

Metody statyczne selekcji i konstruowania cech dla wizualizacji danych wielowymiarowych

Jerzy Stefanowski

(PCA – Robert Susmaga)

Instytut Informatyki Politechniki Poznańskiej

Wykład InfoVis dla TWO 2015

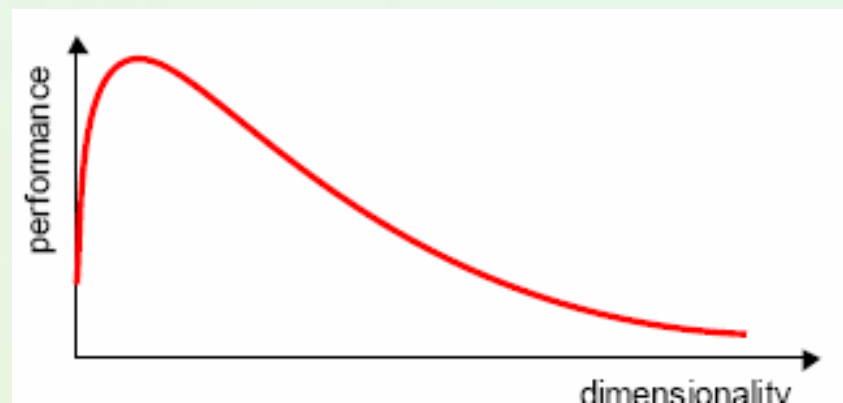
Agregacje i przekształcenia atrybutów

- Połącz dwa lub więcej atrybutów (zmiennych), skonstruuj nowe atrybuty albo dokonaj selekcji atrybutów
- Zmniejsz rozmiary danych (zachowując podstawowe właściwości)
- Cele
 - Wizualizacja w mniejszych przestrzeniach → lepsza percepcja człowieka
 - Mogą zwiększyć czytelność reprezentacji
 - Zmniejszenie wymagań dot. zbierania danych
 - Inne transformacje samych atrybutów
 - Zmień skale / granulacje informacji
 - Cities aggregated into regions, states, countries,

Przekleństwo wymiarowości

„Curse of dimensionality” [Bellman 1961]

- W celu dobrego przybliżenia funkcji (także stworzenia klasyfikatora) z danych:
 - The number of samples required per variable increases exponentially with the number of variables
 - Liczba obserwacji żądanych w stosunku do zmiennej wzrasta wykładniczo z liczbą zmiennych
- Oznacza to konieczność zdecydowanego wzrostu niezbędnych obserwacji przy dodawaniu kolejnych wymiarów



Przykład problemu [D.Mladenic 2005]

F ₁	F ₂	F ₃	F ₄	F ₅	C
0	0	1	0	1	0
0	1	0	0	1	1
1	0	1	0	1	1
1	1	0	0	1	1
0	0	1	1	0	0
0	1	0	1	0	1
1	0	1	1	0	1
1	1	0	1	0	1

- Data set
 - Five Boolean features
 - $C = F_1 \vee F_2$
 - $F_3 = \neg F_2, F_5 = \neg F_4$
 - Optimal subset:
 $\{F_1, F_2\}$ or $\{F_1, F_3\}$
- optimization in space of all feature subsets (2^F possibilities)

(tutorial on genomics [Yu 2004])

Różne podejścia do redukcji wymiarów

- **Selekcja cech**, zmiennych (cech, atrybutów, ..)
 - Wybierz podzbiór zmiennych $F' \subset F$
- **Konstrukcja** nowych zmiennych
 - Metody projekcji – nowe cechy zastępują poprzednie;
 - Statystyczne PCA, MDS vs. sztuczne sieci neuronowe (SOM, GN, ...)
- Wykorzystanie wiedzy dziedzinowej
 - Wprowadzenie nowych cech (oprócz istniejących)
 - Indukcja konstruktywna (R. Michalski)

Selekcja cech vs. konstrukcja nowych cech

Za A.Berge

- In general - two approaches for dimensionality reduction
 - Feature selection: choose a subset of the features

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} \longrightarrow \begin{bmatrix} x_{i_1} \\ x_{i_2} \\ \vdots \\ x_{i_m} \end{bmatrix}$$

- Feature extraction: create a subset of new features by combining existing features

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} \longrightarrow \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = f \left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{bmatrix} \right)$$

Redukcja rozmiarów danych – Selekcja atrybutów

- Dany jest n elementowy zbiór przykładów (obiektów). Każdy przykład x jest zdefiniowany na $V_1 \times V_2 \times \dots \times V_m$ gdzie V_i jest dziedziną i -tego atrybutu. W przypadku nadzorowanej klasyfikacji przykłady zdefiniowane są jako $\langle x, y \rangle$ gdzie y określa pożądaną odpowiedź, np. klasyfikację przykładu.
- **Cel selekcji atrybutów:**
 - *Wybierz minimalny podzbiór atrybutów, dla którego rozkład prawdopodobieństwa różnych klas obiektów jest jak najbliższy oryginalnemu rozkładowi uzyskanemu z wykorzystaniem wszystkich atrybutów.*
- **Nadzorowana klasyfikacja**
 - *Dla danego algorytmu uczenia i zbioru uczącego, znajdź najmniejszy podzbiór atrybutów dla którego system klasyfikujący przewiduje przydział obiektów do klas decyzyjnych z jak największą trafnością.*

Selekcja w trakcie wstępnego przetwarzania danych

- Ocena pojedynczych atrybutów:
 - testy χ^2 i miary siły związku,
 - miary wykorzystujące względną entropię między atrybutem warunkowym a decyzyjnym (ang. *info gain*, *gain ratio*),
 - ...
- Ocena podzbiorów atrybutów (powinny być niezależne wzajemnie a silnie zależne z klasyfikacją):
 - Miara korelacji wzajemnych,
 - Statystyki λ Wilksa, T2-Hotellinga, odległości D2 Mahalanobisa,
 - Redukty w teorii zbiorów przybliżonych,
 - Techniki dekompozycji na podzbiory (ang. *data table templates*)
 - ...
- Model „filter” vs. „wrapper”

WEKA – attribute selection

The screenshot displays the WEKA Explorer interface with the 'Select attributes' tab selected. The 'Attribute Evaluator' section is active, showing a list of evaluators in the left pane and a 'Choose' button with 'ChiSquaredAttributeEval' selected in the right pane. Below the 'Attribute Evaluator' section, the 'Search Method' section is visible, showing a list of search methods in the right pane.

Weka Explorer

Preprocess | Classify | Cluster | Associate | **Select attributes** | Visualize

Attribute Evaluator

- weka
 - attributeSelection
 - CfsSubsetEval
 - ClassifierSubsetEval
 - WrapperSubsetEval
 - ConsistencySubsetEval
 - ReliefFAttributeEval
 - InfoGainAttributeEval
 - GainRatioAttributeEval
 - SymmetricalUncertAttributeEval
 - OneRAttributeEval
 - ChiSquaredAttributeEval**
 - PrincipalComponents
 - SVMAttributeEval

Choose **ChiSquaredAttributeEval**

Search Method

- weka
 - attributeSelection
 - BestFirst
 - ForwardSelection
 - RaceSearch
 - GeneticSearch
 - RandomSearch
 - ExhaustiveSearch
 - Ranker**
 - RankSearch

Ranking with ...? WEKA

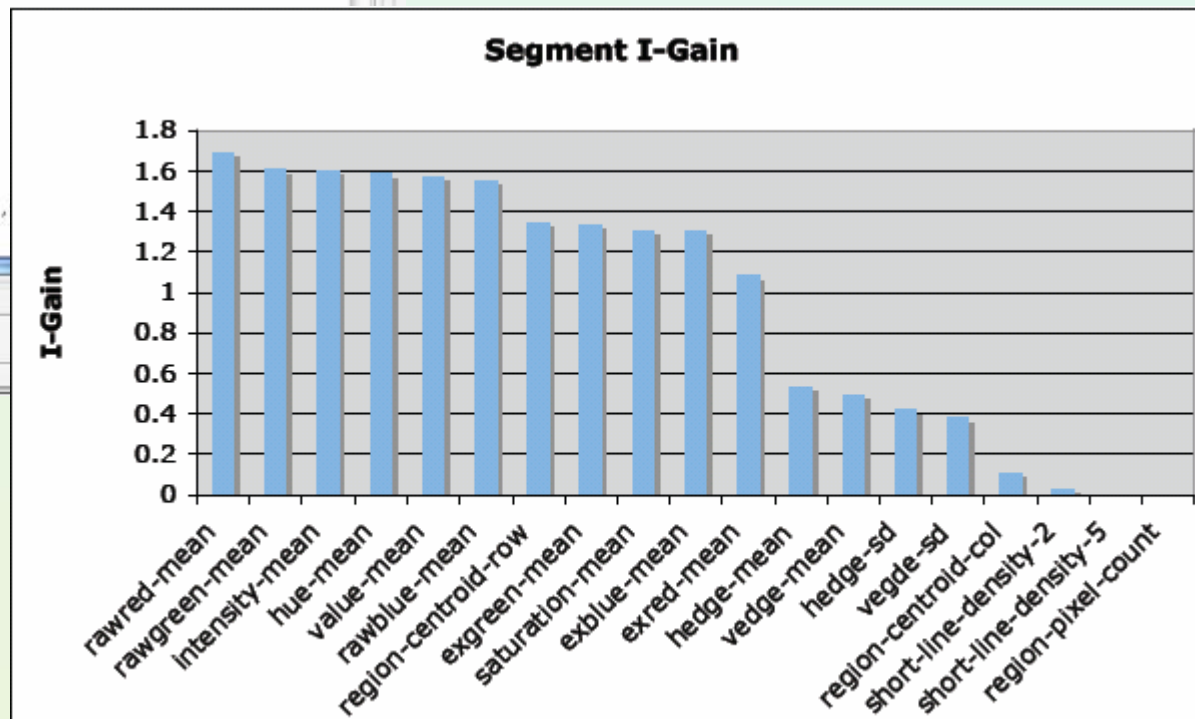
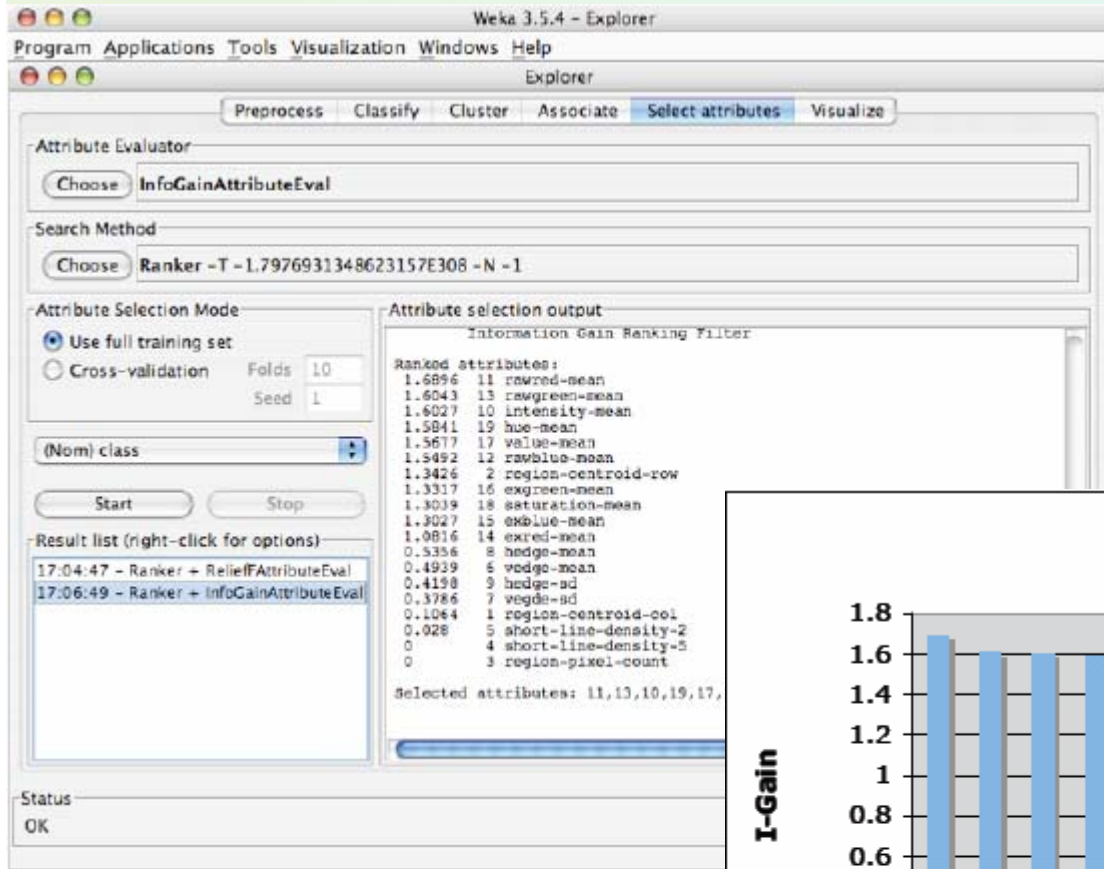
The screenshot shows the Weka Explorer interface with the 'Select attributes' tab active. The 'Attribute Evaluator' is set to 'ChiSquaredAttributeEval' and the 'Search Method' is 'Ranker -T -1.7978931348623157E308 -N -1'. The 'Attribute Selection Mode' is 'Use full training set' with 10 folds and seed 1. The 'Attribute selection output' pane displays the following text:

```
A9:  
D1:  
Evaluation mode: evaluate on all training data  
  
=== Attribute Selection on all input data ===  
  
Search Method:  
Attribute ranking.  
  
Attribute Evaluator (supervised, Class (nominal): 9 D1:):  
Chi-squared Ranking Filter  
  
Ranked attributes:  
71.9035 2 A3:  
68.5634 1 A2:  
67.8595 4 A5:  
67.629 8 A9:  
64.2122 7 A8:  
64.0766 3 A4:  
18.9905 5 A6:  
14.0986 6 A7:  
  
Selected attributes: 2,1,4,8,7,3,5,6 : 8
```

The 'Result list' at the bottom shows a single entry: '21:37:48 - Ranker + ChiSquaredAttributeEval'.

Jak wykorzystać ranking atrybutów

- Wybierz powyżej progu τ
- Mediana czy inny?



Przykład selekcji cech w eksploracji danych

- Rozważmy rzeczywisty przykład z zakresu diagnostyki technicznej
 - Autosan H9-21 (prototyp 1969, produkowany lata 73-84) w Sanockiej Fabryce Autobusów
- Metodycznie – klasyfikacja nadzorowana
- Lecz istotne pytanie o ważność cech i możliwość ich selekcji



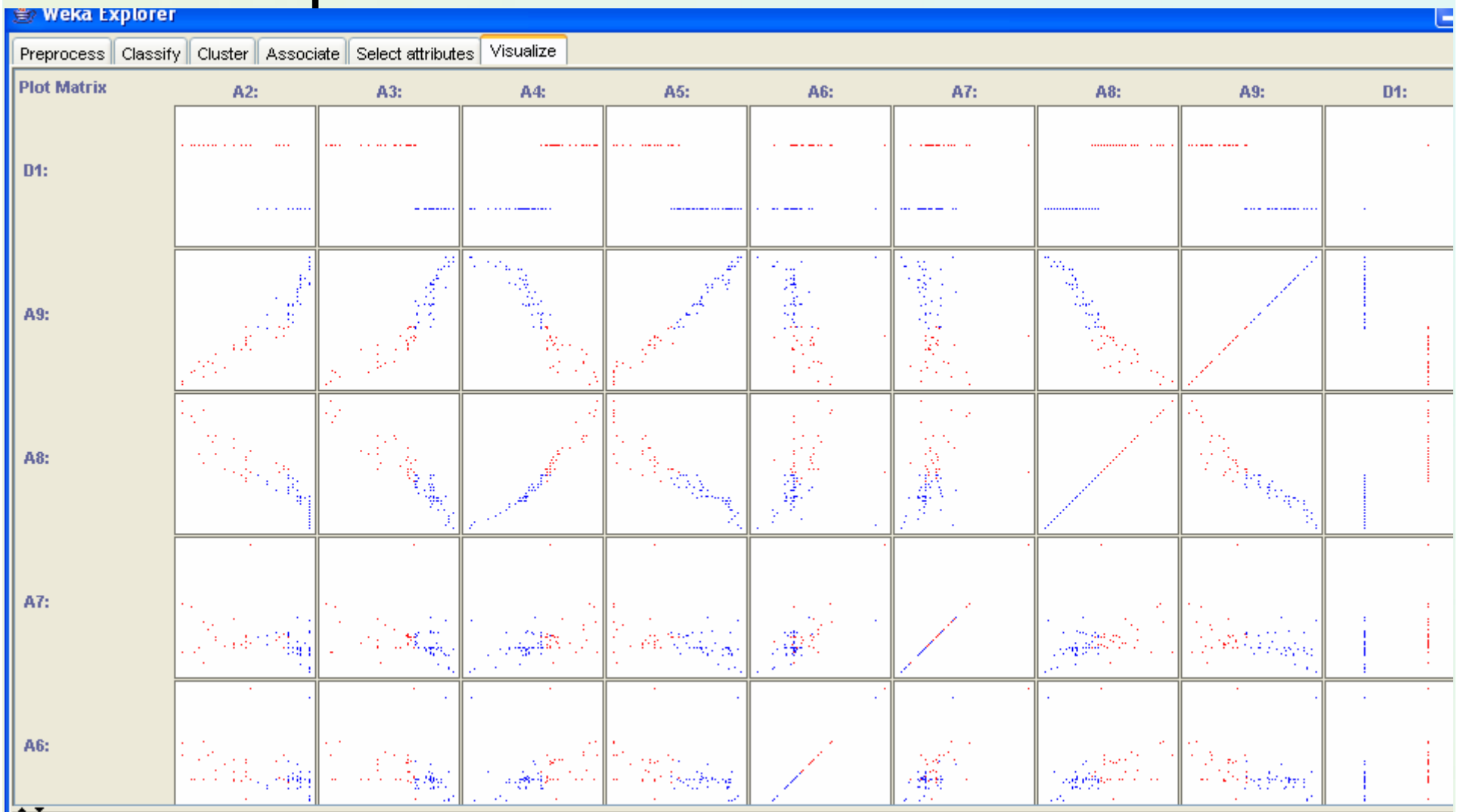
Analiza diagnostycznej bazy danych

- Bada się stan techniczny 80 autobusów tego samego typu (dokładnie ich silników) na podstawie symptomów stanu technicznego - parametrów pochodzących z okresowych badań diagnostycznych
 - Pierwsza klasyfikacja D1: autobusy są podzielone na dwie klasy: dobry i zły stan techniczny pojazdu
 - Możliwa jest druga klasyfikacja D2 + stan przejściowy
- Cel analizy
 - Ocenia się jakość diagnostyczną symptomów stanu technicznego
 - (pośrednio ocena przydatności tzw. wartości granicznych)
 - Ocena ważności poszczególnych symptomów
 - Ewentualność rankingu lub selekcji
 - Poszukuje się zależności pomiędzy wartościami najistotniejszych w tych symptomów a przydziałem do klas
 - Konstruuje się klasyfikator stanu technicznego

Opis danych

- 80 obserwacji (autobusów)
- 8 atrybutów
 - max speed km/h
 - Compression pressure Mpa
 - blacking components in exhaust gas [%]
 - torque Nm
 - summer fuel consumption l/100km
 - winter fuel consumption l/100km
 - oil consumption l/1000km
 - max horsepower kW
- D1: [1,2] | Stan techniczny pojazdu: 1 – dobry , 2 – zły
lub 3 klasy D2: [1,2,3]

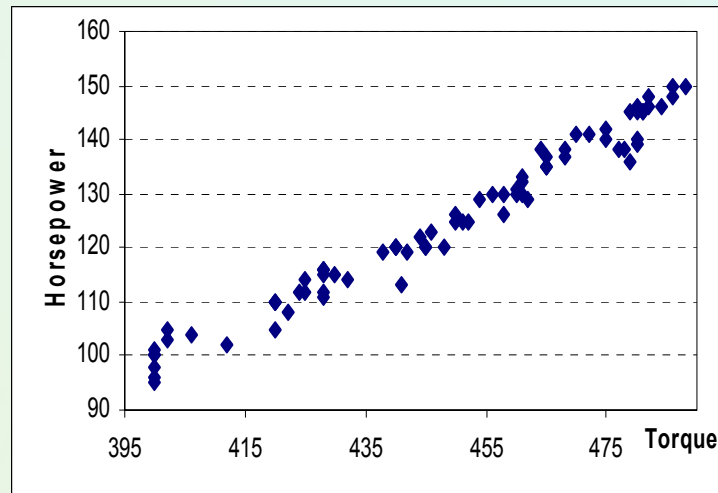
Scatter point matrix



WEKA - Visualization

Powiązanie pewnych symptomów

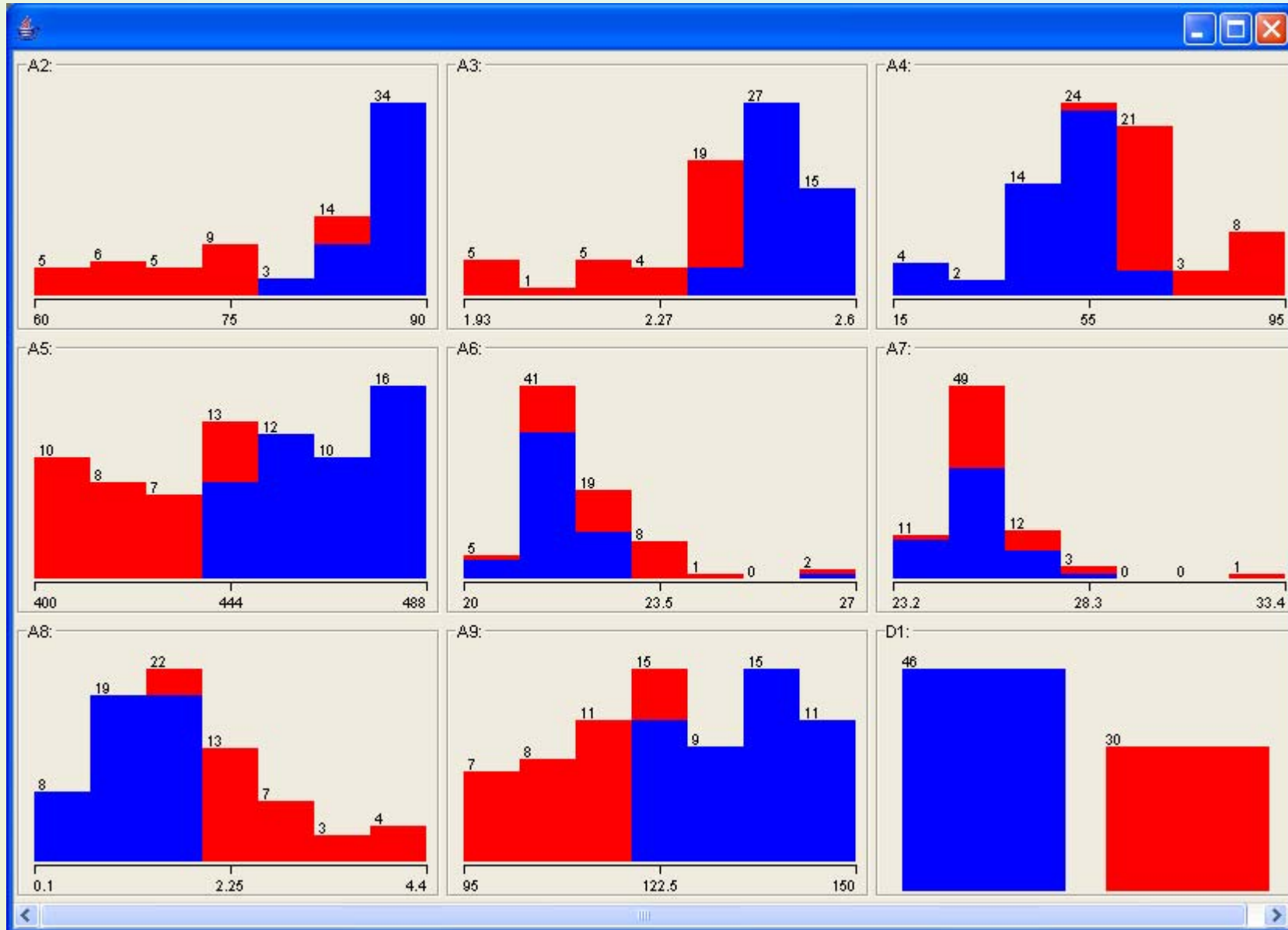
Można zrobić wykresy korelacyjne (rozzutu XY)



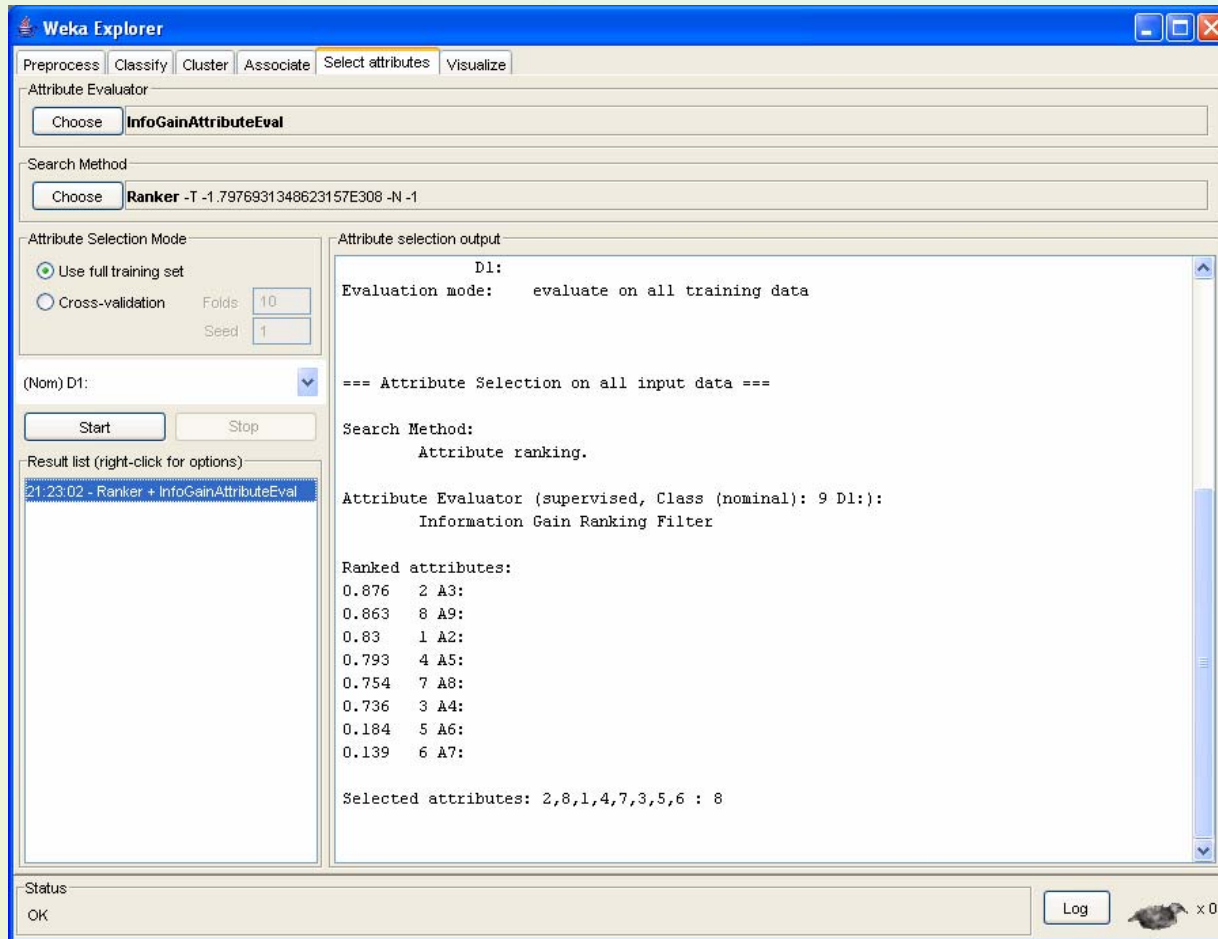
Wiedza dziedzinowa

- ciśnienie sprężania powiązane jest z momentem obrotowym, im większe ciśnienie tym większy można uzyskać moment obrotowy,
- moment obrotowy powiązany jest z mocą pojazdu,
- zawartość elementów smołowych i zużycie oleju świadczyć może o wieku silnika i jego stanie technicznym,
- Mniejsza użyteczność info. o zużyciu paliwa (warunki, styl jazdy,..)

WEKA –visualize



„Select attributes” z WEKA



Różne metody –
lekką zróżnicowane
selekcje

Możliwe jest
ograniczenie do
podzbiorów z 3
atributami

Drzewo J4.8 (WEKA)

The screenshot shows the Weka Explorer interface with the J4.8 classifier selected. The classifier output window displays the following information:

Classifier output

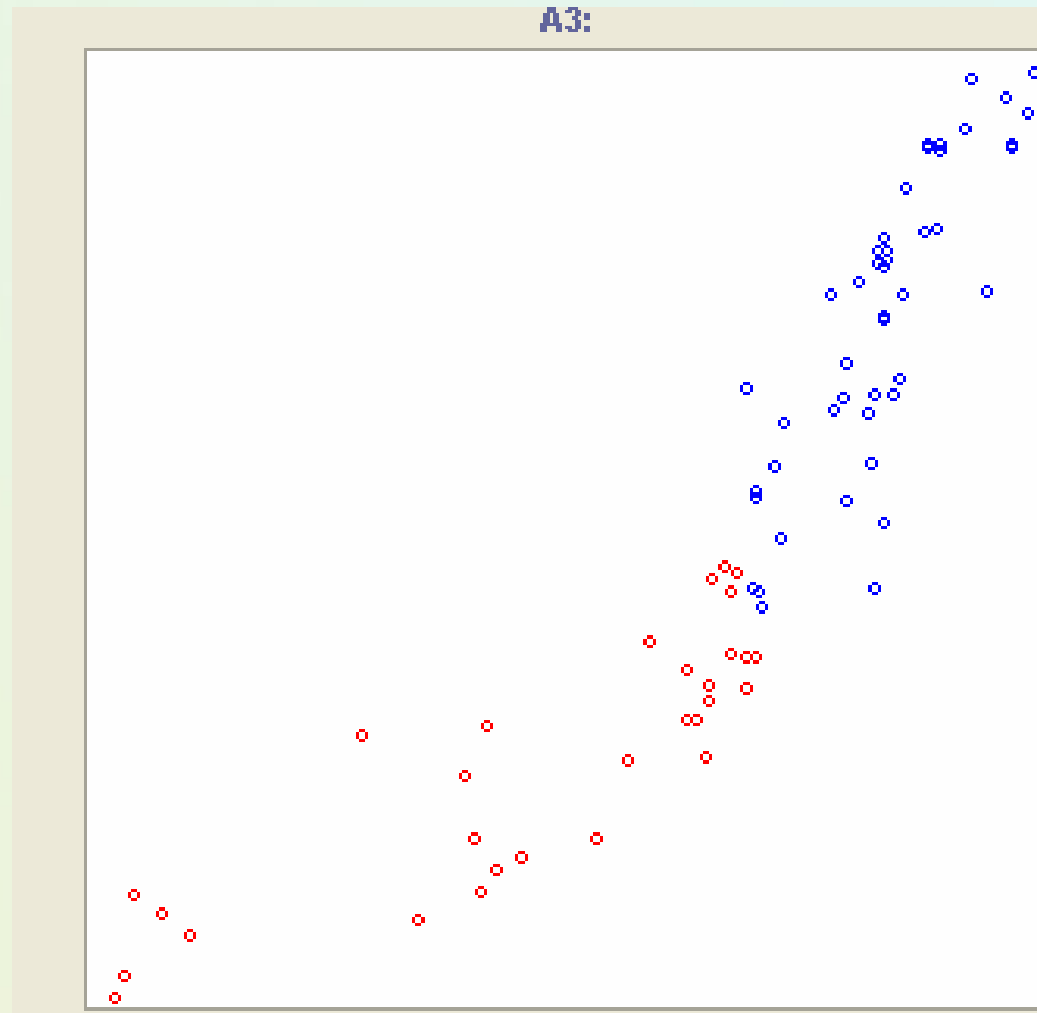
```
-----  
A3: <= 2.39: 2 (29.0)  
A3: > 2.39: 1 (47.0/1.0)  
  
Number of Leaves :    2  
Size of the tree :    3  
  
Time taken to build model: 0.01 seconds  
  
=== Stratified cross-validation ===  
=== Summary ===  
  
Correctly Classified Instances      75           98.6842 %  
Std Dev. of Corr. Class. Inst.     3.75      %  
Incorrectly Classified Instances    1           1.3158 %  
Kappa statistic                     0.9723  
Mean absolute error                  0.0259  
Root mean squared error              0.116  
Relative absolute error              5.4081 %  
Root relative squared error         23.726 %  
Total Number of Instances          76  
  
=== Detailed Accuracy By Class ===  
  
TP Rate  FP Rate  Precision  Recall  F-Measure  Class  
1        0.033    0.979     1       0.989     1  
0.967    0         1         0.967   0.983     2  
  
=== Confusion Matrix ===
```

The interface also shows the following details:

- Classifier:** J48 -U -M 2
- Test options:** Cross-validation (10 folds, 66% split)
- Result list:** 21:27:16 - trees.J48 (selected)
- Status:** OK

WEKA – Scatterplot Visualization

A3 (ciężnienie sprężania) vs. A9 (rzeczywista moc silnika) i 2 klasy

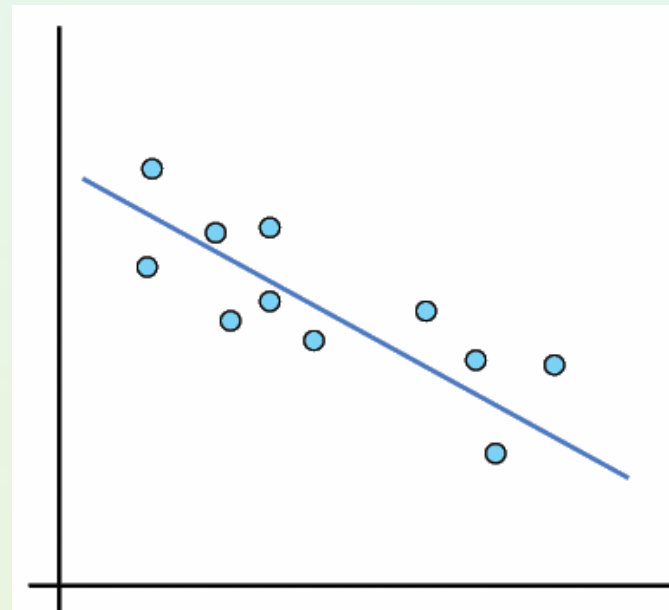


Metody projekcji → tworzenie nowych atrybutów

- **Principal Component Analysis** PCA → **Analiza składowych głównych** / Carl Pearson 1901
- Hotelling (1933) rozwinięte i zastosowane do analizy testów osiągnięć szkolnych
 - Wystarczą 2 składowe zdolności słowne i matematyczne
- Dla n obserwacji z m wymiarowej przestrzeni znajdź c ortogonalnych wektorów ($c \ll m$), które dobrze reprezentują właściwości oryginalnych danych.
- Jak dokonać dobrej konstrukcji / redukcji wymiarów bez znaczącej utraty informacji?
- Używane do kompresji danych i lepszej wizualizacji ogólnych prawidłowości w związkach między danymi
- Przeznaczone do danych liczbowych.
- Oprogramowanie statystyczne → Statsoft i inne
- Więcej opisów i przykładów → książki Larose / Stanisz / Krzyśko, Wołyński

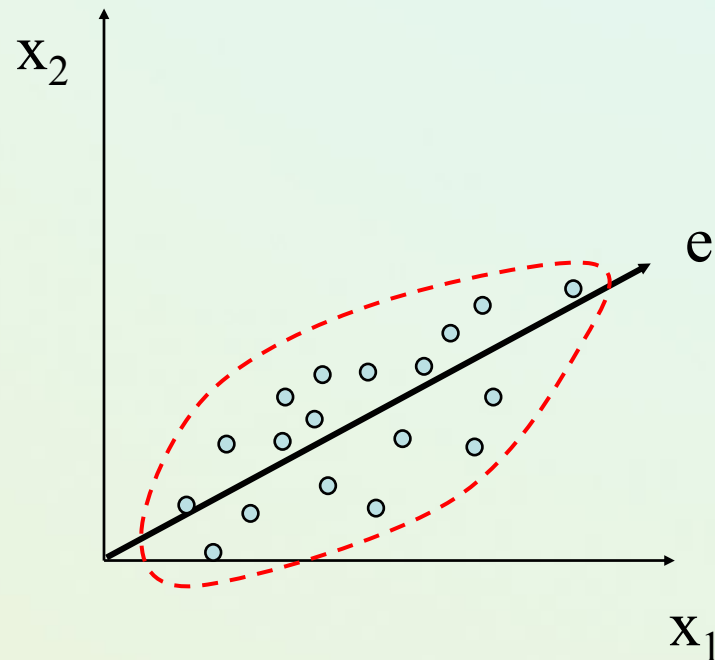
PCA – analiza składowych głównych (1)

- Istota to „ortogonalne” przekształcenie początkowych skorelowanych zmiennych w nowy zbiór **nieskorelowanych zmiennych**
 - Całkowita wariancja zmiennych jest równa sumie wariancji składowych głównych
 - Problem wyboru nowych składowych
- Każdy z nowych wektorów jest **kombinacją liniową** pewnych składowych głównych (odnoszących się do oryginalnych zmiennych)

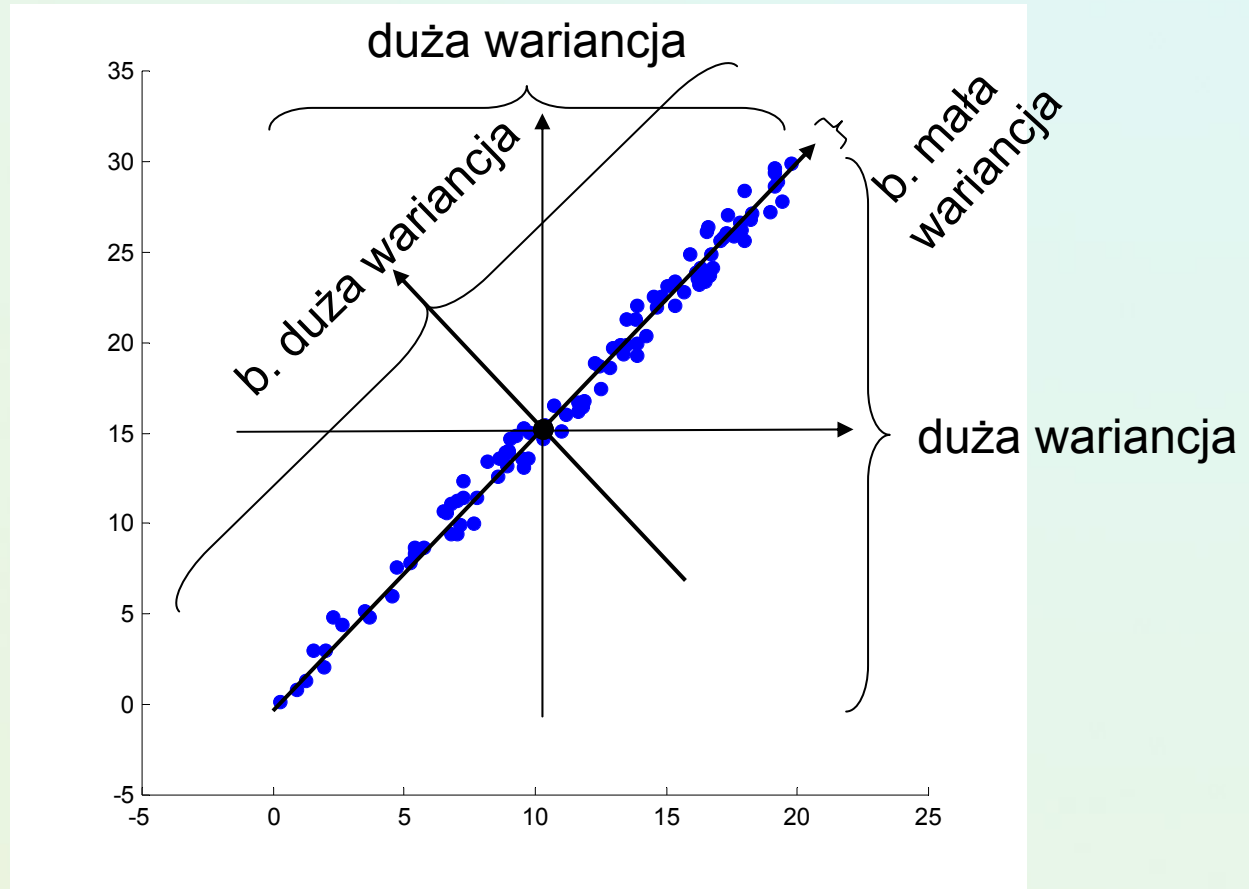


PCA – analiza składowych głównych (2)

- Nowe nieskorelowane zmienne (tzw. składowe główne) powstają z przekształcenia oryginalnych zmiennych skorelowanych, w taki sposób aby w maksymalnym stopniu wyjaśniać całkowitą wariancję w próbie zmiennych oryginalnych.
- Formalnie wariancje składowych głównych są wartościami własnymi macierzy kowariancji oryginalnych zmiennych (the eigenvectors of the covariance matrix).
 - Pierwsza składowa główna redukuje największą część zróżnicowania, druga kolejną której nie redukowała poprzednia itp.



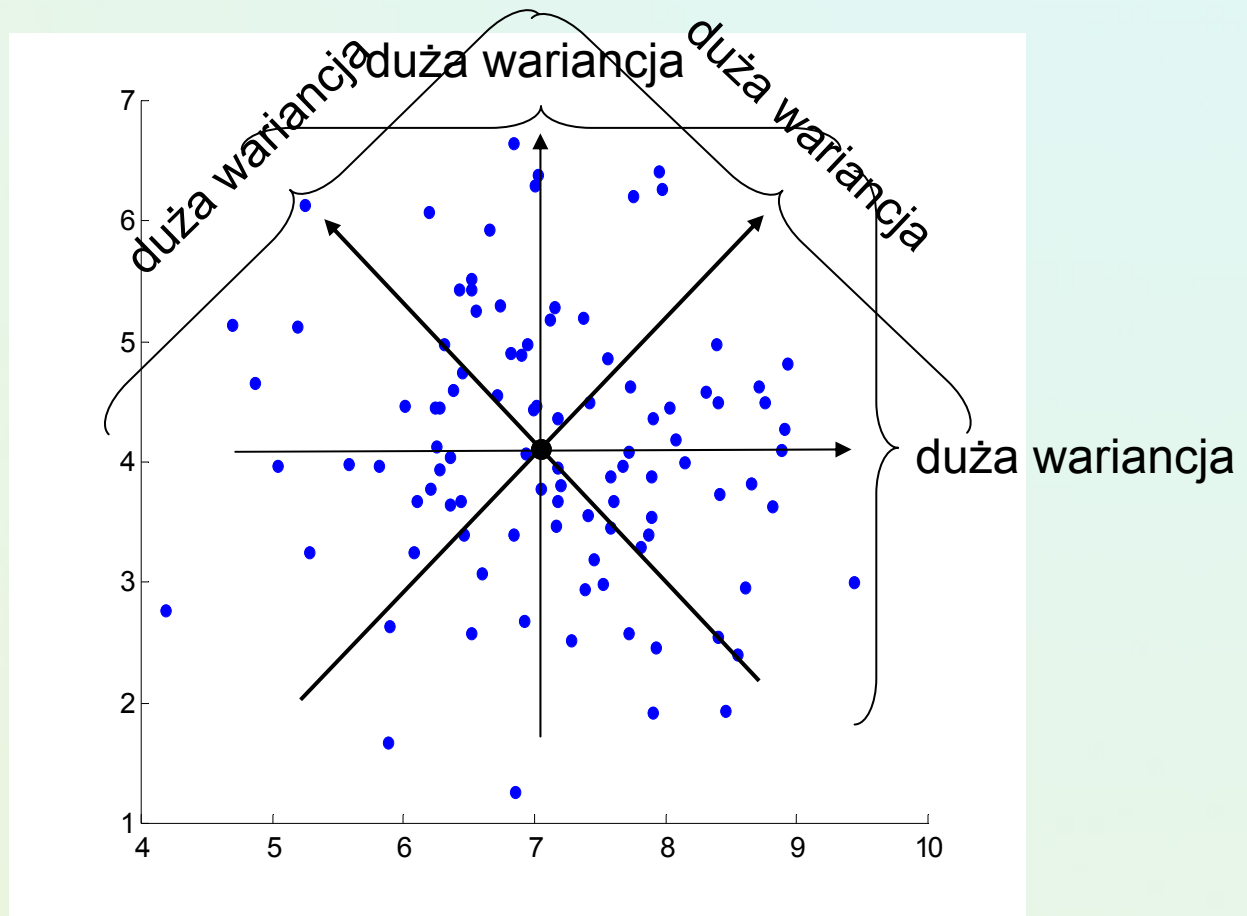
Przykład: PCA dla danych dwuwymiarowych



Pod względem wariancji

- stare zmienne: małe zróżnicowanie
- nowe zmienne: duże zróżnicowanie

Przykład: PCA dla danych dwuwymiarowych



Pod względem wariancji

- stare zmienne: małe zróżnicowanie
- nowe zmienne: małe zróżnicowanie

Procedura PCA – więcej matematyki

- Dana jest macierz danych: \mathbf{X} (obserwacje w wierszach)
 - oblicz macierz kowariancji $\mathbf{S}_x = \mathbf{X}^T \mathbf{X}$
 - można pominąć mnożnik $(1/(m-1))$ czy $(1/m)$
 - macierz \mathbf{S}_x jest symetryczna i pozwala ocenić
 - wariacje zmiennych (elementy na głównej przekątnej)
 - zależności pomiędzy zmiennymi (elementy poza główną przekątną)
 - dokonaj rozkładu: $\mathbf{S}_x = \mathbf{K} \mathbf{L} \mathbf{K}^T$
 - utwórz nowe zmienne wykonując operację $\mathbf{Y} = \mathbf{X} \mathbf{K}$

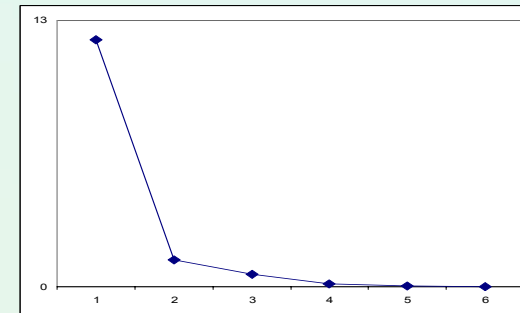
Procedura PCA

- Wniosek 1:
macierz kowariancji \mathbf{S}_Y zmiennych przekształconych \mathbf{Y} wyraża się macierzą diagonalną \mathbf{L} (utworzoną z wartości własnych macierzy \mathbf{S}_X)
 - kowariancje zmiennych przekształconych (elementy pozadiagonalne macierzy \mathbf{S}_Y) są równe zero
 - nowe zmienne są niezależne liniowo
 - wariancje zmiennych przekształconych (elementy diagonalne macierzy \mathbf{S}_Y) są równe wartościom własnym macierzy \mathbf{S}_X (zapisanym na przekątnej macierzy \mathbf{L})
 - można je poznać tuż po wyliczeniu wartości własnych macierzy \mathbf{S}_X , a więc jeszcze przed wyliczeniem nowej macierzy danych $\mathbf{Y} = \mathbf{XK}$

PCA – nowy układ współrzędnych

- Osie uporządkowane wzduż zmienności / wariacji
- Weź k pierwszych składowych PCA

$$V_k = \frac{\sum_{i=1}^k \lambda_k}{\sum_{j=1}^d \lambda_j}$$



- Starając się zachować przynajmniej ~ wystarczająco dużo % zmienności całkowitej

Na podstawie
tzw. wykresu osypiska

PCA – może pomóc zrozumieć strukturę danych

- za książka Elements of Statistical Learning, Friedman, Hastie, Tibushirani ...

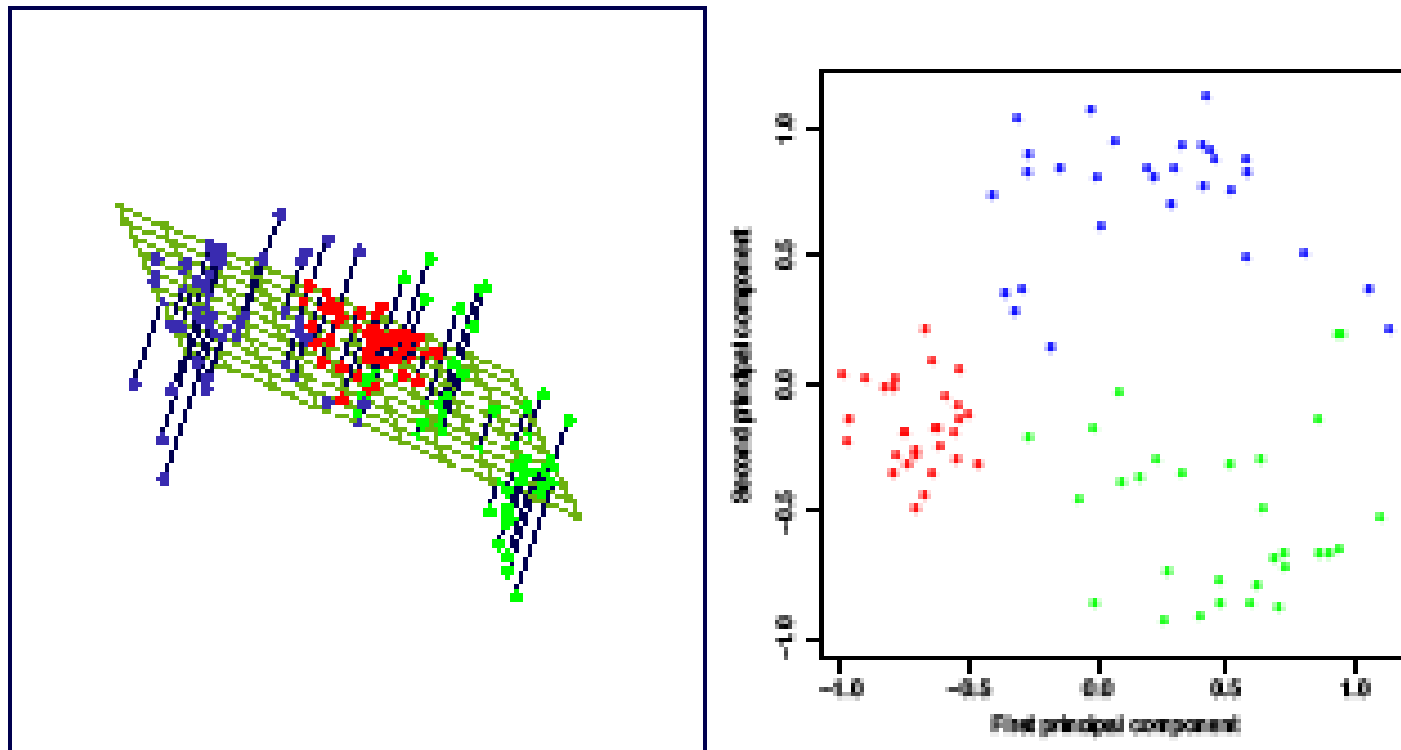


FIGURE 14.21. The best rank-two linear approximation to the half-sphere data. The right panel shows the projected points with coordinates given by $U_2 D_2$, the first two principal components of the data.

Zastosowania PCA: wizualizacja danych wielowymiarowych

- Przykłady
 - 15 samochodów osobowych
- Atrybuty
 - 4 wybrane parametry
 - zużycie paliwa [1/100 l/km], zakres: [6,10], typ: strata
 - przyspieszenie do 100km/h [s], zakres: [9,14], typ: strata
 - cena [tys. PLN], zakres: [24, 94], typ: strata
 - wyposażenie [punkty uznaniowe], zakres: [1, 10], typ: zysk

PCA: wizualizacja danych wielowymiarowych (2)

Oryginalne opisy 15 samochodów (tabela danych)

Marka	Model	Zużycie	Przyspieszenie	Cena	Wyposażenie
Alfa Romeo	156	8,1	9,3	71,14	9
Audi	A4	7,9	11,9	93,35	10
BMW	316I	7,5	12,3	81,79	8
Daewoo	Lanos	8,4	12,2	34,90	3
Honda	Civic	6,7	10,8	48,90	7
Hyunday	Accent	6,4	11,7	35,30	2
Lada	Samara	7,2	13,0	24,90	2
Mitsubishi	Carisma	7,2	12,0	60,60	9
Opel	Astra II	7,0	12,0	56,95	8
Peugeot	206 XR	6,6	13,2	38,36	4
Renault	Megane	7,0	9,9	50,05	9
Saab	9 3S	9,7	11,0	85,35	8
Seat	Cordoba	8,3	10,9	44,99	8
Toyota	Corrola	7,7	10,2	50,36	4
Volkswagen	Golf IV	8,3	9,9	61,62	6

PCA: wizualizacja danych wielowymiarowych (2)

Znormalizowane opisy 15 samochodów

Marka	Model	Zużycie	Przyspieszenie	Cena	Wyposażenie
Alfa Romeo	156	0,57	-1,74	0,76	0,93
Audi	A4	0,34	0,46	1,87	1,31
BMW	316I	-0,11	0,80	1,29	0,56
Daewoo	Lanos	0,92	0,72	-1,05	-1,34
Honda	Civic	-1,03	-0,47	-0,35	0,18
Hyunday	Accent	-1,37	0,29	-1,03	-1,34
Lada	Samara	-0,46	1,40	-1,55	-1,72
Mitsubishi	Carisma	-0,46	0,55	0,23	0,93
Opel	Astra II	-0,69	0,55	0,05	0,56
Peugeot	206 XR	-1,14	1,57	-0,88	-0,96
Renault	Megane	-0,69	-1,23	-0,29	0,93
Saab	9 3S	2,40	-0,30	1,47	0,56
Seat	Cordoba	0,80	-0,39	-0,55	0,56
Toyota	Corrola	0,11	-0,98	-0,28	-0,96
Volkswagen	Golf IV	0,80	-1,23	0,29	-0,20

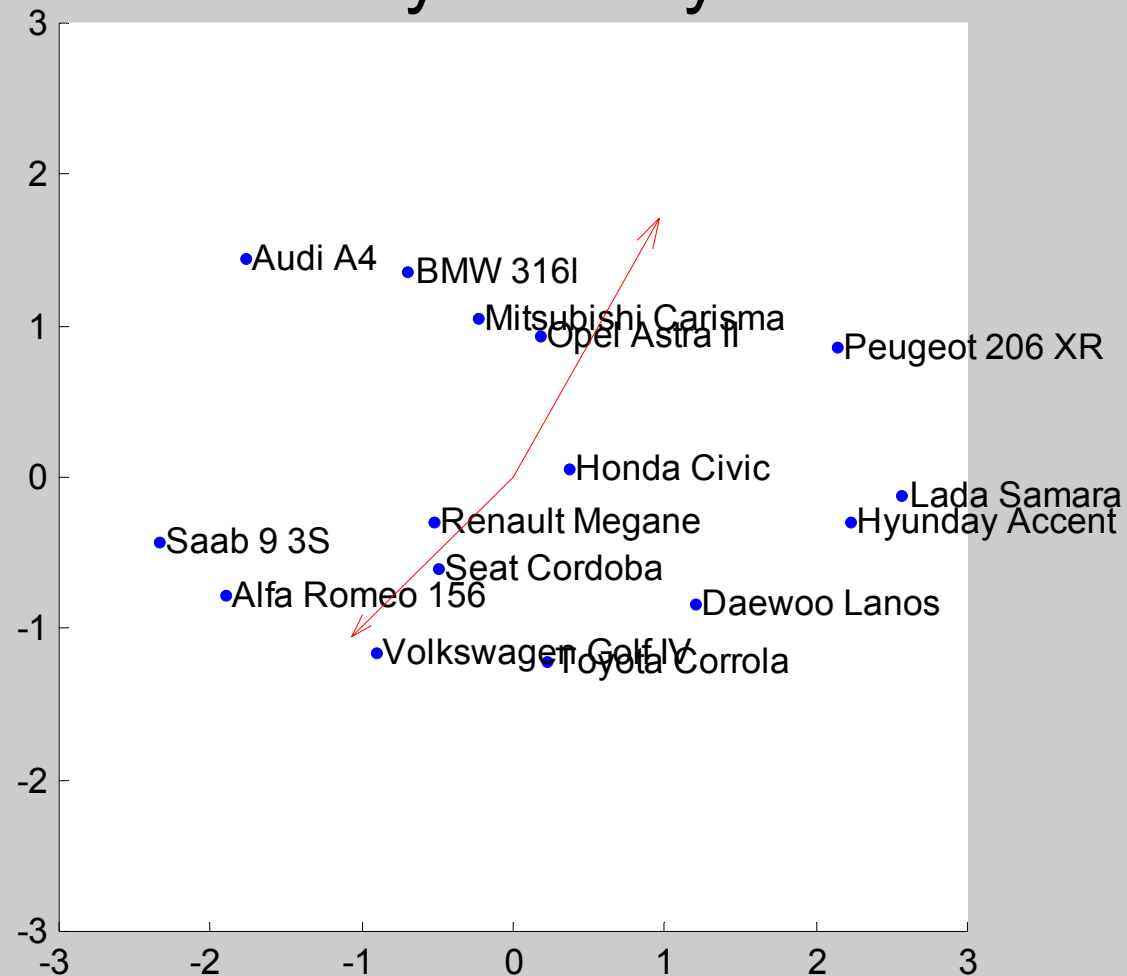
Trochę macierzy

1.0000	-0.3245	0.4872	0.2619
-0.3245	1.0000	-0.2609	-0.3814
0.4872	-0.2609	1.0000	0.7691
0.2619	-0.3814	0.7691	1.0000

Nowe zmienne

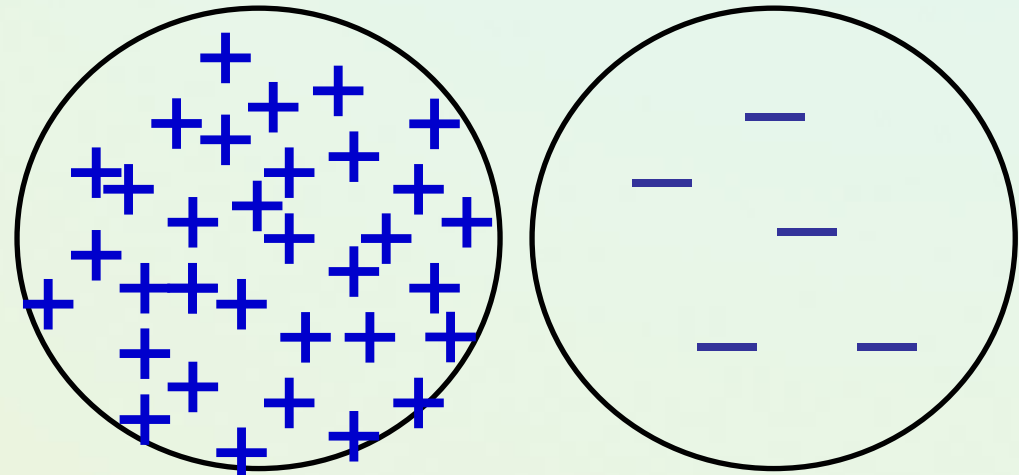
-1.9007	-0.7758
-1.7763	1.4540
-0.6978	1.3548
1.2019	-0.8323
0.3599	0.0557
2.2258	-0.2916
2.5666	-0.1197
-0.2338	1.0564
0.1759	0.9305
2.1353	0.8628
-0.5332	-0.2984
-2.3404	-0.4234
-0.4961	-0.5953
0.2245	-1.2162
-0.9117	-1.1617

Zastosowania PCA: wizualizacja danych wielowymiarowych



Nieźrównoważenie rozkładu w klasach (Class Imbalance)

- Dane są nieźrównoważone, jeśli klasy nie są w przybliżeniu równoliczne
- ▢ • Klasa mniejszościowa (**minority class**) zawiera wyraźnie mniej przykładów niż inne klasy
- Przykłady z klasy mniejszościowej są często najważniejsze, a ich poprawne rozpoznawanie jest głównym celem.
 - Rozpoznawanie rzadkiej, niebezpiecznej choroby,...
- **CLASS IMBALANCE** → powoduje trudności w fazie uczenia i obniża zdolność predykcyjną



„Class imbalance is not the same as COST sensitive learning. In general cost are unknown!”

Przykład niezrównoważonych danych

```
% 10. Class Distribution:
%      Database:      0   1   2   3   4 Total
%      Cleveland: 164  55  36  35  13  303
%      Hungarian: 188  37  26  28  15  294
%      Switzerland:  8  48  32  30  5   123
%      Long Beach UA: 51  56  41  42  10  200

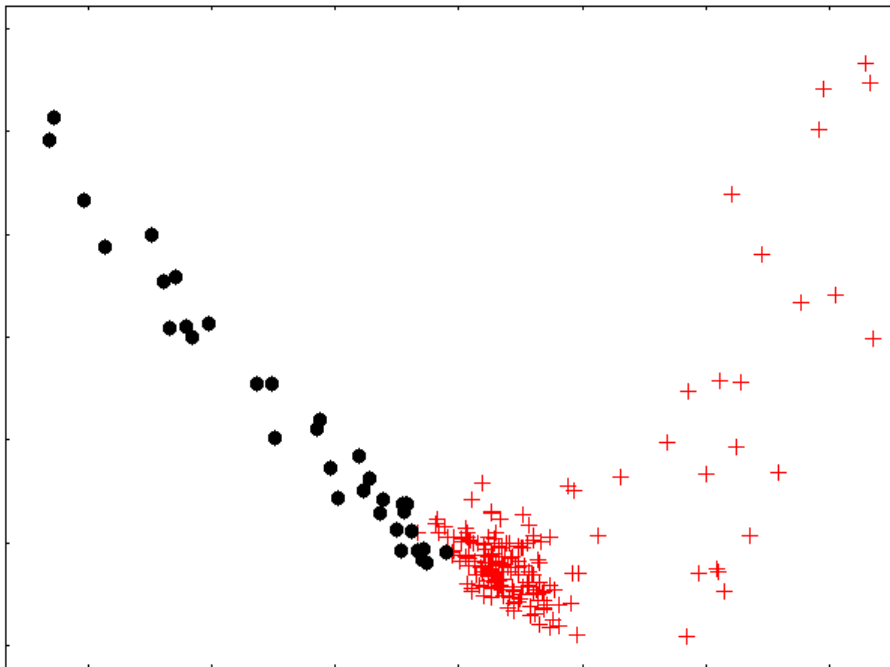
@relation cleveland
@attribute age numeric
@attribute sex {1, 0}
@attribute cp {1, 2, 3, 4}
@attribute trestbps numeric
@attribute chol numeric
@attribute fbs {1, 0}
@attribute restecg {0, 1, 2}
@attribute thalach numeric
@attribute exang {1, 0}
@attribute oldpeak numeric
@attribute slope {1, 2, 3}
@attribute ca numeric
@attribute thal {3, 6, 7}
@attribute num {3, other}

@data
63      1      1      145      233      1      2      150      0      2.3      3      0      6      other
67      1      4      160      286      0      2      108      1      1.5      2      3      3      other
67      1      4      120      229      0      2      129      1      2.6      2      2      7      other
37      1      3      130      250      0      0      187      0      3.5      3      0      3      other
41      0      2      130      204      0      2      172      0      1.4      1      0      3      other
56      1      2      120      236      0      0      178      0      0.8      1      0      3      other
62      0      4      140      268      0      2      160      0      3.6      3      2      3      3
57      0      4      120      354      0      0      163      1      0.6      1      0      3      other
63      1      4      130      254      0      2      147      0      1.4      2      1      7      other
53      1      4      140      203      1      2      155      1      3.1      3      0      7      other
```

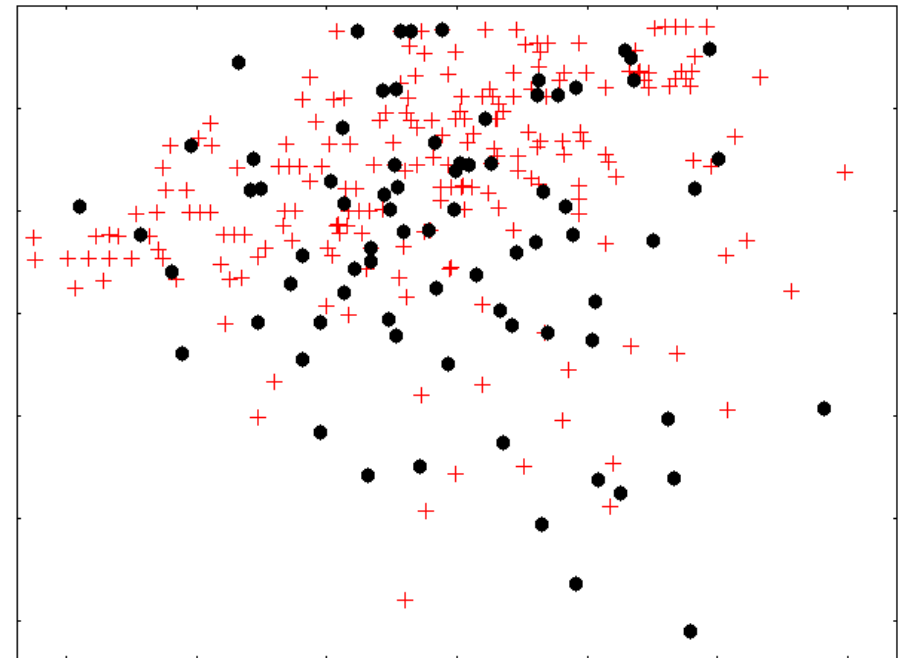
UCI repository – Cleveland medical data

Podglądajmy rzeczywiste dane

Wizualizacja 2 pierwszych składowych w metodzie PCA



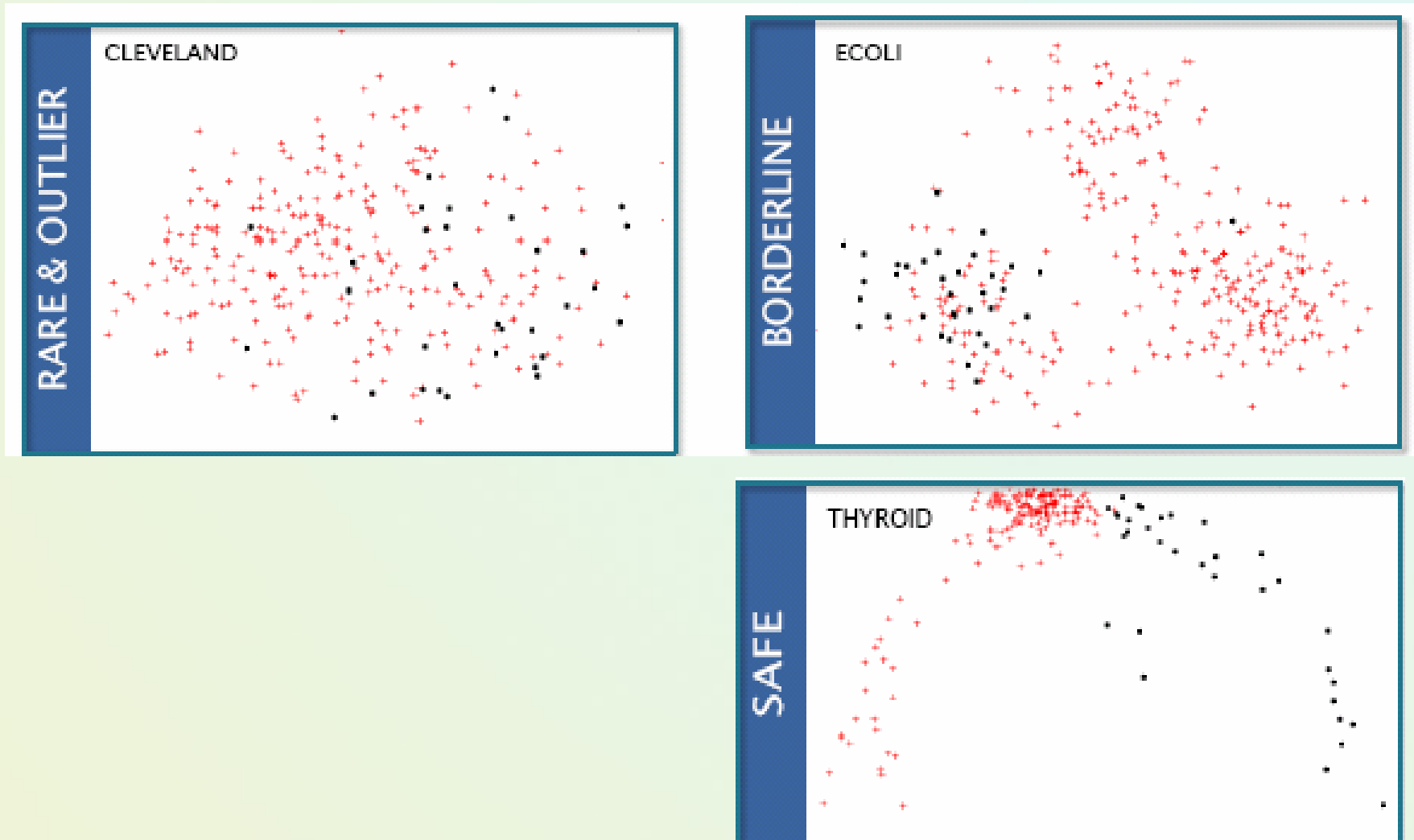
Thyroid
215 ob. / 35



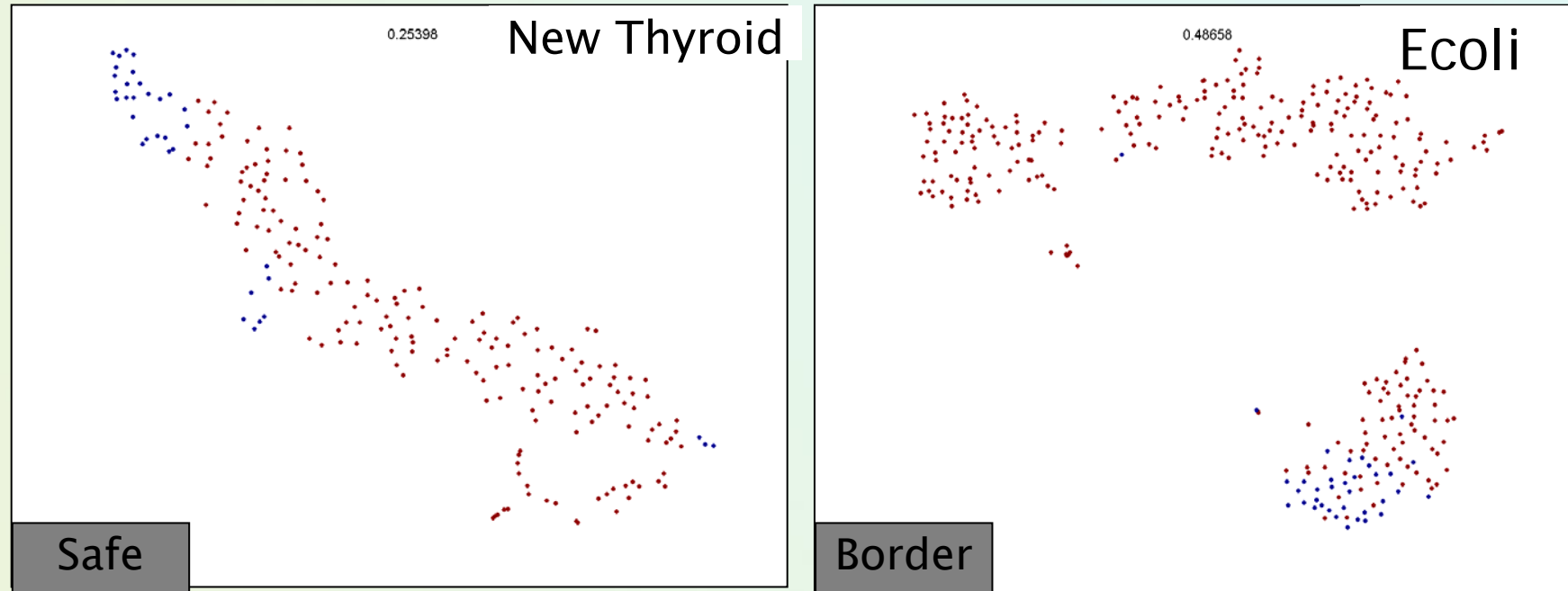
Haberman
306 ob. / 81

MDS wizualizacja danych niezrównoważonych

Skalowanie wielowymiarowe

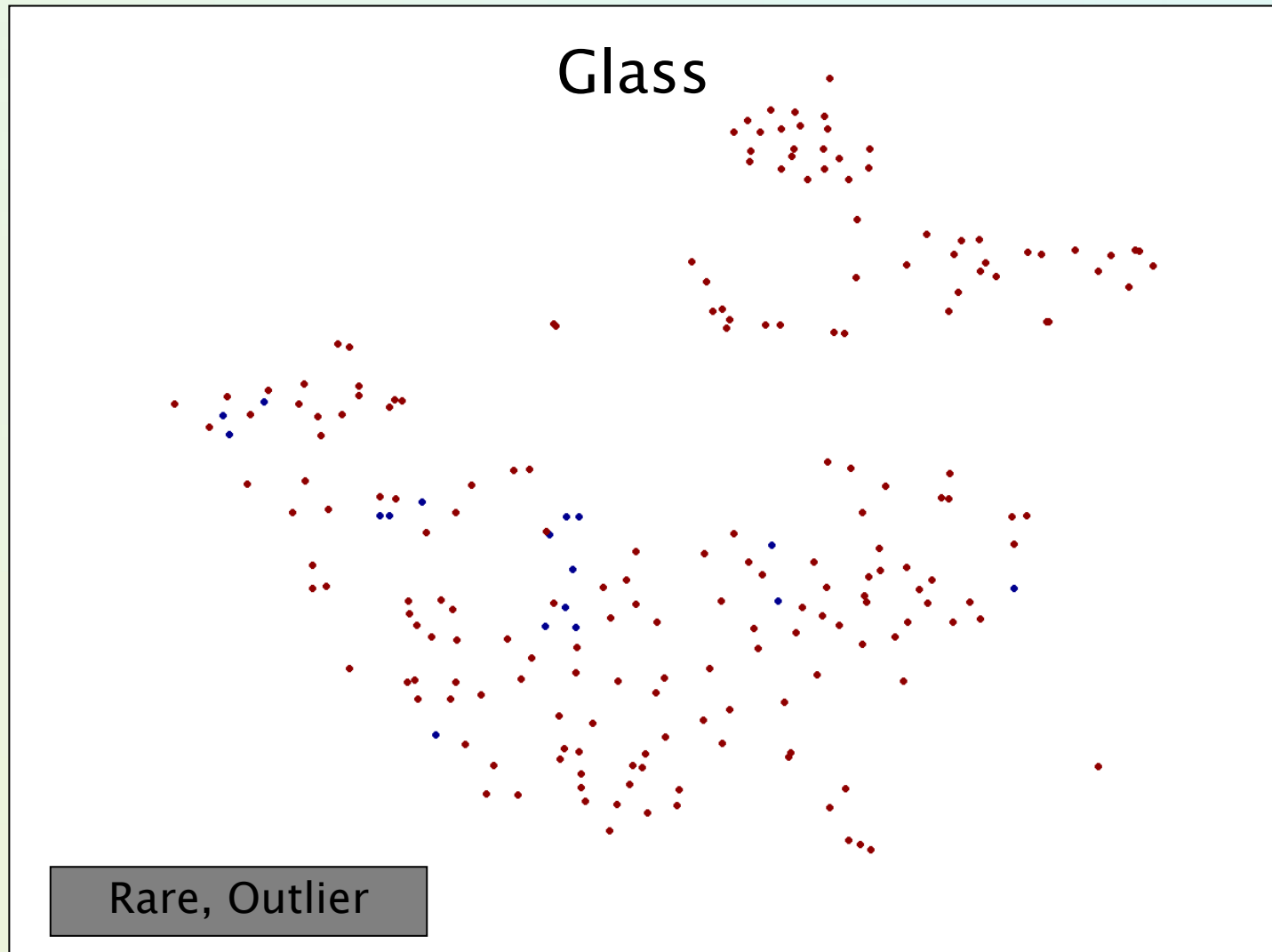


t-SN visualisation



- t-Distributed Stochastic Neighbor Embedding (t-SNE)
- Keeps probability that i is a neighbor of j / local distances

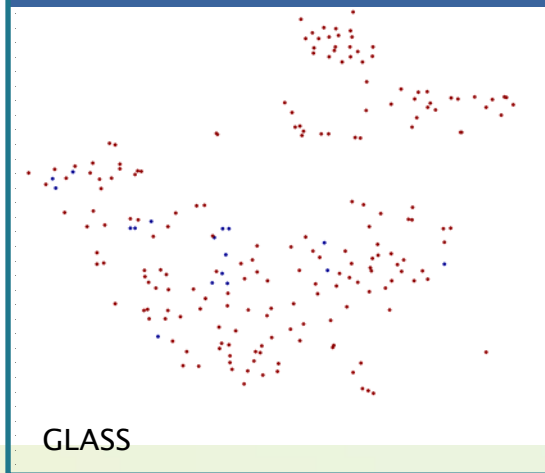
t-SN visualisation



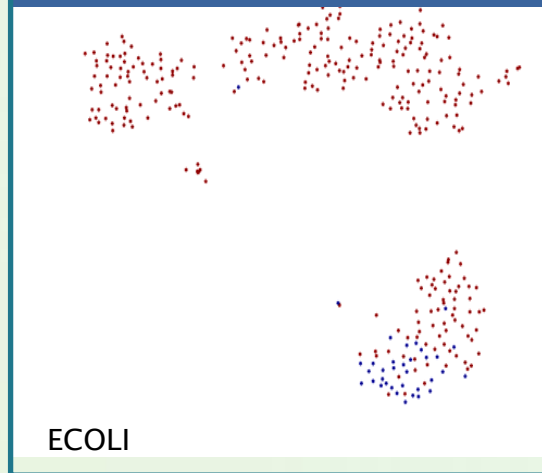
„natura niezrównoważonych danych”

Dataset	Safe	Border	Rare	Outlier	Category
new-thyroid	68,57	31,43	0,00	0,00	S
ecoli	28,57	54,29	2,86	14,29	B
glass	0,00	35,29	35,29	29,41	R, O

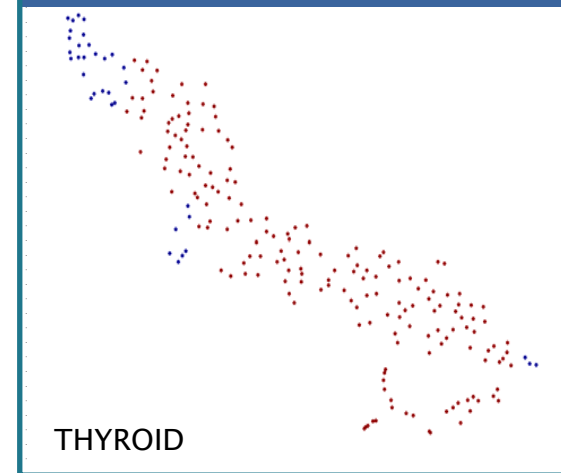
RARE & OUTLIER



BORDERLINE

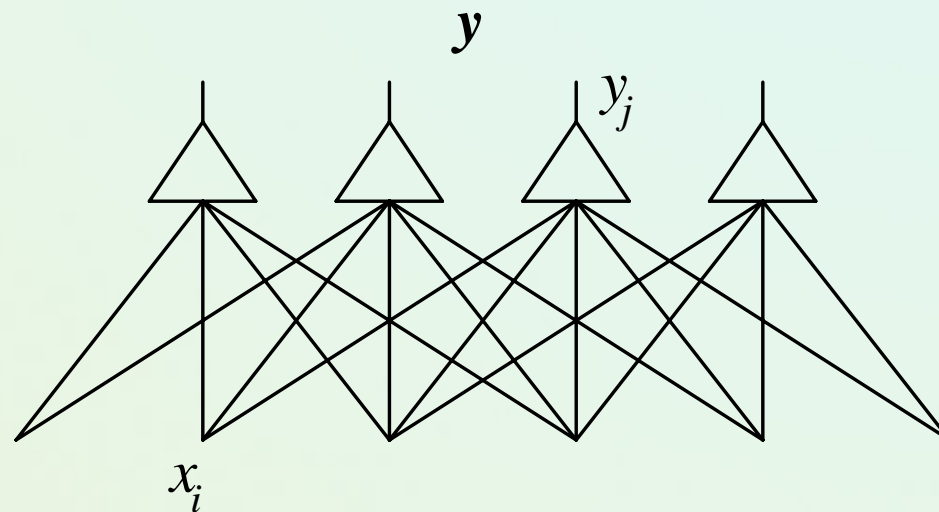


SAFE



Podstawowa sieć Kohonena LVQ

- Celem jest grupowanie wektorów wejściowych \mathbf{x}
- Istota działania \rightarrow podobieństwo
- Podobne wektory powinny pobudzać te same neurony
- Prosta topologia



- gdzie $y_j = \mathbf{w}_j \mathbf{x} = \sum_i w_{ij} x_i$
- Reguła konkurencyjnego uczenia się

Reguła „zwycięzca bierze wszystko”

- Określenie zwycięzcy:

$$\|\mathbf{x} - \hat{\mathbf{w}}_m\| = \min_{i=1, \dots, p} \|\mathbf{x} - \hat{\mathbf{w}}_i\|$$

- Alternatywnie iloczyn skalarny

$$\hat{\mathbf{w}}_m^T \mathbf{x} = \max_{i=1, \dots, p} \hat{\mathbf{w}}_i^T \mathbf{x}$$

- Zwycięzcą jest jeden neuron m . Korekcja wag \mathbf{w}_m odbywa się wyłącznie dla neuronu zwycięzcy według reguły:

$$\Delta \mathbf{w}_m = \eta \cdot (\mathbf{x} - \mathbf{w}_m)$$

- η - stała uczenia (na ogół między 0.1 a 0.7)
- Przykład – interpretacja geometryczna

Sieci SOM

- Podstawą odwzorowania takie uporządkowanie neuronów, takie że położenie zwycięskich neuronów niesie informacje
- Topologia → relacja sąsiedztwa
- Podobne przykłady wejściowe x powinny aktywizować sąsiednie neurony
- „Gęstość” wzorców w zredukowanej przestrzeni musi odpowiadać gęstości przykładów w oryginalnej przestrzeni

Uczenie sieci SOM

Siatka neuronów $i = 1 \dots K$ w 1D-3D, każdy neuron z N wagami.

Neurony z wagami $\mathbf{W}_i(t) = \{W_{i1} \ W_{i2} \dots W_{iN}\}$, wektory $\mathbf{X} = \{X_1, X_2 \dots X_N\}$.

t - dyskretny czas; nie ma połączeń pomiędzy neuronami!

1. Inicjalizacja: przypadkowe $\mathbf{W}_i(0)$ dla wszystkich $i=1..K$.
Funkcja sąsiedztwa $h(|r-r_c|/\sigma(t), t)$ definiuje wokół neuronu położonego w miejscu r_c siatki obszar $O_s(r_c)$.
2. Oblicz odległości $d(\mathbf{X}, \mathbf{W})$, znajdź neuron z wagami W_c najbardziej podobnymi do X (neuron-zwycięzcę).
3. Zmień wagi wszystkich neuronów w sąsiedztwie $O_s(r_c)$
4. Powoli zmniejszaj siłę $h_o(t)$ i promień $\sigma(t)$.
5. Iteruj aż ustaną zmiany.

Efekt: podział (tesselacja) na wieloboki Voronoia.

Przykład – analiza produkcji oliwy we Włoszech

SOM application oil Italian market:

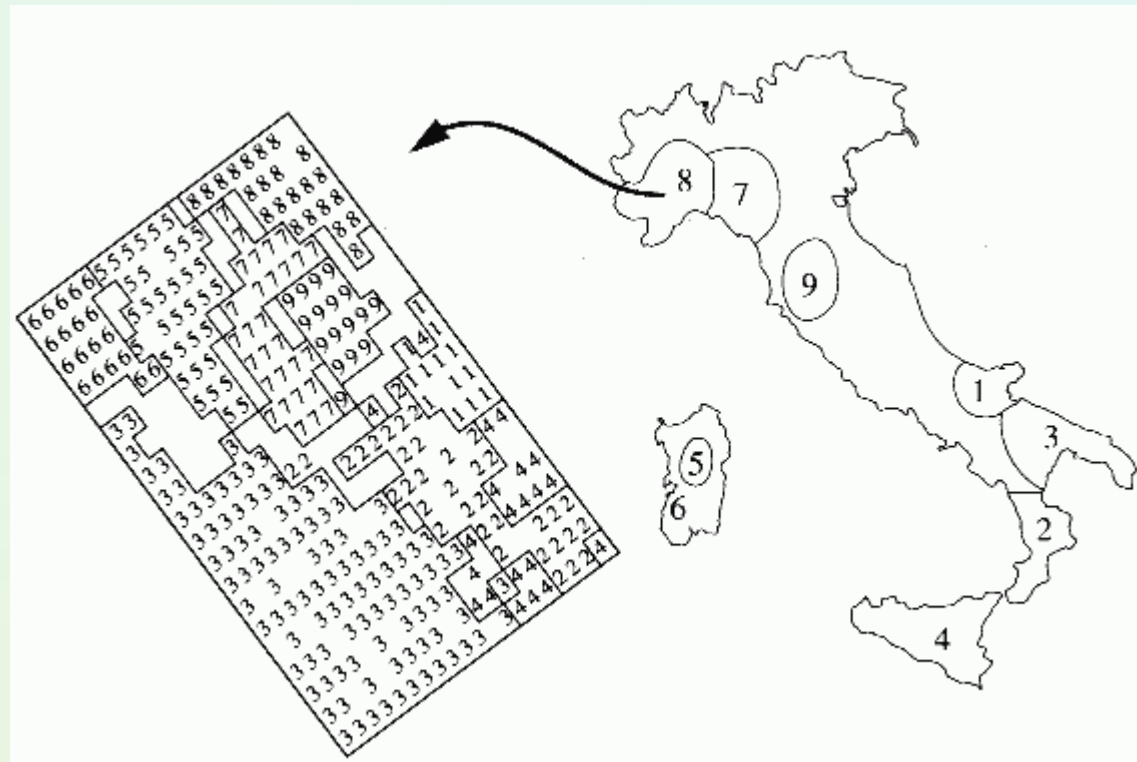
572 samples of olive oil were collected from 9 Italian provinces

Content of 8 fats was determined for each oil

SOM 20 x 20 network,

Maps 8D => 2D

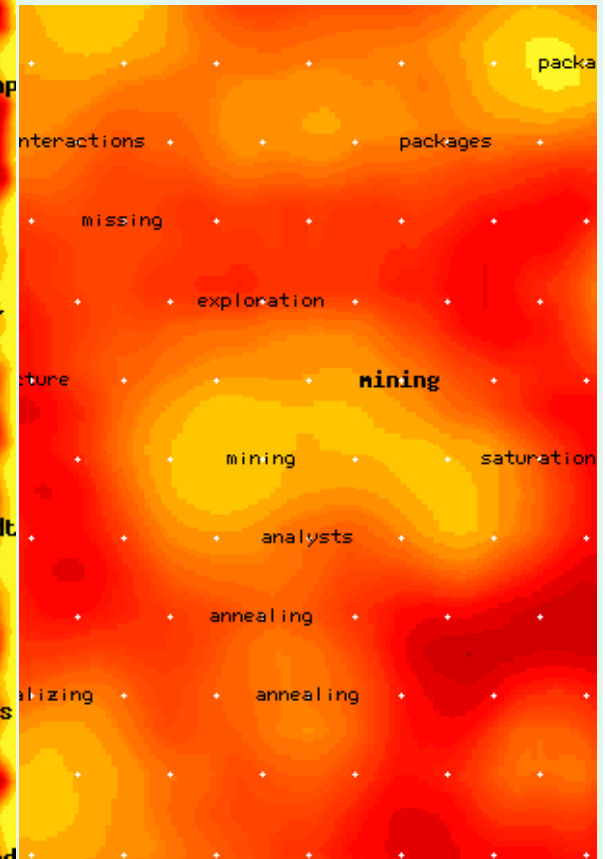
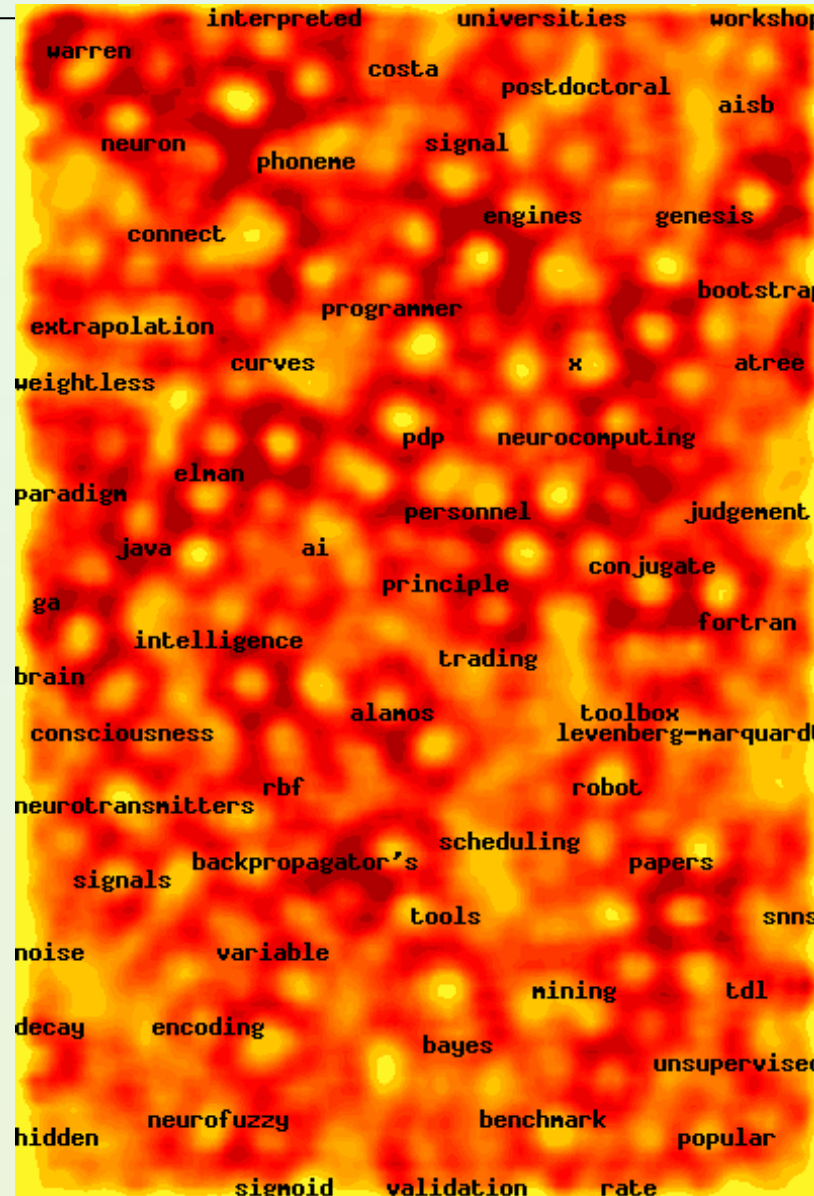
Classification accuracy was around 95-97%



Note that topographical relations are preserved, region 3 is most diverse

Web Document Clustering Using SOM

- The result of SOM clustering of 12088 Web articles
- The picture on the right: drilling down on the keyword “mining”
- Based on websom.hut.fi Web page



Trochę wskazówek literaturowych

- Robert Susmaga – wykład Metody kompresji i wizualizacji danych (PP obieralny przedmiot semX)
- T.Hastie, R.Tibshirani, J.Friedman: The Elements of Statistical Learning. Springer → poszukaj wersji elektronicznej pdf
- J.Koronacki, J.Ćwik: Statystyczne systemy uczące się WNT.
- M.Krzyśko, W.Wołyński, T.Górecki, M.Skorzybut: Systemy uczące się. WNT.
- S.Osowski: Sieci neuronowe w przetwarzaniu informacji.
- K.Napierała, J.Stefanowski: Identification of Different Types of Minority Class Examples in Imbalanced Data. Proc. HAIS 2012, Part II, LNAI vol. 7209, Springer Verlag 2012, 139–150.
- Napierała, K., Stefanowski, J.: Types of Minority Class Examples and Their Influence on Learning Classifiers from Imbalanced Data, to appear in Journal of Intelligent Information Systems (2015)

Don't ..., it is not over!

Jeszcze wiele rozwiązań dla danych
jakościowych

