

Razpoznavanje Vzorcev

Poročilo druge laboratorijske vaje

Avtor: Bor Starčič

Mentorja: izr. prof. dr. Simon Dobrišek, asist. dr. Klemen Grm

Datum: December 3, 2023

Contents

1	Naloge	3
2	Gaussova standardizacija	3
3	Populacijska standardizacija	4
4	Kriterijska funkcija za povprečno razdaljo med razredi	5
5	Kriterijska funkcija uspešnosti razvrščanja	5
6	Povzetek	6

1 Naloge

1. Gaussova standardizacija
2. Populacijska standardizacija
3. Kriterijska funkcija povprečne medrazredne razdalje
4. Kriterijska funkcija uspešnosti razvrščanja

2 Gaussova standardizacija

Izvajamo postopek Gaussove standardizacije značilk za učni in testni nabor podatkov. Namen standardizacije je prilagoditi značilke tako, da imajo povprečje 0 in standardno odstopanje 1 kar omogoča izenačevanje lestvic različnih značilk. Na sliki so prikazane normalizirane vrednosti značilk prvega vzorca in vrednosti standardiziranih značilk.

```
[ -0.65624076   1.00985625   0.53592741  -0.95785223   0.27416913   0.54818157  
    0.47013237  -0.26430802   0.17998983  -0.40720134   0.48764773  -0.68636917  
    0.23592044   0.01856777  -0.36724187  -0.49895491   0.53257885  -0.56125552  
    0.18140166  -1.38004802   0.44992139  -0.85517843   0.27652233  -1.49346137  
    0.37271078  -0.45488748  -0.17635353  -1.08122713   0.82380534   0.6775146  
   -0.60872758  -0.51410549   0.82850212   0.01424277  -0.00206861  -0.24483052  
    0.58190078]  
  
[ -1.46371804e-16   3.25069601e-16  -4.98554901e-16  -1.54243285e-16  
    2.48319883e-17  -3.88709898e-17   3.48147437e-17   6.74136672e-17  
    2.55351296e-19   1.62610666e-17   2.93209901e-17   6.24722496e-17  
   -2.22932783e-16  -3.27642543e-16   4.95011439e-17  -4.74879395e-17  
    5.93710266e-17   1.67945750e-16   1.97399504e-16   7.58578385e-17  
    1.76155387e-18  -2.75064693e-16  -6.51978471e-18  -8.99798754e-17  
    8.46415530e-17  -9.97091298e-17  -4.54743650e-16  -1.73931240e-16  
    7.23495338e-17  -7.31636973e-18  -1.08109817e-16  -5.13885231e-17  
    9.59454738e-17   2.44197255e-16   1.05056704e-16  -7.57949259e-17  
   -1.01237537e-16]  
  
[ 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.  
  1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.]
```

Figure 1: Gaussova standardizacija vzorcev.

3 Populacijska standardizacija

Izvajamo postopek Gaussove standardizacije značilk za učni in testni nabor podatkov. Namen standardizacije je prilagoditi značilke tako, da imajo mediano 0 in interkvartilni razpon 1 kar omogoča izenačevanje lestvic različnih značilk. Na sliki so prikazane normalizirane vrednosti značilk prvega vzorca in vrednosti standardiziranih značilk.

```
[ -0.44      0.55766152  0.46480964 -0.57199373  0.47946537 -0.8071761
  0.68407154 -0.27977849  0.27692067 -0.5657138  0.54825756 -0.58063019
  0.31785472 -0.23055069 -0.04157913 -0.62484298  0.50333983 -0.69609721
  0.25182143 -1.12243778  0.54915899 -1.18413637  0.4962302  -1.29286839
  0.41375996 -0.63190085  0.01554097 -0.89659062  0.82402392  0.38318672
 -0.23699429 -0.52354541  0.72448199 -0.16100597  0.13777109 -0.22572751
  0.68955785]
[ 0.00000000e+00  0.00000000e+00 -1.00619044e-16  3.84641897e-17
  0.00000000e+00  0.00000000e+00  0.00000000e+00  4.28564790e-17
  0.00000000e+00 -1.18551155e-16 -1.72992927e-16  0.00000000e+00
  0.00000000e+00  7.83742910e-17  0.00000000e+00  1.40776876e-17
  0.00000000e+00 -1.13500721e-16  0.00000000e+00  5.22983553e-17
  0.00000000e+00 -8.67903839e-17  0.00000000e+00  0.00000000e+00
  0.00000000e+00 -1.35370688e-16 -1.69904645e-16  0.00000000e+00
  0.00000000e+00  0.00000000e+00  0.00000000e+00 -2.78885598e-17
  0.00000000e+00  0.00000000e+00  0.00000000e+00 -5.30589908e-17
  0.00000000e+00]
[1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.
 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.]
```

Figure 2: Populacijska standardizacija vzorcev.

Uspešnost klasifikatorja z algoritmom najbližjih sosedov za klasifikacijo z populacijsko standardiziranimi značilkami ima podobne rezultate kot pri Gaussovi standardizaciji.

```
Točnost klasifikacije z Gaussovo standardizacijo: 53.50%
Točnost klasifikacije z populacijsko standardizacijo: 53.22%
```

Figure 3: Natančnost klasifikatorja za standardizirane značilke.

4 Kriterijska funkcija za povprečno razdaljo med razredi

Izbira značilke se izvaja za klasifikacijo s pomočjo razdalje med razredi. Uporablja dve metodi: "naprej", kjer postopoma dodaja značilke, ki najbolj izboljšajo razločevanje med razredi, in "nazaj", kjer začne z vsemi značilkami in odstranjuje najmanj informativne. Cilj obeh metod je identificirati optimalno podmnožico značilke, ki povečujejo učinkovitost klasifikacijskega modela.

```
[107.24594679 103.5530022 104.46352086 100.56226022 84.07457419
94.87097112 95.39997072 96.54721996 91.17544414 96.97533025
99.09747642 98.08859643 86.31478595 93.71196083 96.62852034
95.69667631 99.67889095 96.32154588 102.27907508 102.58663521
82.82320066 91.35472765 96.57485899 96.681209 92.55344773
92.47994369 99.84180486 98.83068319 88.70287925 90.07631988
99.16788529 96.46362948 97.49461892 97.71533303 101.92977693
97.35104867 83.99621754]
```

Figure 4: Razdalje med razredi.

```
Izbrane značilke (Naprej): [ 0 2 1 19 18 34 3 26]
Izbrane značilke (Nazaj): [4, 8, 12, 20, 21, 28, 29, 36]
Natančnost klasifikatorja (Naprej): 43.30%
Natančnost klasifikatorja (Nazaj): 47.21%
```

Figure 5: Izbrane značilke in natančnost klasifikatorja.

5 Kriterijska funkcija uspešnosti razvrščanja

Izvajata se metodi "naprej" in "nazaj" za izbiro značilke, ki se uporabljajo v K-Nearest Neighbors klasifikatorju. Po izbiri značilke se klasifikator nauči na izbranih značilkah in preizkusi na testnem naboru podatkov, da se oceni njegova natančnost. Namen je izboljšati učinkovitost klasifikacije z identifikacijo in uporabo najbolj relevantnih značilke.

```
Izbrane značilke (Naprej): [0, 2, 1, 19, 18, 34, 3, 26]  
Izbrane značilke (Nazaj): [0, 1, 2, 3, 18, 19, 26, 34]  
Natančnost klasifikacije z izbranimi značilkami: 51.64%
```

Figure 6: Natančnost klasifikacije z izbranimi značilkami.

6 Povzetek

Uporabili smo značilke povprečna svetlost, varianca, simetričnost in kurtoza ter jih standardizirali z Gaussovo in populacijsko standardizacijo. To je omogočilo učinkovitejše in natančnejše razvrščanje slik v razrede. Za klasifikacijo smo uporabili algoritmom najbližjih sosedov, ki za klasifikacijo uporablja razdalje med podatkovnimi točkami, pri čemer napoved za novo točko temelji na številu najbližjih znanih točkah v učnem naboru podatkov. V nadaljevanju smo implementirali klasifikacijo z pomočjo razdalje med razredi z uporabo metod "naprej" in "nazaj", ki nam je podalo 8 najboljših značilk izmed 5 prvotnih in 32 drugo-rednih značilk.