

Klasyfikacja za pomocą kombinacji estymatorów  
jądrowych: dobór szerokości jądra w zależności  
od lokalizacji w przestrzeni cech, stosowanie  
różnych przekształceń przestrzeni, funkcja błędu  
oparta na "information loss"

Mateusz Kobos

03.12.2013

Seminarium z Metod Inteligencji Obliczeniowej, MiNI PW

# Spis treści

Wyniki poprzednie

Nowa wersja algorytmu z funkcją Information Loss

Wartość optymalizowanego błędu w trakcie działania algorytmu

Stosowanie różnych przekształceń przestrzeni

Porównanie algorytmów z  $N=1,2,3,4,5,7,10$

Dobieranie szerokości jądra do miejsca w przestrzeni

Plany

# Spis treści

Wyniki poprzednie

Nowa wersja algorytmu z funkcją Information Loss

Wartość optymalizowanego błędu w trakcie działania algorytmu

Stosowanie różnych przekształceń przestrzeni

Porównanie algorytmów z  $N=1,2,3,4,5,7,10$

Dobieranie szerokości jądra do miejsca w przestrzeni

Plany



# Działanie algorytmu w skrócie

- ▶ Idea: wykorzystać spojrzenie na dane przy różnych „rozdzielczościach” / „skalach” a następnie wykorzystać uzyskaną w ten sposób wiedzę o strukturze danych do klasyfikacji.
- ▶ Realizacja: estymujemy gęstość danych przy wykorzystaniu estymatorów jądrowych o różnych szerokościach jądra a następnie obliczamy gęstość uśrednioną. Za pomocą wzoru Bayesa obliczamy prawdopodobieństwa przynależności do każdej z klas.
  - ▶ Nauka algorytmu: minimalizacja gradientowa (algorytm L-BFGS-B) błędu MSE

## Co zrobiono: ogólnie

- ▶ Przedstawiliśmy nowe podejście do klasyfikacji opartej na „wielorozdzielczościowym” spojrzeniu na dane.
- ▶ Pokazaliśmy formalnie, że przy pewnych założeniach, ogólna klasa tego typu metod może charakteryzować się niskimi błędami klasyfikacji.
- ▶ Algorytm dobrze działa na sztucznych zbiorach danych o specyficznej własności „wielorozdzielczości” – dane składają się ze skupień o różnej gęstości.
  - ▶ Im większa różnica między gęstościami, tym lepsze wyniki.
- ▶ W eksperymentach na zbiorach referencyjnych (bez doboru pod względem specyficznych własności) przewaga podejścia wielorozdzielczościowego nad wersją podstawową jest mała. Mimo to, algorytm dobrze radzi sobie w porównaniu z innymi klasyfikatorami.

# Własności teoretyczne

- ▶ By uzasadnić w sposób formalny, że stosowanie naszego algorytmu może dawać dobre wyniki, zmodyfikowaliśmy model algorytmu uśredniającego pr. a posteriori  $n$  klasyfikatorów
  - ▶ Oryginalny model zaproponowany w
    - ▶  K. Tumer, J. Ghosh: Analysis of decision boundaries in linearly combined neural classifiers, **Pattern Recognition** 29 (1996) 341–348.
    - ▶  K. Tumer, J. Ghosh: Error correlation and error reduction in ensemble classifiers, **Connection Science** 8 (1996) 385–404.
  - ▶ Pokazaliśmy, że przy pewnych założeniach stosowanie  $N$  estymatorów prowadzi do  $N$ -krotnego zredukowania tzw. błędu dodanego klasyfikacji.
- ▶ Proponowanej podejście jest równoważne takiemu, gdzie używamy pojedynczego KDE o szczególnej postaci funkcji jądra. Obliczyliśmy tzw. „efektywność” jądra (osiąga ona maksimum, gdy szerokości jąder różnych estymatorów są sobie równe).


# Eksperymenty na sztucznych danych

- ▶ Intuicja: by jak najlepiej wykorzystać podejście wielorozdzielczościowe, badany problem klasyfikacyjny musi mieć naturę wielorozdzielczościową, tj. dane muszą być dobrze aproksymowane przy użyciu różnych rozdzielczości.
- ▶ By potwierdzić tę intuicję, zaproponowaliśmy sztuczne zbiory danych posiadające wspomnianą własność i wykonaliśmy na nich eksperymenty.
- ▶ Porównaliśmy proponowany algorytm (wiele estymatorów na klasę) z algorytmem podstawowym (z 1 estymatorem na klasę).
- ▶ Główny wniosek: proponowany algorytm zwraca lepsze wyniki niż wersja podstawowa.

## Eksperymenty na referencyjnych danych - porównanie wersji algorytmu

- ▶ Algorytm był testowany na 19 popularnych zbiorach danych (nie dobieranych pod względem posiadania jakichkolwiek specjalnych własności).
- ▶ Porównaliśmy wersje algorytmu:  $N = 2$  z  $N = 1$ .
- ▶ Proponowany algorytm  $N = 2$  osiągnął statystycznie istotnie lepsze wyniki niż wersja podstawowa  $N = 1$  (rangowy test Wilcoxsona [Demsar06],  $p \approx .02$ ).
- ▶ Różnica między błędami klasyfikacji jest względnie duża dla części zbiorów danych, a dla innych praktycznie znika, czasami proponowany algorytm osiąga wynik gorszy niż wersja podstawowa.

---

 [Demsar06] J. Demsar: Statistical comparisons of classifiers over multiple data sets, **Journal of Machine Learning Research** 7 (2006) 1–30.



# Eksperymenty na referencyjnych danych - porównanie z popularnymi algorytmami

- ▶ Porównaliśmy wersję  $N = 2$  algorytmu z Naive Bayes, Nearest Neighbor, LDA, SVM
  - ▶ Ogólna skuteczność algorytmu stat. istotnie lepsza niż algorytmów Naive Bayes i 1-NN, wyniki algorytmów SVM, LDA i 5-NN nie były stat. istotnie różne (test Holma).
  - ▶ Bezpośrednie porównanie z alg SVM: proponowany alg. lepszy na *image segmentation* ( $p \approx 5 \cdot 10^{-4}$ ), *satellite image* ( $p \approx 1 \cdot 10^{-3}$ ) i *vowel Deterding* ( $p \approx 1 \cdot 10^{-3}$ ), SVM lepszy na: *vehicle silhouette* ( $p \approx 6 \cdot 10^{-5}$ ), *waveform* ( $p \approx 1 \cdot 10^{-3}$ ).
- ▶ Porównaliśmy z wynikami literaturowymi

# Spis treści

Wyniki poprzednie

**Nowa wersja algorytmu z funkcją Information Loss**

Wartość optymalizowanego błędu w trakcie działania algorytmu

Stosowanie różnych przekształceń przestrzeni

Porównanie algorytmów z  $N=1,2,3,4,5,7,10$

Dobieranie szerokości jądra do miejsca w przestrzeni

Plany

# Zmiany w algorytmie

- ▶ Poprawione 3 (dość subtelne) błędy w algorytmie
- ▶ Nowa wersja algorytmu (na niej przeprowadzaliśmy wszystkie kolejne eksperymenty) w porównaniu ze starą:
  - ▶ Cechy:
    - ▶ W pełni deterministyczna (stosujemy Leave One Out Cross-Validation zamiast 10-fold)
    - ▶ Optymalizujemy miarę Information Loss zamiast MSE (ta pierwsza powinna lepiej przybliżać błąd klasyfikacji)
    - ▶ Brak arbitralnie przyjętego parametru określającego dolne ograniczenie na możliwe wartości parametru wygładzania
  - ▶ Nie ma kolosalnej różnicy w porównaniu z poprzednimi wynikami eksperymentalnymi, ale algorytm jest prostszy
- ▶ Ciekawostka techniczna: eksperymenty przeprowadzane w sposób rozproszony na klastrze

## Originalna funkcja błędu MSE

$$\text{MSE}(\mathcal{D}^v, \mathbf{h}) = \frac{1}{|\mathcal{D}^v|} \sum_{\mathbf{x} \in \mathcal{D}^v} \sum_{k=1}^c (\hat{p}_k(\mathbf{x}; \mathbf{h}) - \mathbf{t}_k(\mathbf{x}))^2,$$

gdzie

$$\hat{p}_k(\mathbf{x}; \mathbf{h}) = \frac{\hat{f}_k(\mathbf{x}; \mathbf{h}) P_k}{\hat{f}(\mathbf{x}; \mathbf{h})}.$$

Występujące w powyższym wzorze estymatory gęstości opisane są wzorami

$$\hat{f}(\mathbf{x}; \mathbf{h}) = \sum_l \hat{f}_l(\mathbf{x}; \mathbf{h}) P_l,$$

oraz

$$\hat{f}_k(\mathbf{x}; \mathbf{h}) = \frac{1}{N_k} \sum_{i=1}^{N_k} \hat{f}(\mathbf{x}; h_{k,i}),$$

gdzie

$$\hat{f}(\mathbf{x}; h_{k,i}) = \frac{1}{|\mathcal{D}_k|} \sum_{\mathbf{x}' \in \mathcal{D}_k} \frac{1}{(h_{k,i})^d} \phi\left(\frac{\mathbf{x} - \mathbf{x}'}{h_{k,i}}\right).$$

## Funkcja błędu Information Loss

$$\begin{aligned} \text{IL}(\mathcal{D}^v, \mathbf{h}) &= \frac{1}{|\mathcal{D}^v|} \sum_{\mathbf{x} \in \mathcal{D}^v} \sum_{k=1}^c \mathbf{t}_k(\mathbf{x}) (-\log_a \hat{p}_k(\mathbf{x}; \mathbf{h})) \\ &= \frac{1}{|\mathcal{D}^v|} \sum_{\mathbf{x} \in \mathcal{D}^v} -\log_a \hat{p}_{\text{class}(\mathbf{x})}(\mathbf{x}; \mathbf{h}), \end{aligned}$$

gdzie

$$\hat{p}_k(\mathbf{x}; \mathbf{h}) = \frac{\hat{f}_k(\mathbf{x}; \mathbf{h}) P_k}{\hat{f}(\mathbf{x}; \mathbf{h})} \cdot (1 - r) + \frac{r}{c}.$$

oraz

$$\hat{f}(\mathbf{x}; \mathbf{h}) = \sum_l \hat{f}_l(\mathbf{x}; \mathbf{h}) P_l,$$

Natomiast  $\text{class}(\mathbf{x})$  oznacza klasę punktu  $\mathbf{x}$  ze zbioru treningowego, a  $r \in [0, 1]$  to pewna mała stała, np.  $r = 0.001$  lub  $r = 0.01$ .

# Spis treści

Wyniki poprzednie

Nowa wersja algorytmu z funkcją Information Loss

Wartość optymalizowanego błędu w trakcie działania algorytmu

Stosowanie różnych przekształceń przestrzeni

Porównanie algorytmów z  $N=1,2,3,4,5,7,10$

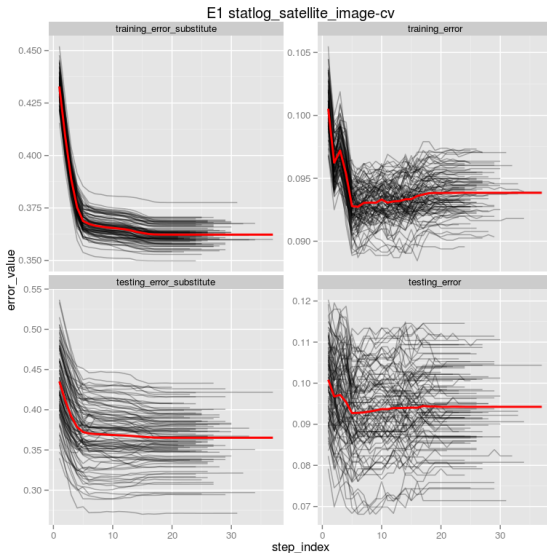
Dobieranie szerokości jądra do miejsca w przestrzeni

Plany

# Opis

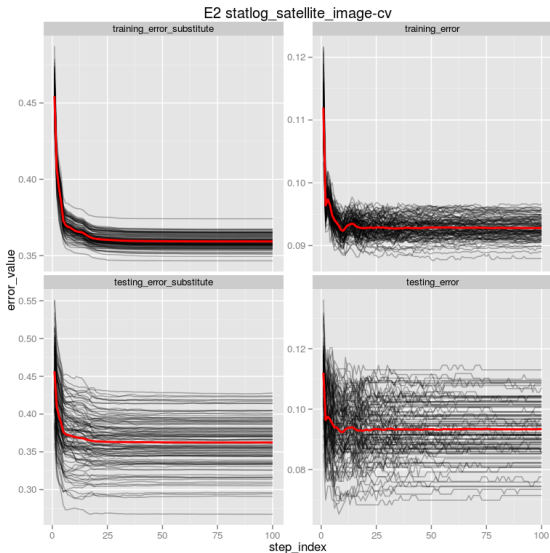
- ▶ Przeprowadziliśmy eksperymenty na 19 popularnych referencyjnych zb. danych. Dla każdego z nich zrobiliśmy 10-krotnie powtórzony eksperyment 10-fold cross validation (w sumie 100 eksperymentów na zb. danych).
- ▶ Oznaczenia:
  - ▶  $E1$ ,  $E2$  – wersja z 1 lub 2 estymatorami
  - ▶ błędy:
    - ▶ training\_error\_substitute – błąd Information Loss na zb. treningowym
    - ▶ training\_error – błąd klasyfikacji na zb. treningowym
    - ▶ testing\_error\_substitute, testing\_error – analogicznie na zb. testowym
  - ▶ oś X: numer kroku iteracyjnego, oś Y – wielkość błędu
  - ▶ na wykresach dla pojedynczej wersji algorytmu: czerwona linia – błąd uśredniony
  - ▶ na wykresach porównujących 2 wersje algorytmu: długość „wąsa”:  $2 \times$  odchylenie standardowe pomiarów dla danego kroku iteracyjnego

# Przykładowy wykres

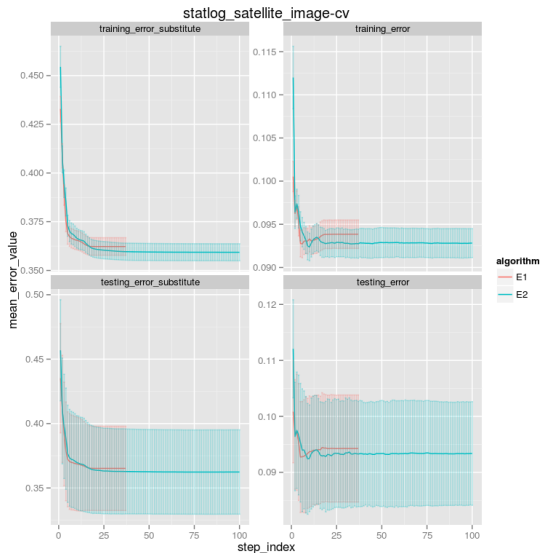




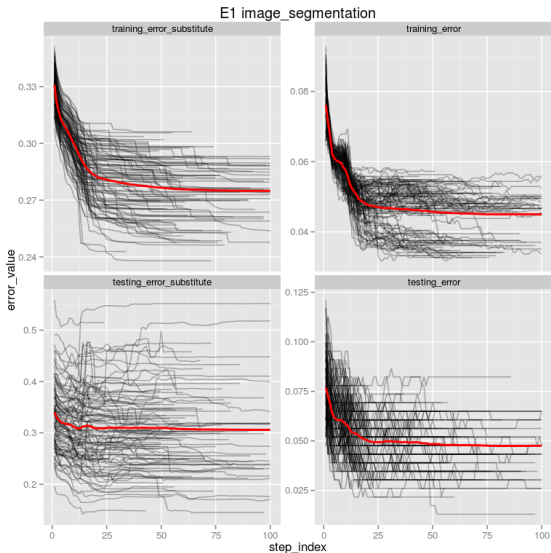
# Przykładowy wykres



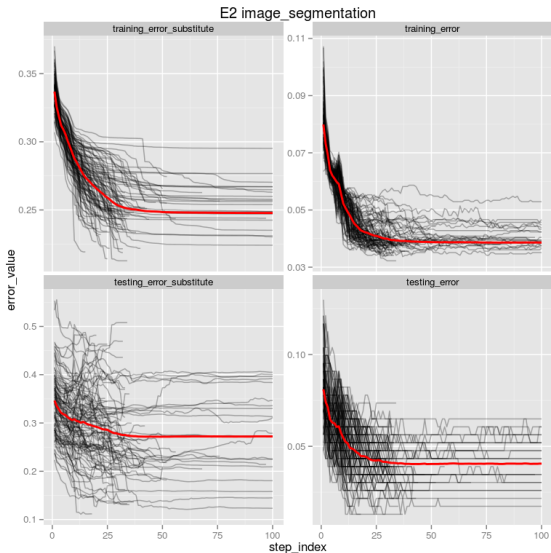
# Przykładowy wykres



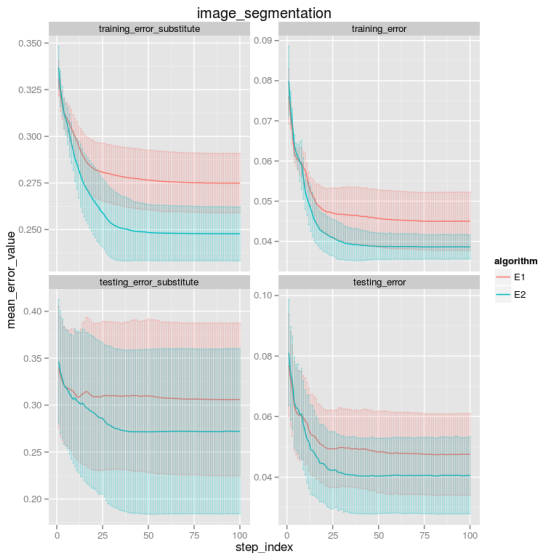
# Przykładowy wykres



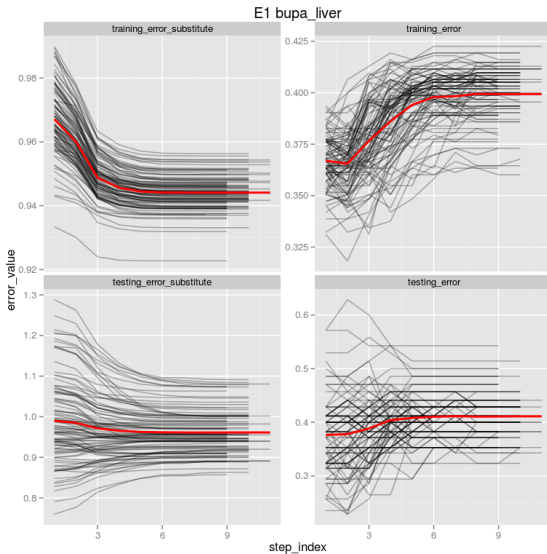
# Przykładowy wykres



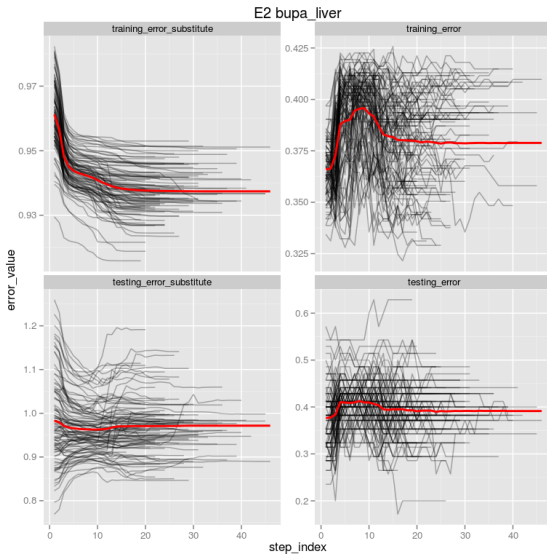
# Przykładowy wykres



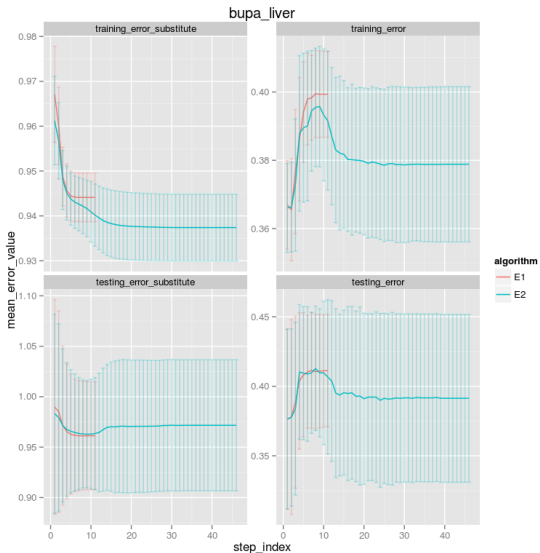
# Przykładowy wykres



# Przykładowy wykres



# Przykładowy wykres





# Wnioski

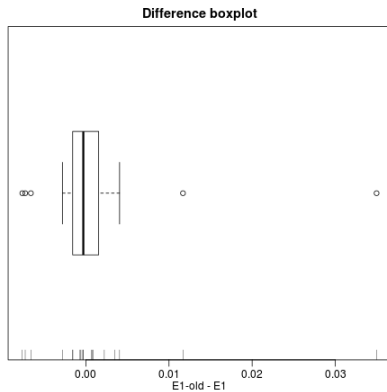
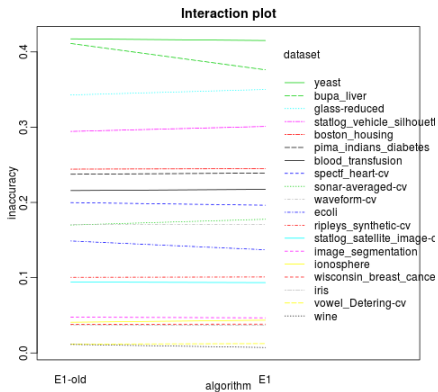
- ▶ Zgodnie z oczekiwaniami błąd Information Loss na zb. treningowym zawsze maleje
- ▶ Błędy na zb. treningowym są mniej chaotyczne niż na testowym
- ▶ Średnie błędy na zb. treningowym wydają się dobrym przybliżeniem średnich błędów na zb. testowym - czy to jednak przekłada się na wartości nieuśrednione?
- ▶ Zastosowanie metody „early stopping” może potencjalnie dać dobre rezultaty

## Zastosowanie „early stopping”

- ▶ Wykonujemy normalną optymalizację, ale jako najlepsze wybieramy te parametry dla tej iteracji, dla której błąd klasyfikacji był minimalny.
- ▶ Wyniki:
  - ▶ Na niektórych zb. danych jest lepiej a na innych gorzej (w nawiasach różnica; pomijam zbiory o różnicy mniejszej niż 0.005):
    - ▶ nowa wersja  $N = 1$  lepsza na: *bupa\_liver* (0.035), *ecoli* (0.012)
    - ▶ nowa wersja  $N = 1$  gorsza na: *glass-reduced* (0.007), *sonar-averaged* (0.01), *sonar-averaged-cv* (0.008), *spectf\_heart* (0.016), *statlog\_vehicle\_silhouette* (0.007)
    - ▶ nowa wersja  $N = 2$  lepsza na: *boston\_housing* (0.015), *spectf\_heart* (0.012), *vowel\_Detering* (0.186)
    - ▶ nowa wersja  $N = 2$  gorsza na: *glass-reduced* (0.009), *spectf\_heart* (0.01)
  - ▶ Ogólnie, nie ma stat. istotnej różnicy między algorytmami (biorąc pod uwagę wszystkie zb. danych)
- ▶ Czemu wyniki nie są lepsze?
  - ▶ prawdopodobnie ze względu na chaotyczność błędu klasyfikacji na zb. treningowym: wersja uśredniona jest gładka, ale poszczególne przebiegi są dość „poszarpane”.

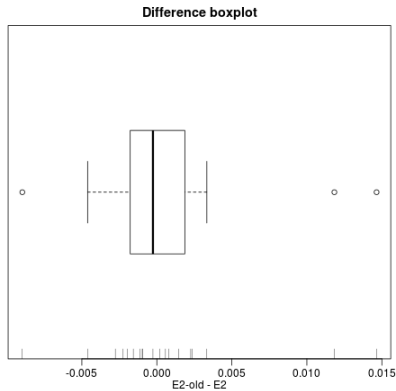
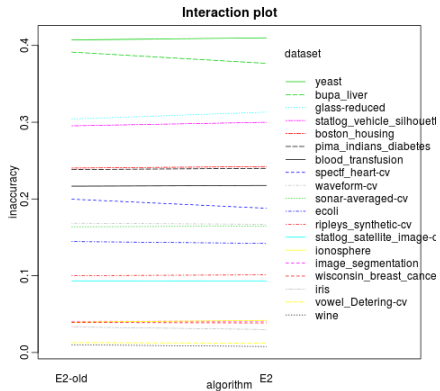
# Porównanie dla N=1

Non-parametric 2 paired groups comparison  
Wilcoxon test: p-value=1, (pseudo)median=2.01e-06, conf.int=[-0.0019, 0.00214]



# Porównanie dla N=2

Non-parametric 2 paired groups comparison  
Wilcoxon test: p-value=0.953, (pseudo)median=-6.1e-05, conf.int=[-0.00155, 0.00186]



# Spis treści

Wyniki poprzednie

Nowa wersja algorytmu z funkcją Information Loss

Wartość optymalizowanego błędu w trakcie działania algorytmu

**Stosowanie różnych przekształceń przestrzeni**

Porównanie algorytmów z  $N=1,2,3,4,5,7,10$

Dobieranie szerokości jądra do miejsca w przestrzeni

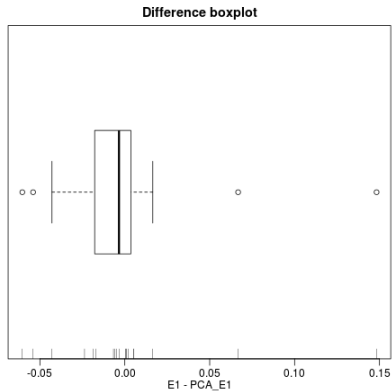
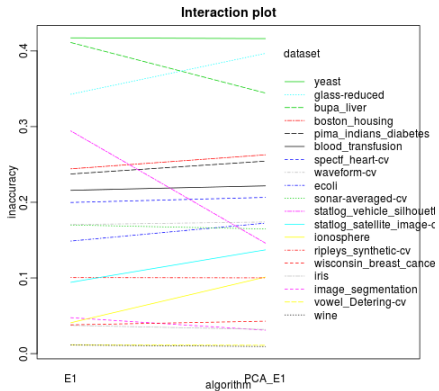
Plany

## Opis podejścia

- ▶ Porównujemy wersję z przekształceniem PCA i ze standardyzacją.
- ▶ Tworzymy algorytm „łączony”, w którym sprawdzamy na zb. treningowym (za pomocą CV), który z algorytmów sprawuje się najlepiej i jego używamy do końcowej ewaluacji na zb. testowym.

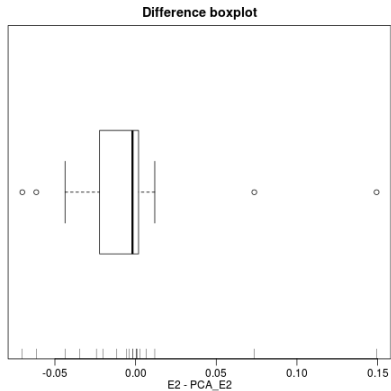
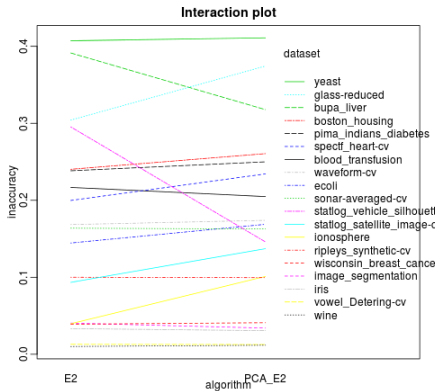
# Porównanie dla N=1

Non-parametric 2 paired groups comparison  
Wilcoxon test: p-value=0.395, (pseudo)median=-0.0033, conf.int=[-0.0203, 0.00529]



# Porównanie dla N=2

Non-parametric 2 paired groups comparison  
Wilcoxon test: p-value=0.312, (pseudo)median=-0.00498, conf.int=[-0.0227, 0.00402]





# Spostrzeżenia i wnioski

- ▶ Spostrzeżenia
  - ▶ Na jednych zbiorach pierwsza wersja jest lepsza, a na innych druga.
  - ▶ Żadna wersja nie jest „generalnie” lepsza (brak stat. istotnej różnicy).
- ▶ Wniosek: należałoby dobierać przekształcenie do zbioru danych.

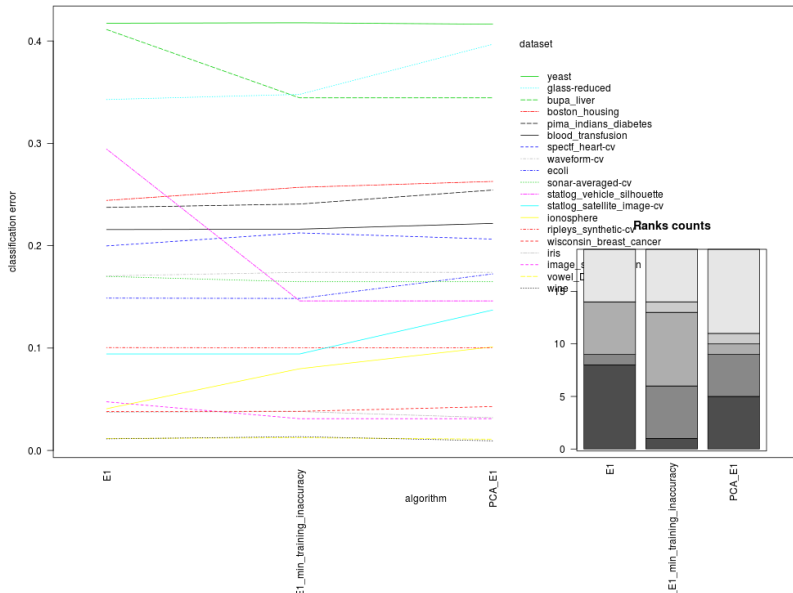
# Eksperymenty

Przetestujmy algorytm „łączony” w różnych wariantach:

- ▶ Optymalizacja Information Loss vs. optymalizacja błędu klasyfikacji
  - ▶ bezpośrednio porównanie błędów: brak stat. istotnych różnic
  - ▶ w kontekście wyników literaturowych: **optymalizacja błędu klasyfikacji daje lepsze rezultaty** niż optymalizacja Information Loss (ta druga miara jest  $\leq$  od tej drugiej na na wszystkich zb. danych)
- ▶ Porównanie wersji podstawowych algorytmu z wersją „łączoną”
  - ▶ bezpośrednio porównanie błędów: brak stat. istotnych różnic (ale wersja ze standardyzacją ma najlepszy „ranking”)
  - ▶ w kontekście wyników literaturowych: wersja „łączona” plasuje się gdzieś między wynikiem Std i PCA, ale bliżej tego lepszego, a czasami ma ten sam wynik co lepsza opcja. Dzięki temu:
    - ▶ „łączona” wersja  $N = 2$  ma 7 wyników w górnym kwantylu rzędu 0.5 (o 2 lepiej niż Std  $N = 1$  i o 1 lepiej niż PCA  $N = 2$ ),
    - ▶ „łączona” wersja  $N = 1$  ma 5 (o 1 lepiej niż Std  $N = 1$  i o 2 lepiej niż PCA  $N = 1$ )

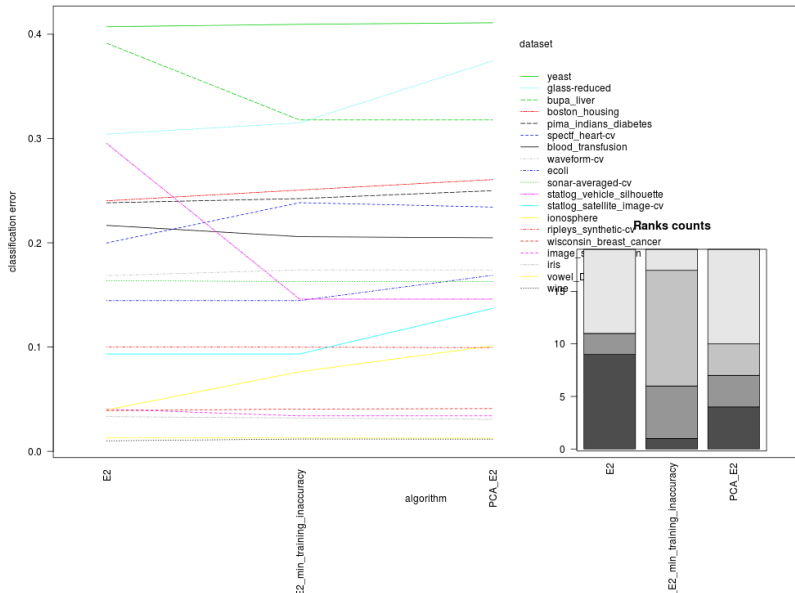
# Porównanie dla N=1

Non-parametric n blocked groups comparison (standard tests)  
 (Friedman test: p-value=0.5894, chi-squared\_2=1.057; Iman & Davenport test: p-value=0.6018, F\_(2,36)=0.5151)  
 Mean ranks of consecutive groups: 1.816, 2.105, 2.079

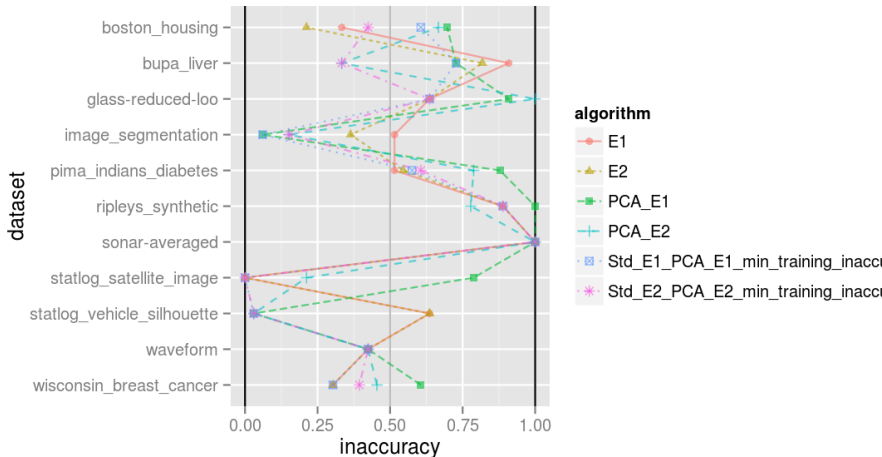


# Porównanie dla N=2

Non-parametric n blocked groups comparison (standard tests)  
 (Friedman test: p-value=0.5939, chi-squared\_2=1.042; lman & Davenport test: p-value=0.6062, F\_(2,36)=0.5076)  
 Mean ranks of consecutive groups: 1.895, 1.921, 2.184



# Wersja „łączona” w kontekście wyników literaturowych



# Wnioski

- ▶ W zasadzie można używać tej metody żeby polepszyć trochę wyniki (na to by wskazywały wyniki w kontekście literaturowym), ale z drugiej strony wyniki bezpośredniego porównania błędów wskazują, że nie ma stat. istotnej różnicy.
- ▶ Do do analizy porównawczej  $N = 1$  i  $N = 2$ , ta metoda się nie nadaje, bo za bardzo zaburza wyniki – nie wiemy, czy zmiany są wynikiem dodania dodatkowego estymatora, czy nie.

# Spis treści

Wyniki poprzednie

Nowa wersja algorytmu z funkcją Information Loss

Wartość optymalizowanego błędu w trakcie działania algorytmu

Stosowanie różnych przekształceń przestrzeni

**Porównanie algorytmów z  $N=1,2,3,4,5,7,10$**

Dobieranie szerokości jądra do miejsca w przestrzeni

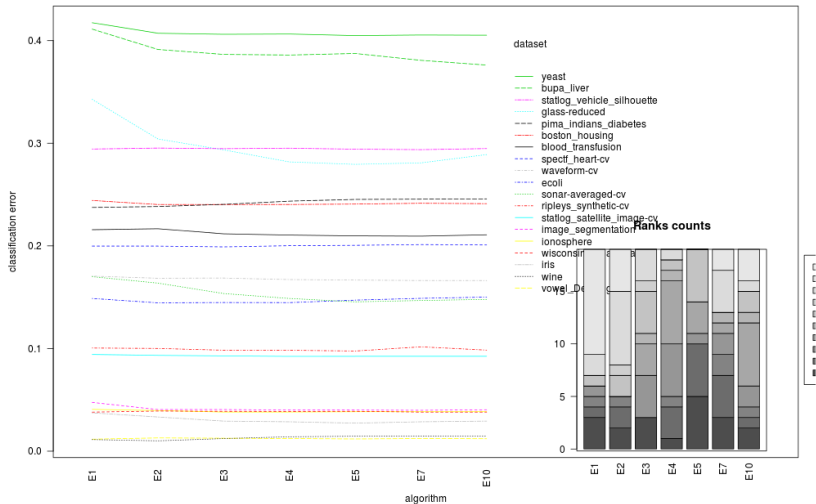
Plany

- ▶ Wykonaliśmy eksperymenty dla  $N = 1, 2, 3, 4, 5, 7, 10$ 
  - ▶ Ogólne porównanie algorytmów (ANOVA): są stat. istotnie różne (na poziomie  $p = 0.009$ ).
    - ▶ Rankingi w kolejności od najlepszego do najgorszego: E5, E4, E7, E3, E10, E2, E1.
  - ▶ Porównanie wszystkich par algorytmów (za pomocą różnych multiple testing procedures) pokazuje, że różnica między E5 i E1 jest istotna, a reszta różnic jest nieistotna.
  - ▶ Porównanie par: algorytm  $N = 1$  vs inny algorytm dla wszystkich takich par:  $N = 1$  jest stat. istotnie gorsze od innych i to na poziomach  $p < 0.01$  - wyjątkiem jest  $N = 2$  ( $p = 0.012$ ),  $N = 7$  ( $p = 0.032$ ),  $N = 10$  ( $p = 0.026$ ).
- ▶ Podsumowanie:
  - ▶  $N = 5$  ma najlepsze wyniki
  - ▶ To, że algorytmy  $N > 2$  też mają wyniki lepsze i od  $N = 1$  (a nawet lepsze również od  $N = 2$ ) potwierdza, że podejście dodawania dodatkowych estymatorów ma sens.



# Porównanie

Non-parametric n blocked groups comparison (standard tests)  
 (Friedman test: p-value=0.01219, chi-squared\_6=16.31; Iman & Davenport test: p-value=0.009367, F\_{(6,108)}=3.005)  
 Mean ranks of consecutive groups: 5.132, 4.947, 3.947, 3.368, 2.895, 3.605, 4.105



# Spis treści

Wyniki poprzednie

Nowa wersja algorytmu z funkcją Information Loss

Wartość optymalizowanego błędu w trakcie działania algorytmu

Stosowanie różnych przekształceń przestrzeni

Porównanie algorytmów z  $N=1,2,3,4,5,7,10$

Dobieranie szerokości jądra do miejsca w przestrzeni

Plany

# Opis metody

Metoda modyfikacji param. wygładzania opisana w [Kulczycki05].

Ogólnie:

- ▶ Dla wstępnej estymacji gęstości (za pomocą KDE), określ średnią geometryczną estymacji  $\tilde{s}$  ze wszystkich punktów uczących
- ▶ Na podstawie powyższej średniej dobierz parametry  $s_i$  modyfikujące dla każdego punktu uczącego parametr wygładzania  $h$ . Jeśli gęstość jest większa niż średnia - zmniejsz  $h$  (wyostrz estymację), w p.p. zwiększ  $h$  (wygładź estymację).
  - ▶ Jest tu stosowany parametr  $c$ , najczęściej  $c \in [0, 1]$ , im większy, tym większa intensywność modyfikacji.

## Modyfikacje wzorów

$$\hat{f}_k(\mathbf{x}; \mathbf{a}) = \frac{1}{N_k} \sum_{i=1}^{N_k} \hat{f}_k(\mathbf{x}; h_{k,i}, c_k),$$

gdzie

$$\hat{f}_k(\mathbf{x}; h_{k,i}, c_k) = \frac{1}{|\mathcal{D}_k|} \sum_{\mathbf{x}' \in \mathcal{D}_k} \frac{1}{(h_{k,i} s_k(\mathbf{x}', c_k))^d} \phi\left(\frac{\mathbf{x} - \mathbf{x}'}{h_{k,i} s_k(\mathbf{x}', c_k)}\right).$$

Natomiast

$$s_k(\mathbf{x}', c_k) = \left(\frac{\hat{f}_{0k}(\mathbf{x}')}{\tilde{s}_k}\right)^{-c_k},$$

gdzie

$$\tilde{s}_k = \exp\left(\frac{1}{|\mathcal{D}_k|} \sum_{\mathbf{x}' \in \mathcal{D}_k} \ln(\hat{f}_{0k}(\mathbf{x}'))\right).$$

## Opis eksperymentów

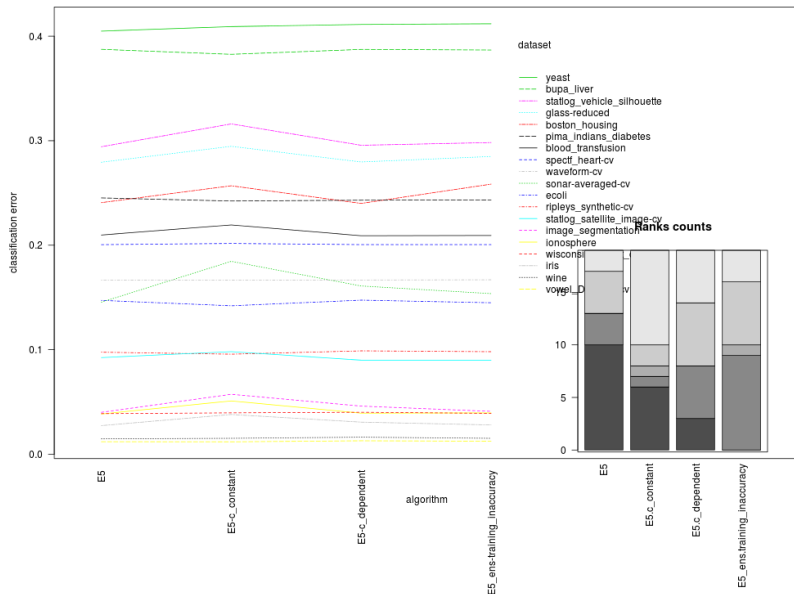
- ▶ Porównanie wersji standardowych  $N = 1$ ,  $N = 2$ ,  $N = 5$  z modyfikacjami, gdy:
  - ▶  $c = 1/2$
  - ▶  $c$  jest dobierane dla każdej klasy osobno i optymalizowane (startując z punktu  $1/2$ )

# Eksperymenty

- ▶ Czy któraś z wersji algorytmu dla danego  $N$  jest generalnie lepsza?
  - ▶ Nie, brak stat. istotnych różnic (test Friedmana)
- ▶ Porównanie par: wersja podstawowa  $N = k$  vs. modyfikacja z  $c$  (test Wilcoxsona)
  - ▶ Średnio, wersja podstawowa jest lepsza od wersji zmodyfikowanych,
    - ▶ ale dla  $c$  optymalizowanego oprócz  $N = 5$  **nie jest to nigdy różnica stat. istotna**,
    - ▶ dla  $c = 1/2$  jest to **zawsze różnica stat. istotna** ( $p < 0.05$ ).
  - ▶ Mimo to, dla niektórych zb. danych wersje zmodyfikowane są trochę lepsze.
- ▶ Porównaliśmy też wyniki algorytmu „łączonego” (jak poprzednio), gdzie dla danego  $N = k$  łączymy wszystkie wersje z  $k$ , np. dla  $N = 1$ : ( $N = 1$ ) z ( $N = 1, c$  optymalizowane) z ( $N = 1, c = 1/2$ ) z poszczególnymi wersjami.
  - ▶ Wyniki: **różnice między wersją łączoną i wersjami podstawowymi są stat. nieistotne**. Dodatkowo, wersja  $E = k$  zawsze była najlepsza (ale oczywiście stat. nieistotne).

# Wyniki porównania alg. łączonego dla $N = 5$

Non-parametric n blocked groups comparison (standard tests)  
 (Friedman test: p-value=0.129, chi-squared\_3=5.667; Iman & Davenport test: p-value=0.1268,  $F_{(3,54)}=1.987$ )  
 Mean ranks of consecutive groups: 1.895, 2.763, 2.684, 2.658



# Spis treści

Wyniki poprzednie

Nowa wersja algorytmu z funkcją Information Loss

Wartość optymalizowanego błędu w trakcie działania algorytmu

Stosowanie różnych przekształceń przestrzeni

Porównanie algorytmów z  $N=1,2,3,4,5,7,10$

Dobieranie szerokości jądra do miejsca w przestrzeni

Plany



# Plany

- ▶ Przyjrzeć się rozkładowi gęstości odległości między punktami, i  $k$ -najbliższymi sąsiadami na zbiorach danych (cel: automatyczny dobór ilości estymatorów).
- ▶ Porównać algorytm z wersją SVM z bardziej wyszukany doborem parametrów

# Plany

Dziękuję za uwagę!