

Daugiamačių duomenų vizualizacija

3 laboratorinio darbo ataskaita

Ignas Jatulis 1815628

Trečioji užduotis:

Atlikite pagrindinių komponentų analizę daugiamačiams duomenims. Nustatykite, kokia procentinė dalis nuo visos dispersijos tenka kiekvienai komponentei, kiek tenka dviem pagrindinėms. Vizualizuokite dvi pagrindines komponentes. Galima naudoti [Orange](#), Matlab, octave ar kitas programas.

Sprendimas:

Prieš nustatant kokią procentinę dalį tenka kiekvienai komponentai, klasės atributo duomenys buvo diskretizuoti pasitelkiant WEKA programinę įrangą. Klasės atributas *medv* buvo sutrauktas į penkis intervalus.

Toliau programinio paketo WEKA pagalba, buvo apskaičiuota kiekvieno atributų dispersija:

```
=== Attribute Selection on all input data ===

Search Method:
  Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 14 medv):
  Information Gain Ranking Filter

Ranked attributes:
0.804 13 LSTAT
0.603 6 RM
0.524 5 NOX
0.443 3 INDUS
0.388 11 PTRATIO
0.342 10 TAX
0.331 1 CRIM
0.252 8 DIS
0.207 9 RAD
0.197 7 AGE
0.167 2 ZN
0.104 12 B
0      4 CHAS

Selected attributes: 13,6,5,3,11,10,1,8,9,7,2,12,4 : 13
```

Toliau, pritaikius *PrincipalComponents* įvertinimo metodą, buvo gauta:

Koreliacijos matrica:

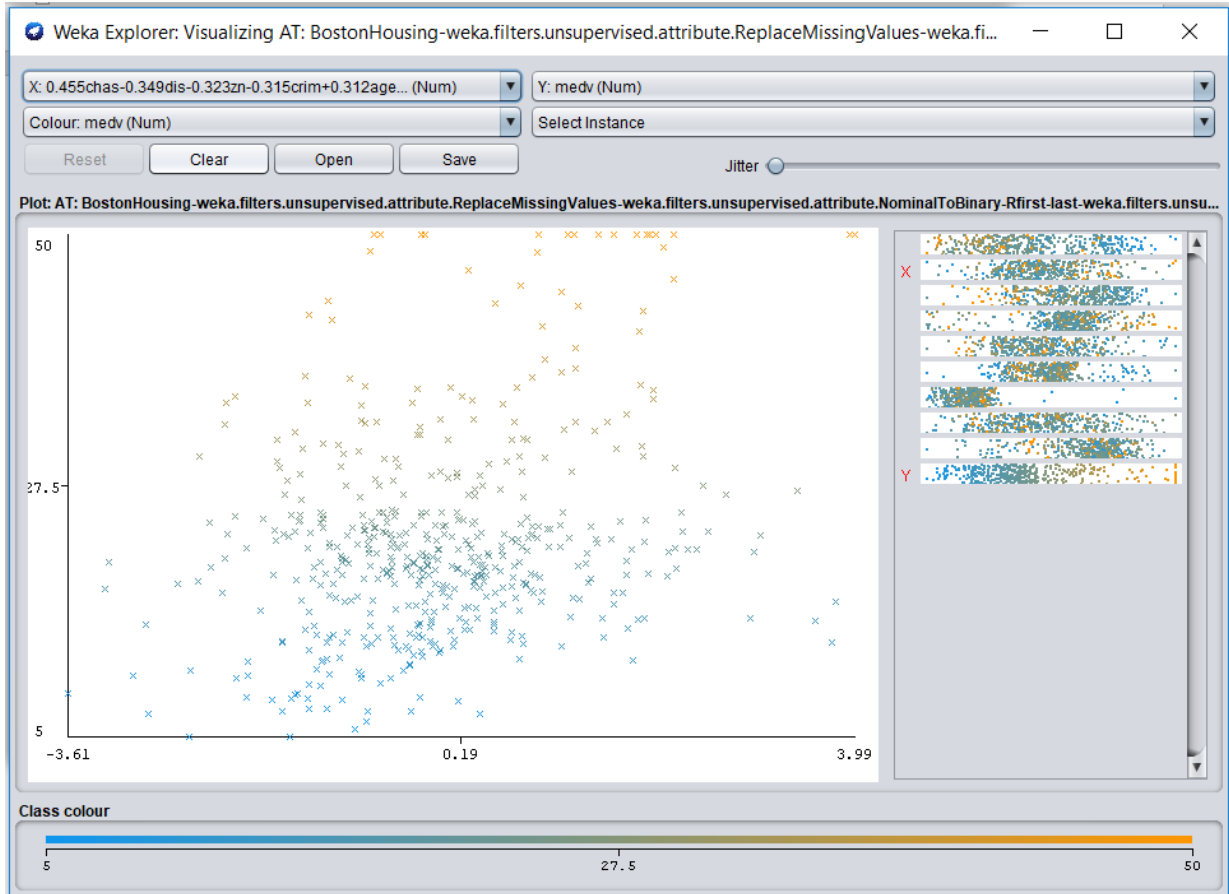
```
Correlation matrix
 1      -0.2    0.41 -0.06    0.42 -0.22    0.35 -0.38    0.63    0.58    0.29 -0.39    0.46
-0.2     1     -0.53 -0.04   -0.52    0.31 -0.57    0.66   -0.31   -0.31   -0.39    0.18   -0.41
 0.41   -0.53     1     0.06    0.76   -0.39    0.64   -0.71    0.6     0.72    0.38   -0.36    0.6
-0.06  -0.04    0.06     1     0.09    0.09    0.09   -0.1   -0.01   -0.04   -0.12    0.05   -0.05
 0.42   -0.52    0.76    0.09     1    -0.3    0.73   -0.77    0.61    0.67    0.19   -0.38    0.59
-0.22    0.31   -0.39    0.09   -0.3     1   -0.24    0.21   -0.21   -0.29   -0.36    0.13   -0.61
 0.35   -0.57    0.64    0.09    0.73   -0.24     1   -0.75    0.46    0.51    0.26   -0.27    0.6
-0.38    0.66   -0.71   -0.1   -0.77    0.21   -0.75     1   -0.49   -0.53   -0.23    0.29   -0.5
 0.63   -0.31    0.6   -0.01    0.61   -0.21    0.46   -0.49     1     0.91    0.46   -0.44    0.49
 0.58   -0.31    0.72   -0.04    0.67   -0.29    0.51   -0.53    0.91     1     0.46   -0.44    0.54
 0.29   -0.39    0.38   -0.12    0.19   -0.36    0.26   -0.23    0.46    0.46     1    -0.18    0.37
-0.39    0.18   -0.36    0.05   -0.38    0.13   -0.27    0.29   -0.44   -0.44   -0.18     1    -0.37
 0.46   -0.41    0.6   -0.05    0.59   -0.61    0.6   -0.5    0.49    0.54    0.37   -0.37     1
```

Buvo nustatyta, kokia procentinė dalis nuo visos dispersijos tenka kiekvienai komponentei

eigenvalue	proportion	cumulative	
6.12685	0.4713	0.4713	0.347indus+0.343nox+0.338tax-0.322dis+0.32 rad...
1.43328	0.11025	0.58155	0.455chas-0.349dis-0.323zn-0.315crim+0.312age...
1.24262	0.09559	0.67713	-0.594rm+0.323ptratio+0.3 b-0.296zn-0.29chas...
0.85758	0.06597	0.7431	-0.816chas-0.283ptratio+0.281rm-0.215dis+0.175age...
0.83482	0.06422	0.80732	-0.584ptratio-0.423rm+0.3951stat-0.346b+0.321zn...
0.65741	0.05057	0.85789	-0.803b-0.323zn+0.273ptratio-0.22crim-0.193tax...
0.53536	0.04118	0.89907	0.778crim-0.34indus-0.315tax-0.275zn-0.2nox...
0.3961	0.03047	0.92954	-0.601age-0.4241stat-0.403zn-0.327rm-0.318ptratio...
0.27694	0.0213	0.95084	-0.644indus+0.471rad-0.358zn-0.26crim-0.254ptratio...

Eigenvectors

Pagrindinių komponentių vizualizacija:



Išvada:

Matome, kad labiausiai yra atsiskyrę intervalai, kur būsto kainos yra > 27.5 (tūkst.)