



Wstęp do chemometrii. Podstawy metody PCA



Definicja chemometrii

„Chemometria jest dziedziną nauki i techniki zajmującą się wydobywaniem użytecznej informacji z wielowymiarowych danych pomiarowych, wykorzystującą metody statystyki i matematyki.”¹

¹⁾ Jan Mazerski *Podstawy chemometrii*, Wydawnictwo Politechniki Gdańskiej, 2000



zmienne
próbk

zmienne

próbk

- wiarygodne wyniki można uzyskać już na podstawie 20 ÷ 30 próbek i będą to wyniki oparte na całym lub prawie całym zasobie informacji zawartym w próbkach
- możliwa jest jednoczesna analiza dużej liczby zmiennych



Analiza składowych głównych (PCA)

PCA - **P**roincipal **C**omponent **A**nalysis (Analiza głównych składowych)

„Statystyka daje informacje o podobieństwie wśród badanych próbek, PCA – o różnicach wśród próbek”

Przykład: Kwestionariusz z pytaniami – 50 ludzi (próbek), 100 pytań (zmiennych)
Chcemy wiedzieć, w odpowiedziach na które pytania ludzie najbardziej się różnią
Statystyka nie da odpowiedzi na to pytanie!



**To może posłużyć do
klasyfikacji – można
podzielić ludzi na grupy w
zależności od poglądów**



Analiza składowych głównych (PCA)

Mieliśmy do czynienia z danymi, w których zmienne objaśniające nie są od siebie niezależne. Oznacza to, że tylko część informacji wnoszonej przez każdą ze zmiennych jest swoista, a reszta jest powtórzeniem informacji wnoszonej przez inne zmienne. W takiej sytuacji można się pokusić o przekształcenie zmiennych pierwotnych w nowe, wzajemnie **ortogonalne** zmienne zwane czynnikami.



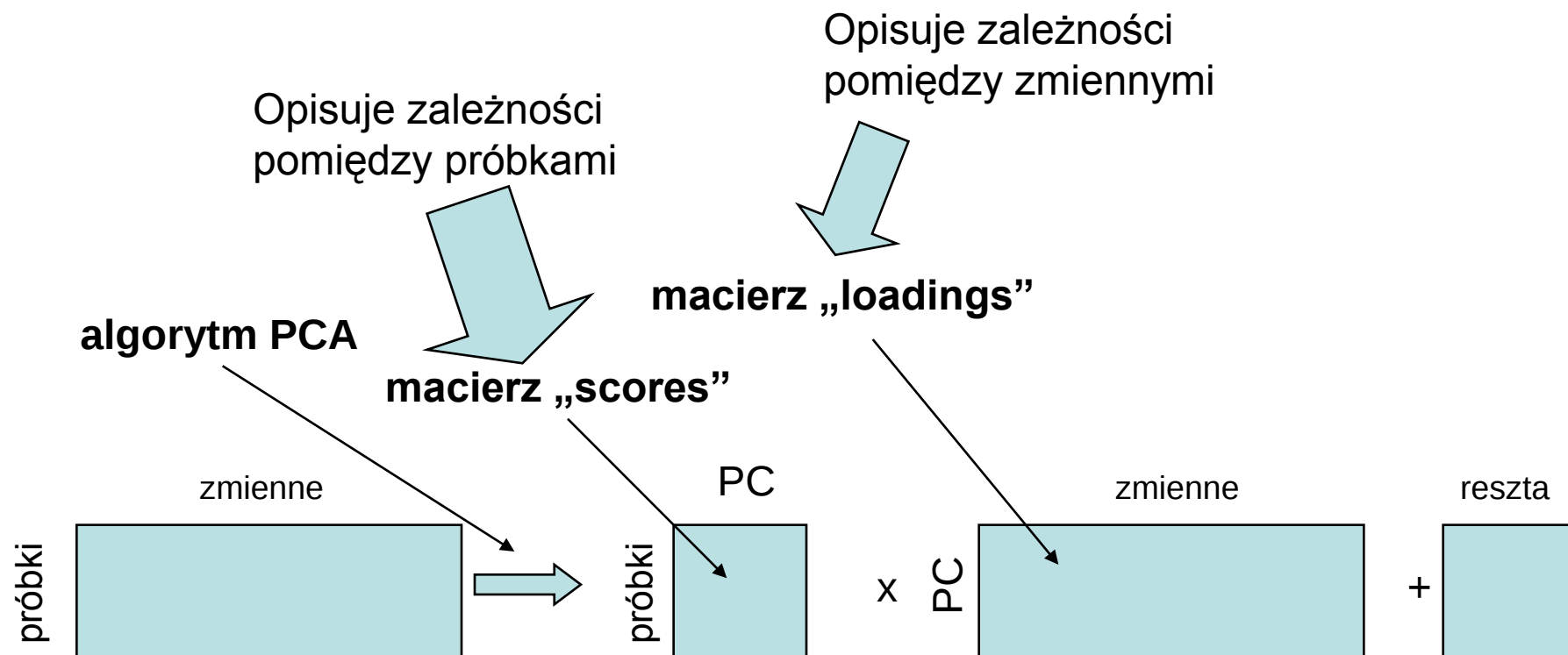
czyli już wzajemnie
niezależne

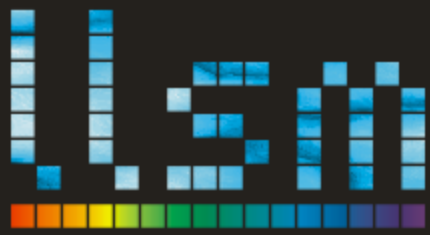
Zastosowanie metody głównych składowych (PCA) w ramach analizy zależności obejmuje dwa podstawowe zadania:

- prezentację graficzną zależności wielowymiarowych
- redukcję wymiarowości problemu



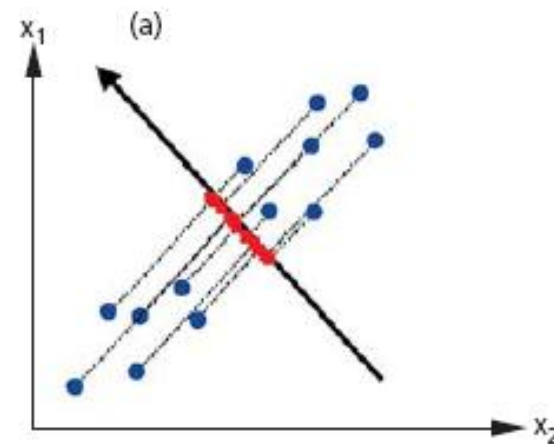
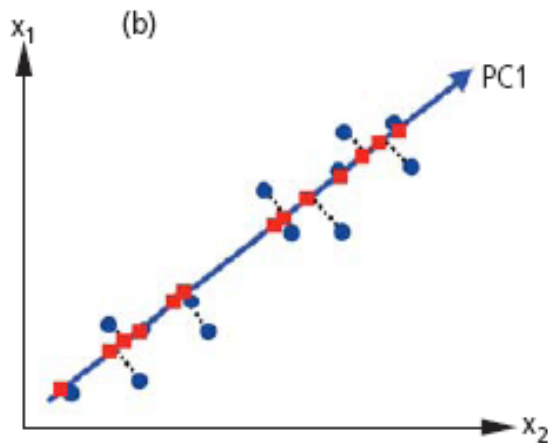
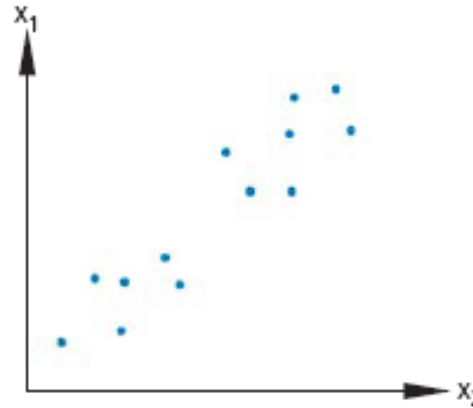
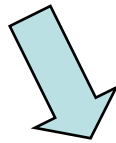
Analiza składowych głównych (PCA)





Analiza składowych głównych (PCA)

Rzuty próbek na główne składowe (PCs) – elementy macierzy wyników (scores)

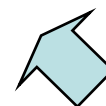




Analiza składowych głównych (PCA)

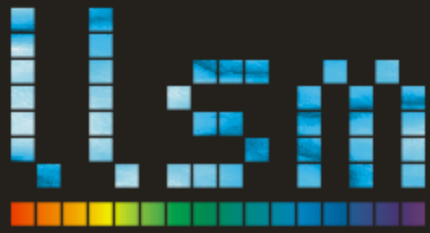
Przygotowanie danych wejściowych:

- dane muszą być wycelowane (średnia równa zero)
- często stosuje się autoskalowanie (średnia równa zero i odchylenie standardowe równe jeden)



Uwaga na szumy, duchy, sklejki itp.!

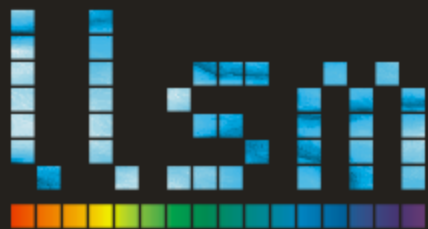
Gdy obie zmienne są istotne, a mają różną zmienność, różne rzędy wielkości, różne jednostki



Analiza składowych głównych (PCA)

Wybór istotnej liczby składowych

- procent opisanej wariancji
- rysunek wartości własnych
- krosvalidacja



Jan Mazerski *Podstawy chemometrii*, Wydawnictwo Politechniki Gdańskiej, 2000

www.chemometria.us.edu.pl – dr hab. Michał Daszykowski

Prof. Max Diem, Northeastern University Boston

Prof. D.L. Massart, Department of Pharmaceutical and Biomedical Analysis, Vrije Universiteit Brussel, Brussels, Belgium

Principal Component Analysis in Spectroscopy, University of Otago, New Zealand:

<http://neon.otago.ac.nz/chemlect/chem306/pca/index.html>

Andrzej Stanisław *Przystępny kurs statystyki. Tom3. Analizy wielowymiarowe*, Kraków 2007

<http://www.statsoft.pl/textbook/stathome.html>