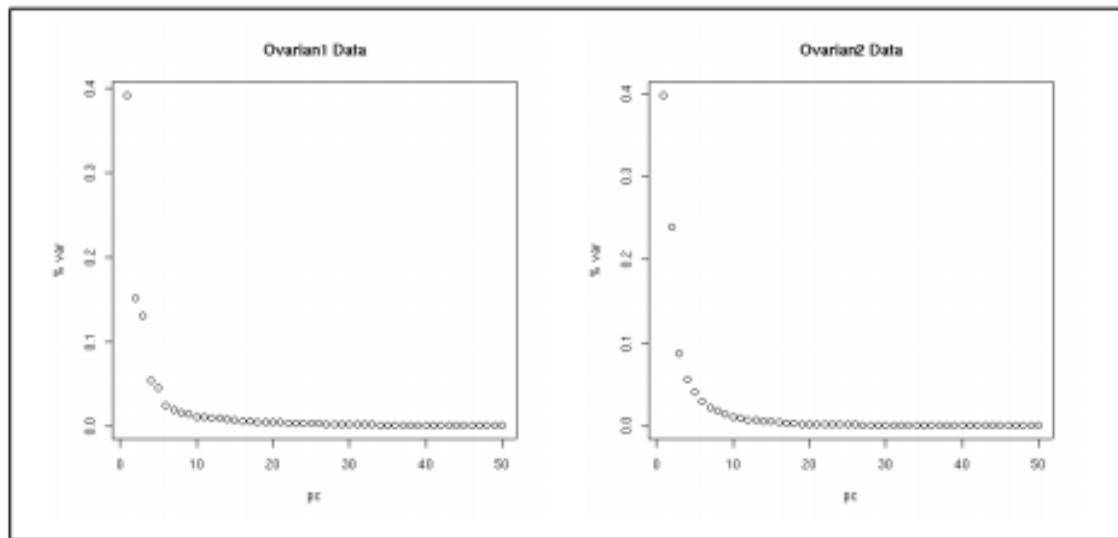
- wybór 5 głównych składowych
- rzut na płaszczyznę wyznaczoną przez ich pary
- dzięki parze składowych nr 2 i 3 można przeprowadzić klasyfikację.



Rysunek 1: Analiza składowych głównych dla zbioru danych z Clinical Proteomics Program Databank dotyczących raka jajnika. Rzutowanie obserwacji na układy współrzędnych wyznaczone przez pary pierwszych pięciu głównych składowych. Czerwone punkty odpowiadają chorym pacjentom, szare zdrowym. W tym przypadku rzutowanie na drugą i trzecią główną składową dobrze rozdziela obserwacje z różnych klas.

Przykładowe zastosowanie (2)

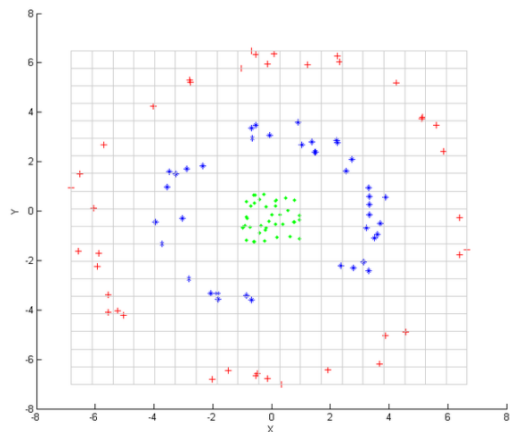
- wariancja skupiona w kilku pierwszych głównych składowych
- dalsze składowe można częściowo pominąć - mają znikomy wpływ na wynik



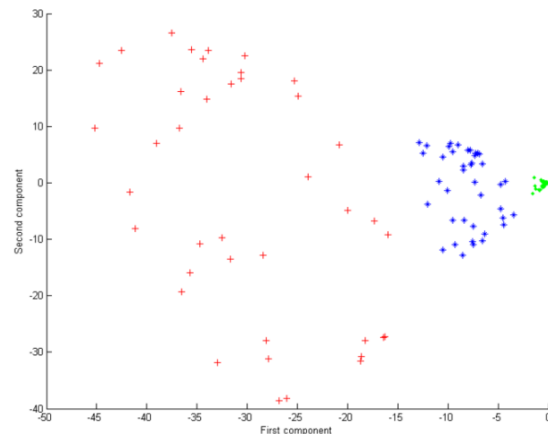
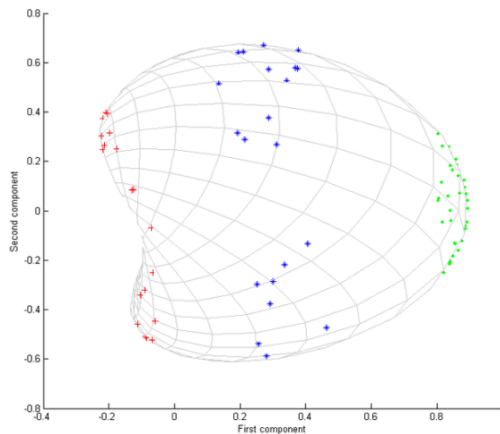
Rysunek 2: Procent wariancji danych wzdłuż kolejnych głównych składowych dla dwóch zbiorów danych z Clinical Proteomics Program Databank. Prawie cała wariancja jest zawarta w kilku pierwszych głównych składowych.

Kernel-PCA

- W przypadku gdy dane nie są liniowo podzielne PCA nam nie pomoże
- Kernel-PCA pozwala na rzutowanie danych na wyższe wymiary przy wykorzystaniu kernel methods
- Na przykład: dane początkowo dwuwymiarowe lecz niemożliwe do prostej klasyfikacji



a



Użyta biblioteka python - sklearn

Korzystamy z zawartego w bibliotece modułu 'decomposition'

<http://scikit-learn.org/stable/modules/classes.html#module-sklearn.decomposition>

A w szczególności z procedury:

```
sklearn.decomposition.PCA
```

Szczegółów można poszukać w dokumentacji scikit-learn.

Źródła i przydatne linki

Step by step tutorial:

http://sebastianraschka.com/Articles/2014_pca_step_by_step.html

PCA Irysy w 2D: [http://scikit-](http://scikit-learn.org/stable/auto_examples/decomposition/plot_pca_vs_lda.html)

[learn.org/stable/auto_examples/decomposition/plot_pca_vs_lda.html](http://scikit-learn.org/stable/auto_examples/decomposition/plot_pca_vs_lda.html)

Kernel PCA: [http://scikit-](http://scikit-learn.org/stable/auto_examples/decomposition/plot_kernel_pca.html)

[learn.org/stable/auto_examples/decomposition/plot_kernel_pca.html](http://scikit-learn.org/stable/auto_examples/decomposition/plot_kernel_pca.html)

PCA irysy w 3D: [http://scikit-](http://scikit-learn.org/stable/auto_examples/decomposition/plot_pca_iris.html)

[learn.org/stable/auto_examples/decomposition/plot_pca_iris.html](http://scikit-learn.org/stable/auto_examples/decomposition/plot_pca_iris.html)