

Rozpoznawanie obrazów

Laboratorium nr 2

Tomasz Gałęcki

21 marca 2019

1 Zmiana etykiet

```
>> labels = unique(train(:,1));  
>> labels'  
ans =  
     1     2     3     4     5     6     7     8
```

2 Usunięcie odstających próbek

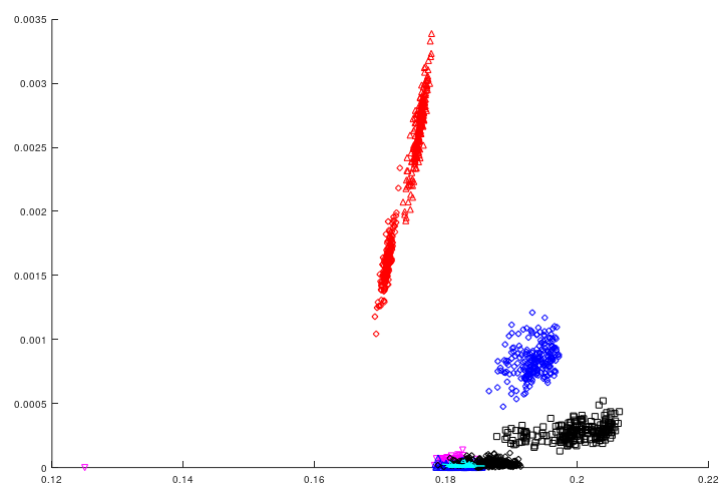
Usunięte próbki: 186 (odstająca w górę), 641 (odstająca w dół).

2.1 Pierwsza próbka

Większość cech ma ogromne różnice między średnią a medianą (np. dla cechy nr 5 to aż 12 rzędów wielkości).

```
>> [mv midx] = max(train)  
mv =  
      8.00000      5.12127      25.82285      382.13906      380.85837      145296.77100  
      ↳ 1934.12276      16.57025  
midx =  
      77      186      186      186      186      186      186      186
```

2.2 Druga próbka



Rysunek 1: Poszukiwanie drugiej próbki do usunięcia; plot2features(train, 2, 3)

```
>> [mv midx] = min(train)  
mv =
```

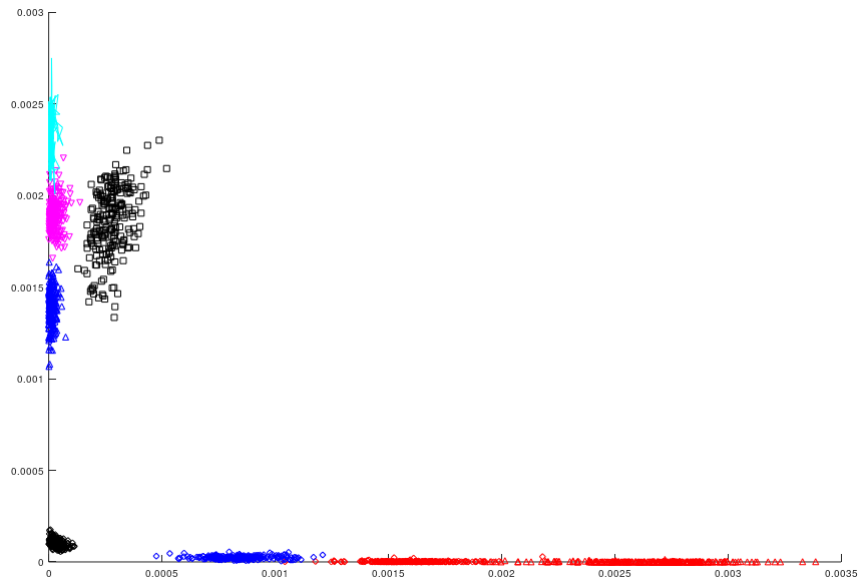
```

1.00000  0.12500  0.00000  0.00000  0.00000  -0.00000  -0.00000  -0.00000
midx =
    58    641    641    641    641    392    25    538

```

3 Wybór cech do klasyfikacji

Para cech nr 2 i 3 (kolumny 3 i 4).



Rysunek 2: Para cech do klasyfikacji; plot2features(train, 3, 4)

3.1 Wyniki

Zastosowana szerokość okna: 0.001.

```

>> base_ercf
base_ercf =
% pdf_indep pdf_multi pdf_parzen
0.021382 0.020833 0.016996

```

4 Redukcja zbioru uczącego

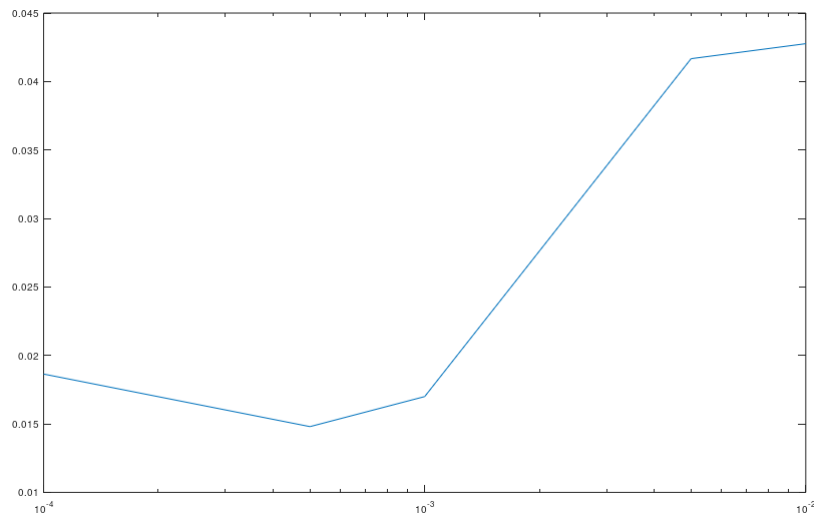
Wyniki dla *pdf_indep* oraz *pdf_multi*.

<i>Test</i>	<i>min</i>	<i>max</i>	<i>mean</i>	<i>std</i>
0.1				
<i>pdf_indep</i>	$2.0833e-02$	$2.9605e-02$	$2.5110e-02$	$4.0362e-03$
<i>pdf_multi</i>	$2.0833e-02$	$3.1250e-02$	$2.5439e-02$	$4.0026e-03$
<i>pdf_parzen</i>	$3.5088e-02$	$3.9474e-02$	$3.7719e-02$	$1.9919e-03$
0.25				
<i>pdf_indep</i>	$2.1382e-02$	$2.4123e-02$	$2.2588e-02$	$1.1886e-03$
<i>pdf_multi</i>	$2.0285e-02$	$2.4123e-02$	$2.2368e-02$	$1.5213e-03$
<i>pdf_parzen</i>	$2.9057e-02$	$3.1250e-02$	$2.9825e-02$	$9.1739e-04$
0.5				
<i>pdf_indep</i>	$2.0285e-02$	$2.4123e-02$	$2.2039e-02$	$1.4191e-03$
<i>pdf_multi</i>	$1.9189e-02$	$2.2478e-02$	$2.1272e-02$	$1.3651e-03$
<i>pdf_parzen</i>	$1.7544e-02$	$2.0833e-02$	$1.9737e-02$	$1.2857e-03$

Wynika z tego, że klasyfikacja metodą Parzena jest bardzo nieefektywna dla mocno okrojonych zbiorów uczących. Z drugiej strony, *pdf_multi* sprawdza się najlepiej dla małych zbiorów uczących (albo identycznie z *pdf_indep* dla najmniejszych zbiorów).

5 Dobór parametru h_1

h_1	Error
0.0001	0.018640
0.0005	0.014803
0.001	0.016996
0.005	0.041667
0.01	0.042763



Rysunek 3: Wykres zależności szerokości okna od stopy błędu.

6 Zmiana prawdopodobieństwa *apriori*

Szerokość okna dla *pdf_parzen* wynosiła 0.001.

```
apriori = [0.165 0.085 0.085 0.165 0.165 0.085 0.085 0.165];
parts = [1.0 0.5 0.5 1.0 1.0 0.5 0.5 1.0];
>> base_ercf
base_ercf =
% pdf_indep pdf_multi pdf_parzen
0.017544 0.016082 0.013889
>> cxfmRes
cxfmRes =
{
[1,1] = % pdf_indep
228 0 0 0 0 0 0 0
0 113 0 0 0 1 0 0
6 0 105 0 0 0 3 0
1 0 2 225 0 0 0 0
0 0 0 0 227 0 0 1
0 5 0 0 0 109 0 0
1 0 3 0 0 0 110 0
0 1 0 0 0 0 0 227
[1,2] = % pdf_multi
228 0 0 0 0 0 0 0
0 113 0 0 0 1 0 0
```

```

5      0    106      0      0      0      3      0
2      0      1    225      0      0      0      0
0      0      0      0    227      0      0      1
0      4      0      0      0    110      0      0
0      0      4      0      0      0    110      0
0      1      0      0      0      0      0    227
[1,3] = % pdf_parzen
224      0      1      3      0      0      0      0
0    112      0      0      0      1      0      1
2      0    108      1      0      0      3      0
0      0      0    228      0      0      0      0
0      0      0      0    228      0      0      0
0      3      0      0      0    111      0      0
0      0      3      0      0      0    111      0
0      1      0      0      0      0      0    227
}

```

Z samej confMx dla zmienionych prawdopodobieństw nie doprowadziła mnie do żadnych wniosków. Wykonałem analogiczny eksperyment dla pierwotnych prawdopodobieństw:

```

apriori = repmat(0.165, 1, 8);
parts = repmat(1.0, 1, 8);
cxfmRes =
{
    [1,1] = % pdf_indep
    228      0      0      0      0      0      0      0
    0    226      0      0      0      1      0      1
    12      0    207      0      0      0      9      0
    1      0      2    225      0      0      0      0
    0      0      0      0    227      0      0      1
    0      5      0      0      0    223      0      0
    1      0      5      0      0      0    222      0
    0      1      0      0      0      0      0    227
    [1,2] = % pdf_multi
    228      0      0      0      0      0      0      0
    0    226      0      0      0      1      0      1
    12      0    207      0      0      0      9      0
    1      0      2    225      0      0      0      0
    0      0      0      0    227      0      0      1
    0      4      0      0      0    224      0      0
    0      0      6      0      0      0    222      0
    0      1      0      0      0      0      0    227
    [1,3] = % pdf_parzen
    224      0      1      3      0      0      0      0
    0    226      0      0      0      1      0      1
    2      0    215      1      0      0      10      0
    0      0      0    228      0      0      0      0
    0      0      0      0    228      0      0      0
    0      3      0      0      0    225      0      0
    0      0      6      0      0      0    222      0
    0      3      0      0      0      0      0    225
}

```

Wynikałoby z tego, że mniejsza liczba klasyfikowanych próbek w klasach, które zawierały najwięcej pomyłek, skutkuje zmniejszeniem ogólnej stopy błędów w tych klasach.

7 Porównanie z *cls1nn*

Normalizacja **nie** była potrzebna. Widać to na wykresie na rysunku 2 - cechy są rozrzucone podobnie w obu osiach. Ponadto odchylenia standardowe były zbliżone ($9,2062e-04$, $9,5129e-04$, różnica rzędu 3,2%). Wynik

klasyfikatora *cls1nn* uplasował się pośrodku między *pdf_multi* a *pdf_parzen*.

```
>> mean(classResult != test(:,1))
%      cls1nn
ans = 0.018092
>> confMx(test(:, 1), classResult(:, 1))
ans = % cls1nn
228    0    0    0    0    0    0    0
  0   227    0    0    0    1    0    0
  1    0   211    2    0    2   12    0
  1    0    0   227    0    0    0    0
  0    0    0    0   228    0    0    0
  0    4    0    0    0   224    0    0
  0    0    8    0    0    0   220    0
  0    2    0    0    0    0    0   226
```

Dla szybkiego porównania:

```
>> base_ercf
base_ercf =
% pdf_indep pdf_multi pdf_parzen
0.021382 0.020833 0.016996
cxfmRes =
{
  [1,1] = % pdf_indep
    228    0    0    0    0    0    0    0
      0   113    0    0    0    1    0    0
      6    0   105    0    0    0    3    0
      1    0    2   225    0    0    0    0
      0    0    0    0   227    0    0    1
      0    5    0    0    0   109    0    0
      1    0    3    0    0    0   110    0
      0    1    0    0    0    0    0   227
  [1,2] = % pdf_multi
    228    0    0    0    0    0    0    0
      0   113    0    0    0    1    0    0
      5    0   106    0    0    0    3    0
      2    0    1   225    0    0    0    0
      0    0    0    0   227    0    0    1
      0    4    0    0    0   110    0    0
      0    0    4    0    0    0   110    0
      0    1    0    0    0    0    0   227
  [1,3] = % pdf_parzen
    224    0    1    3    0    0    0    0
      0   112    0    0    0    1    0    1
      2    0   108    1    0    0    3    0
      0    0    0   228    0    0    0    0
      0    0    0    0   228    0    0    0
      0    3    0    0    0   111    0    0
      0    0    3    0    0    0   111    0
      0    1    0    0    0    0    0   227
}
```