Data mining Assignment #1

7106056054 黃筱真

(1) Install R and WEKA

(2) Do PCA and SVD on IRIS and Arrhythmia datasets and analyze the results

A. PCA on lirs dataset

In Rstudio, The built-in function **'prcomp()'** will return the results of PCA on datasets.

```
iris.pca <- prcomp(iris[c(1,2,3,4)])
```

```
> iris.pca <- prcomp(iris[c(1,2,3,4)])
> iris.pca
Standard deviations (1, .., p=4):
[1] 2.0562689 0.4926162 0.2796596 0.1543862

Rotation (n x k) = (4 x 4):
                    PC1         PC2         PC3         PC4
Sepal.Length  0.36138659 -0.65658877  0.58202985  0.3154872
Sepal.Width  -0.08452251 -0.73016143 -0.59791083 -0.3197231
Petal.Length  0.85667061  0.17337266 -0.07623608 -0.4798390
Petal.Width   0.35828920  0.07548102 -0.54583143  0.7536574
```

PCA results on Iris

Then, use **'summary()'** function in console section to see the **standard deviation, proportion of variance** and **cumulative proportion** of PCA results.

```
> iris.pca <- prcomp(iris[c(1,2,3,4)])
> summary(iris.pca)
Importance of components:
                          PC1     PC2    PC3     PC4
Standard deviation     2.0563 0.49262 0.2797 0.15439
Proportion of Variance 0.9246 0.05307 0.0171 0.00521
Cumulative Proportion  0.9246 0.97769 0.9948 1.00000
```

Summary of PCA results on Iris

If we do normalization,

```
iris.pca <- prcomp(iris[c(1,2,3,4)],scale=TRUE)
```

```
> iris.pca <- prcomp(iris[c(1,2,3,4)],scale=TRUE)
> iris.pca
Standard deviations (1, .., p=4):
[1] 1.7083611 0.9560494 0.3830886 0.1439265

Rotation (n x k) = (4 x 4):
                    PC1         PC2        PC3        PC4
Sepal.Length  0.5210659 -0.37741762  0.7195664  0.2612863
Sepal.Width  -0.2693474 -0.92329566 -0.2443818 -0.1235096
Petal.Length  0.5804131 -0.02449161 -0.1421264 -0.8014492
Petal.Width   0.5648565 -0.06694199 -0.6342727  0.5235971
```

PCA results on Iris (scale=true)

```
> iris.pca <- prcomp(iris[c(1,2,3,4)],scale=TRUE)
> summary(iris.pca)
Importance of components:
                          PC1    PC2     PC3     PC4
Standard deviation     1.7084 0.9560 0.38309 0.14393
Proportion of Variance 0.7296 0.2285 0.03669 0.00518
Cumulative Proportion  0.7296 0.9581 0.99482 1.00000
```

Summary of PCA results on Iris (scale=true)

B. SVD on lirs dataset

The built-in function **'svd()'** will return the results of SVD on datasets.

```
iris.svd <- svd(iris[c(1,2,3,4)])
```

```
> iris.svd <- svd(iris[c(1,2,3,4)])
> iris.svd
$d
[1] 95.959914 17.761034  3.460931  1.884826
```

the matrix d of SVD on Iris

```
$u
              [,1]          [,2]          [,3]          [,4]
 [1,] -0.06161685  1.296114e-01  0.0021385967  0.0016381914
 [2,] -0.05807094  1.110198e-01  0.0706723871  0.0517569646
 [3,] -0.05676305  1.179665e-01  0.0043425491  0.0095570243
 [4,] -0.05665344  1.053081e-01  0.0059246720 -0.0416438911
 [5,] -0.06123020  1.310898e-01 -0.0318810953 -0.0322148124
 [6,] -0.06750317  1.308848e-01 -0.0685371918 -0.0113642477
 [7,] -0.05748208  1.165982e-01 -0.0664136685 -0.0267433923
 [8,] -0.06097263  1.209431e-01  0.0054302657 -0.0240566566
 [9,] -0.05376120  9.994149e-02  0.0176366479 -0.0165153852
[10,] -0.05882666  1.120431e-01  0.0649689136 -0.0304719804
[11,] -0.06529182  1.365781e-01  0.0049358593 -0.0064694222
[12,] -0.05994178  1.137531e-01 -0.0252977575 -0.0836045085
[13,] -0.05711323  1.113548e-01  0.0716796746 -0.0051546938
[14,] -0.05159575  1.153248e-01  0.0042621200 -0.0137409532
[15,] -0.06800719  1.641831e-01  0.0095700911  0.0876360142
[16,] -0.07076259  1.590565e-01 -0.1191426516  0.0065500061
[17,] -0.06536474  1.468448e-01 -0.0616990077  0.0906600216
[18,] -0.06179182  1.276765e-01 -0.0133779293  0.0415289575
[19,] -0.06928030  1.345414e-01  0.0100175034  0.0166397745
[20,] -0.06351469  1.329215e-01 -0.0736188358 -0.0344734444
[21,] -0.06517277  1.193631e-01  0.0600481277 -0.0069852222
[22,] -0.06329358  1.279082e-01 -0.0696249084  0.0222494332
[23,] -0.05596085  1.406498e-01 -0.0830798653  0.0017258878
[24,] -0.06295342  1.056799e-01 -0.0105187125  0.0784565109
[25,] -0.06154560  1.017832e-01 -0.0304263956 -0.1601227105
[26,] -0.05992289  1.046398e-01  0.0817625336  0.0177657222
[27,] -0.06185719  1.130832e-01 -0.0273123325  0.0302188083
```

the matrix u of SVD on Iris

```
$v
             [,1]       [,2]       [,3]       [,4]
[1,] -0.7511082  0.2841749  0.50215472  0.3208143
[2,] -0.3800862  0.5467445 -0.67524332 -0.3172561
[3,] -0.5130089 -0.7086646 -0.05916621 -0.4807451
[4,] -0.1679075 -0.3436708 -0.53701625  0.7518717
```

the matrix v of SVD on Iris

C. PCA on arrhythmia dataset

```
> arrhythmia.pca
Standard deviations (1, .., p=68):
 [1] 1.471528e+02 1.282411e+02 9.970231e+01 8.490835e+01 7.873402e+01 7.201073e+01 5.872948e+01 5.218652e+01 4.635776e+01
[10] 4.314596e+01 3.945547e+01 3.609289e+01 3.500400e+01 3.389285e+01 3.240807e+01 3.004417e+01 2.598760e+01 2.519893e+01
[19] 2.480817e+01 2.277841e+01 2.229312e+01 2.169787e+01 2.093748e+01 1.876727e+01 1.842120e+01 1.788489e+01 1.716605e+01
[28] 1.697709e+01 1.639463e+01 1.560430e+01 1.468192e+01 1.441668e+01 1.262054e+01 1.212960e+01 1.195960e+01 1.128541e+01
[37] 1.079889e+01 9.985788e+00 9.299583e+00 9.118698e+00 8.851099e+00 8.454577e+00 8.104618e+00 7.896199e+00 7.366856e+00
[46] 7.182697e+00 6.931697e+00 6.578783e+00 6.279927e+00 5.966644e+00 5.634127e+00 5.305312e+00 4.689696e+00 4.545383e+00
[55] 4.322969e+00 4.180801e+00 4.020017e+00 3.886463e+00 3.350402e+00 3.269039e+00 3.138770e+00 3.006075e+00 2.913164e+00
[64] 2.383147e+00 2.109089e+00 1.981599e+00 1.367037e+00 6.534898e-14

Rotation (n x k) = (280 x 68):
              PC1           PC2           PC3           PC4           PC5           PC6           PC7           PC8
V1    1.538853e-03 -5.219692e-02 -8.353067e-03 -1.541480e-03  4.920649e-02  4.098411e-03  2.036181e-02  1.186757e-01
V2   -6.241153e-04 -1.500595e-06  3.719178e-04  1.168734e-03  9.137799e-04  3.920392e-04  1.282337e-03 -2.008637e-03
V3    9.299799e-03 -7.866285e-04 -1.317186e-02 -2.169651e-03  4.522849e-02 -3.023800e-02  2.952142e-02  9.118007e-02
V4    2.115914e-02 -1.232383e-02 -2.333433e-03  2.166107e-02  4.740612e-02 -4.366758e-02  1.120070e-02  1.163092e-01
V5    2.560012e-02 -1.179086e-01 -1.849162e-02 -2.115833e-03 -1.404997e-01  1.667385e-01 -6.669676e-02 -1.628850e-01
V6    2.705888e-02  5.052927e-03  4.189241e-02 -4.872923e-02 -7.200612e-02  3.939547e-02  2.734266e-03 -5.765602e-02
V7   -4.337082e-02 -1.837945e-01 -1.800180e-01  6.923541e-03 -1.117548e-01  2.181183e-01 -4.226067e-02  1.061375e-01
V8    4.193528e-02 -1.908396e-01 -1.970244e-02  8.465143e-02 -9.244920e-02  3.115317e-01 -1.325442e-01  2.324835e-01
V9    4.806555e-03  4.464560e-03  2.337101e-02 -3.597021e-02 -3.975194e-02  3.314845e-02  2.672768e-02  6.740882e-02
V10  -2.091755e-01  1.050994e-01 -1.281912e-01  2.665697e-01 -2.312879e-01  8.726344e-02 -2.037786e-01  4.302290e-01
V11   2.949408e-01  6.348484e-02 -8.067730e-01 -3.885521e-01  5.093473e-02 -1.062461e-02 -3.597546e-03  1.374329e-01
V12  -8.123556e-02  5.896672e-02  9.472922e-03 -6.205575e-02 -1.909748e-01 -4.603545e-02  1.848854e-01 -1.239888e-01
V13  -8.215492e-02  1.537663e-01 -2.162682e-01  1.772107e-02 -3.483461e-01 -1.763062e-01 -1.131606e-01 -3.703162e-01
V14   8.170239e-01  8.452789e-02  8.280548e-02  4.771255e-01 -1.874837e-01 -1.229520e-01 -2.820361e-02  6.488645e-02
              PC9          PC10          PC11          PC12          PC13          PC14          PC15          PC16
V1    8.306486e-02 -8.320554e-02 -5.395468e-03 -8.690104e-03  4.082351e-02  7.502217e-02  2.922141e-02  8.152852e-02
V2   -1.629725e-03 -1.515801e-03  3.050703e-03 -1.195936e-03 -2.177170e-03  1.020333e-03  8.465702e-04 -1.768114e-03
```

PCA results on arrhythmia

```
> summary(arrhythmia.pca)
Importance of components:
                           PC1      PC2     PC3      PC4      PC5      PC6      PC7      PC8      PC9     PC10     PC11     PC12
Standard deviation     147.153 128.2411 99.7023 84.90835 78.73402 72.01073 58.72948 52.18652 46.35776 43.14596 39.4555 36.09289
Proportion of Variance   0.235   0.1785  0.1079  0.07825  0.06728  0.05628  0.03744  0.02956  0.02333  0.02021  0.0169  0.01414
Cumulative Proportion    0.235   0.4135  0.5214  0.59968  0.66696  0.72324  0.76068  0.79024  0.81356  0.83377  0.8507  0.86481
                          PC13    PC14    PC15    PC16    PC17    PC18    PC19    PC20    PC21    PC22    PC23    PC24
Standard deviation     35.0040 33.89285 32.4081 30.0442 25.98760 25.19893 24.80817 22.77841 22.29312 21.69787 20.93748 18.76727
Proportion of Variance  0.0133  0.01247  0.0114  0.0098  0.00733  0.00689  0.00668  0.00563  0.00539  0.00511  0.00476  0.00382
Cumulative Proportion   0.8781  0.89057  0.9020  0.9118  0.91910  0.92599  0.93267  0.93830  0.94370  0.94881  0.95357  0.95739
                          PC25    PC26    PC27    PC28    PC29    PC30    PC31    PC32    PC33    PC34    PC35
Standard deviation     18.42120 17.88489 17.1661 16.97709 16.39463 15.60430 14.68192 14.41668 12.62054 12.1296 11.95960
Proportion of Variance  0.00368  0.00347  0.0032  0.00313  0.00292  0.00264  0.00234  0.00226  0.00173  0.0016  0.00155
Cumulative Proportion   0.96107  0.96454  0.9677  0.97087  0.97379  0.97643  0.97877  0.98103  0.98276  0.9843  0.98591
                          PC36    PC37    PC38    PC39    PC40    PC41    PC42    PC43    PC44    PC45    PC46    PC47    PC48
Standard deviation     11.28541 10.79889 9.98579 9.29958 9.1187 8.85110 8.45458 8.10462 7.89620 7.36686 7.18270 6.93170 6.57878
Proportion of Variance  0.00138  0.00127 0.00108 0.00094 0.0009 0.00085 0.00078 0.00071 0.00068 0.00059 0.00056 0.00052 0.00047
Cumulative Proportion   0.98729  0.98855 0.98964 0.99057 0.9915 0.99233 0.99310 0.99382 0.99449 0.99508 0.99564 0.99616 0.99663
                          PC49    PC50    PC51    PC52    PC53    PC54    PC55    PC56    PC57    PC58    PC59    PC60    PC61
Standard deviation     6.27993 5.96664 5.63413 5.30531 4.68970 4.54538 4.3230 4.18080 4.02002 3.88646 3.35040 3.26904 3.13877
Proportion of Variance 0.00043 0.00039 0.00034 0.00031 0.00024 0.00022 0.0002 0.00019 0.00018 0.00016 0.00012 0.00012 0.00011
Cumulative Proportion  0.99706 0.99745 0.99779 0.99810 0.99834 0.99856 0.9988 0.99895 0.99913 0.99929 0.99941 0.99953 0.99964
                          PC62    PC63    PC64    PC65    PC66    PC67      PC68
Standard deviation     3.0061 2.91316 2.38315 2.10909 1.98160 1.36704 6.535e-14
Proportion of Variance 0.0001 0.00009 0.00006 0.00005 0.00004 0.00002 0.000e+00
Cumulative Proportion  0.9997 0.99983 0.99989 0.99994 0.99998 1.00000 1.000e+00
```

Summary of PCA results on arrhythmia

D. SVD on arrhythmia dataset

```
> svd(na.omit(arrhythmia))
$d
  [1] 4771.74480 1204.27120  994.86098  786.02486  690.95436  634.16275  524.50413  468.40440  426.90656  379.44235  351.48031  303.95685
 [13]  294.94339  282.37742  268.07647  248.85528  240.59615  211.00869  206.25059  195.35658  186.29466  178.15081  173.67731  161.41221
 [25]  152.13327  146.52996  142.24446  139.11385  135.38737  133.93886  121.32500  119.29623  116.70596  101.35273   99.25971   93.12149
 [37]   90.62221   86.02328   80.29151   76.11071   73.97805   69.90980   69.11655   66.30427   64.53317   60.29214   57.56583   56.46181
 [49]   53.49175   49.20884   48.41916   45.08332   40.27414   38.30656   35.75526   35.14268   33.99454   32.72100   30.59316   27.26931
 [61]   26.69514   24.98220   24.31928   23.73997   19.12958   17.22345   15.17178   10.79870
```

the matrix d of SVD on Arrhythmia

```
$u
             [,1]         [,2]         [,3]         [,4]         [,5]         [,6]         [,7]         [,8]         [,9]         [,10]
 [1,] -0.12890972 -0.009662942  0.166247482 -0.0733963064  8.067838e-03 -0.1868728033  0.049584363 -0.047688128  0.046408038 -0.2016151504
 [2,] -0.10578356  0.039408370  0.078534392  0.0452908780 -8.923539e-02 -0.1113108058 -0.039815545  0.067756146  0.027050375 -0.0109253531
 [3,] -0.12573672  0.217469694  0.009671981  0.0256192651  7.670511e-02  0.1541938598 -0.049201330 -0.083423444  0.266318322 -0.0145663785
 [4,] -0.10488660 -0.088110674 -0.035981085  0.0658868041 -6.332956e-02  0.0442073481  0.004422862 -0.003199581 -0.041140275 -0.1419282365
 [5,] -0.12535147  0.080896244  0.070027723 -0.0561156859 -1.909919e-01 -0.0275852884 -0.122406856  0.102321124 -0.007606896  0.1121116930
 [6,] -0.15117433 -0.162362088 -0.372642013 -0.0817793255 -9.178189e-03 -0.3411283046  0.218650356  0.064964164 -0.054800144 -0.1013416011
 [7,] -0.11993184 -0.084999224  0.127386589  0.1903276095  5.848833e-02 -0.0216579894  0.031833227 -0.053784460  0.002227412  0.0447768462
 [8,] -0.14471714 -0.095762564 -0.283884093 -0.2058115009  5.135356e-02 -0.1046721729  0.218581790  0.017968092  0.297015693  0.1733747415
 [9,] -0.14206242  0.146644478 -0.100689444  0.1429806478 -1.746289e-01 -0.0612185569  0.097065369 -0.050416868  0.025187984  0.0259919065
[10,] -0.10336096 -0.041056050  0.025572965 -0.0721424775  8.003459e-02  0.1556499440 -0.070225967  0.078479299 -0.036387617 -0.0648717342
[11,] -0.08922593  0.058786428  0.002622926 -0.2298616001 -2.844943e-01 -0.0176971357 -0.078608186 -0.018055593 -0.052938127  0.0182541497
[12,] -0.11372684  0.042940721  0.088834399  0.0051554268 -6.374534e-02 -0.1067678583 -0.073475474  0.005519190  0.056626713 -0.0023601205
[13,] -0.11518576 -0.121884318  0.030566492  0.0809882730  2.317488e-02  0.1314407180 -0.008619704  0.147496469 -0.066013159 -0.0365017454
[14,] -0.11598475  0.011763366  0.056090787 -0.0057863603 -4.037336e-02 -0.0830343565 -0.051298631  0.050004130 -0.049860850 -0.1018573670
             [,11]        [,12]        [,13]        [,14]        [,15]        [,16]        [,17]        [,18]        [,19]        [,20]
 [1,]  0.019981415  0.287727757 -0.1697712302  0.046374090 -0.0353569952  0.230837833 -0.399701705  0.083594566 -0.1820979019  0.083587720
 [2,]  0.059132615 -0.044257784  0.0743108323  0.174801605  0.0509165523 -0.046927854  0.108921523  0.001373217  0.0654897563  0.140433386
 [3,]  0.174411441 -0.148350049  0.0798422480  0.065670158 -0.2626345299 -0.048269003 -0.065992202 -0.090490340  0.0474176936 -0.049165742
 [4,]  0.228916941 -0.116438165 -0.0980586736  0.227467657  0.0744261790  0.328011411  0.142729327 -0.003838676  0.0959743087  0.113878220
 [5,] -0.208782257  0.161752395  0.2666390374 -0.057127474 -0.0603294055 -0.141470123 -0.097309222  0.238269465 -0.0194723394 -0.117395966
 [6,] -0.110059705 -0.148232575  0.0374727017 -0.134946111 -0.1118623112 -0.079956478  0.289941665 -0.054609633 -0.1559967773  0.261676098
 [7,] -0.061249529  0.153951403 -0.1047324021  0.083894677  0.1059391774 -0.120109276  0.126719904 -0.114755171  0.0334568849 -0.096392723
 [8,] -0.060558314  0.381845605  0.1823352731  0.140079853 -0.0580270493  0.038295377  0.027497391 -0.142552750  0.1380338681  0.065728905
```

the matrix u of SVD on Arrhythmia

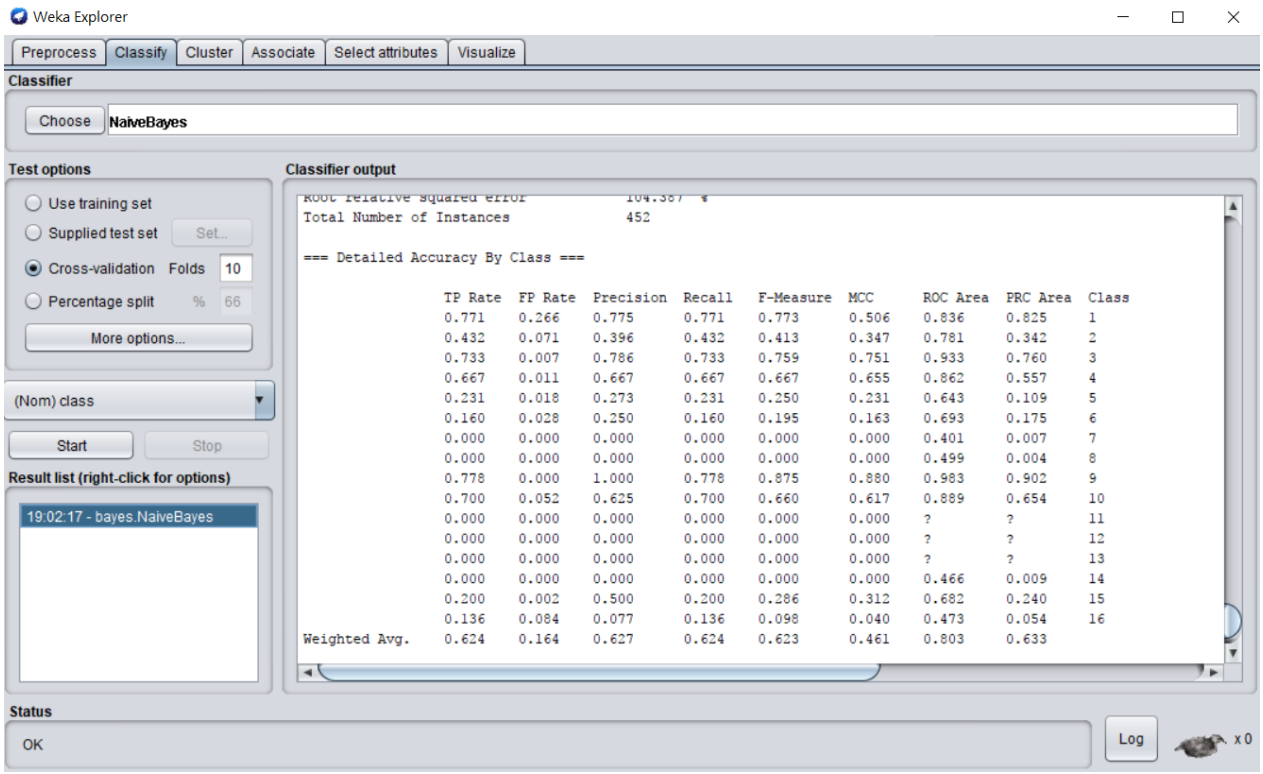
```
$v
              [,1]          [,2]          [,3]          [,4]          [,5]          [,6]          [,7]          [,8]          [,9]
  [1,] -8.137468e-02  3.758075e-03 -2.982406e-02  2.869328e-03 -7.794319e-04  5.451491e-02 -3.844329e-02  5.376634e-02  1.137331e-01
  [2,] -7.998305e-04 -6.139567e-04  1.220853e-04 -5.513758e-04 -1.178383e-03  9.230287e-04  6.150409e-04  9.682620e-04 -1.997259e-03
  [3,] -2.784507e-01  1.372534e-02  7.958793e-02 -5.075493e-02  1.858132e-02  7.009498e-02 -1.796847e-01  1.559359e-01  7.254097e-02
  [4,] -1.232761e-01  2.333115e-02  2.289510e-02 -2.280045e-02 -1.922585e-02  4.735628e-02 -1.336111e-01  1.033677e-01  1.025339e-01
  [5,] -1.802521e-01  3.114185e-02 -4.841718e-02 -4.088504e-02  4.623827e-02 -6.227319e-02  1.263453e-01 -1.152644e-01 -1.565847e-01
  [6,] -2.723762e-01  3.158941e-02  9.049219e-02 -1.337543e-01  9.816982e-02  3.093963e-03 -1.665277e-01  1.578103e-01 -8.531831e-02
  [7,] -6.416051e-01 -2.841627e-02  5.132644e-02  1.181035e-03  6.723507e-02  3.058516e-02 -5.632803e-02  8.243983e-02  8.434135e-02
  [8,] -3.419274e-01  5.176415e-02 -6.280414e-02 -9.405189e-02 -1.802214e-02  4.401183e-02  1.438914e-01 -1.494739e-01  2.359702e-01
  [9,] -1.600381e-01  7.427373e-03  5.410540e-02 -7.646159e-02  6.480184e-02  6.980655e-03 -7.855012e-02  1.124842e-01  5.157921e-02
 [10,] -3.681732e-02 -2.092220e-01  1.603264e-01  2.711516e-02 -2.228022e-01 -2.108066e-01  1.160757e-01 -2.697492e-01  4.481303e-01
 [11,] -7.394236e-02  2.974316e-01  2.083155e-01  8.135849e-01  3.104098e-01  2.072373e-02  5.889051e-02 -4.384330e-02  1.441275e-01
 [12,] -7.992142e-02 -8.097263e-02  8.261312e-02 -4.800738e-02  9.296841e-02 -1.765032e-01  1.759401e-02  2.034154e-01 -1.351979e-01
 [13,] -4.542954e-02 -8.323019e-02  2.096371e-01  1.510940e-01  6.694974e-03 -3.906600e-01 -6.964119e-02 -1.265388e-01 -3.631332e-01
 [14,]  1.275862e-02  8.154531e-01  1.042182e-01 -1.648424e-01 -4.407554e-01 -2.469672e-01 -3.876121e-02 -2.195075e-02  6.617605e-02
              [,10]         [,11]         [,12]         [,13]         [,14]         [,15]         [,16]         [,17]         [,18]
  [1,]  8.262431e-02 -9.312268e-02 -6.808560e-02 -1.405931e-02  2.783559e-02  4.741057e-02  7.692734e-02 -2.330435e-02 -1.610635e-02
  [2,] -1.609544e-03 -1.119148e-03 -2.812528e-03  9.502735e-04  2.903503e-03 -1.454650e-04  1.295935e-03  5.085826e-03  4.962374e-03
  [3,]  3.577160e-02 -6.134269e-02 -7.957145e-02 -5.497755e-02  4.700640e-03  3.067099e-02  1.124968e-01  9.325872e-02 -8.422889e-02
  [4,]  1.067033e-01 -5.911024e-02  6.110752e-02 -4.564808e-02 -3.034255e-02  1.091548e-01  1.949668e-01 -1.222403e-01 -9.916684e-02
  [5,]  5.460264e-02 -2.387321e-02  1.294761e-01 -9.449557e-02 -1.496663e-02 -2.067075e-01  1.454457e-01  1.213039e-02 -1.557722e-02
  [6,]  1.765182e-01 -3.377612e-01  2.113741e-01  2.290913e-01 -1.654820e-01 -7.082948e-02 -3.836808e-01 -2.261697e-01  1.369282e-01
  [7,] -2.135817e-01  1.595542e-01 -1.043554e-01  9.299769e-02  1.334594e-01  2.273806e-01  1.402914e-01  2.693200e-01  2.196044e-01
  [8,]  1.779761e-01  2.484709e-01 -3.823451e-01 -5.364459e-02  1.074638e-01 -8.442772e-02 -1.827206e-01 -4.327750e-01 -2.996080e-01
  [9,]  6.742363e-02 -2.095718e-01  1.443279e-01  4.522626e-02 -2.227170e-01  6.812382e-02 -2.656725e-01  1.894900e-03  4.065674e-02
 [10,] -1.426447e-01  2.816395e-02  1.340675e-01 -6.456901e-02  1.400361e-02  1.810957e-01 -2.580490e-01  1.146216e-01  1.781544e-01
 [11,]  3.085134e-02 -6.242913e-02  8.486200e-02  5.790330e-02 -3.873272e-02 -5.386546e-02  4.422108e-02  1.865011e-03 -6.383353e-02
```

the matrix v of SVD on Arrhythmia

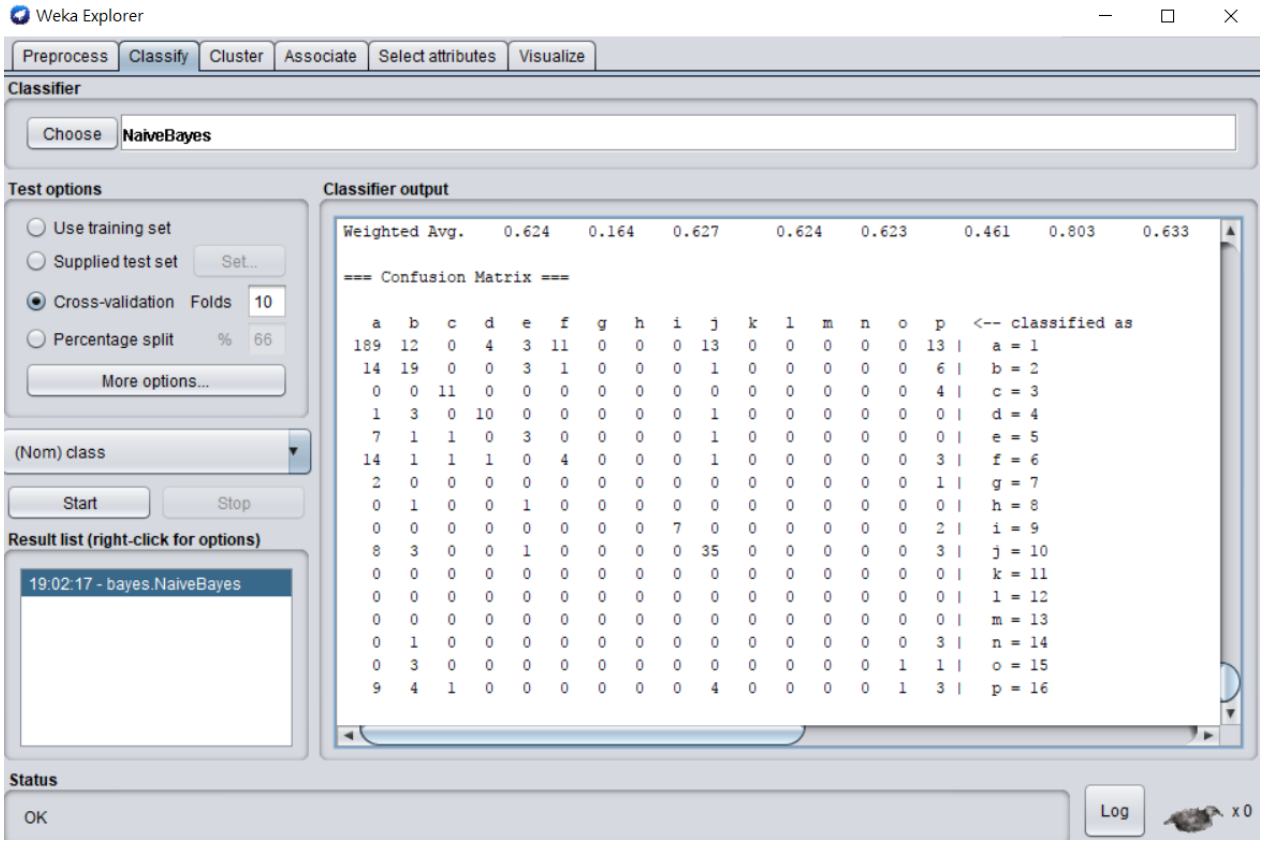
(3) Show how the results of PCA and SVD can be further used for classification

When dimensions is a lot for classifier, we can use PCA and SVD to reduce these features to a manageable size, while maintaining most of the information in the dataset.
For example, we have a dataset composed by a set of properties from cars. These properties describe each car by its size, color, circularity, compactness, radius, number of seats, number of doors, size of trunk and so on. However, many of these features will measure related properties and so will be redundant. Therefore, we should remove these redundancy and describe each car with less properties.

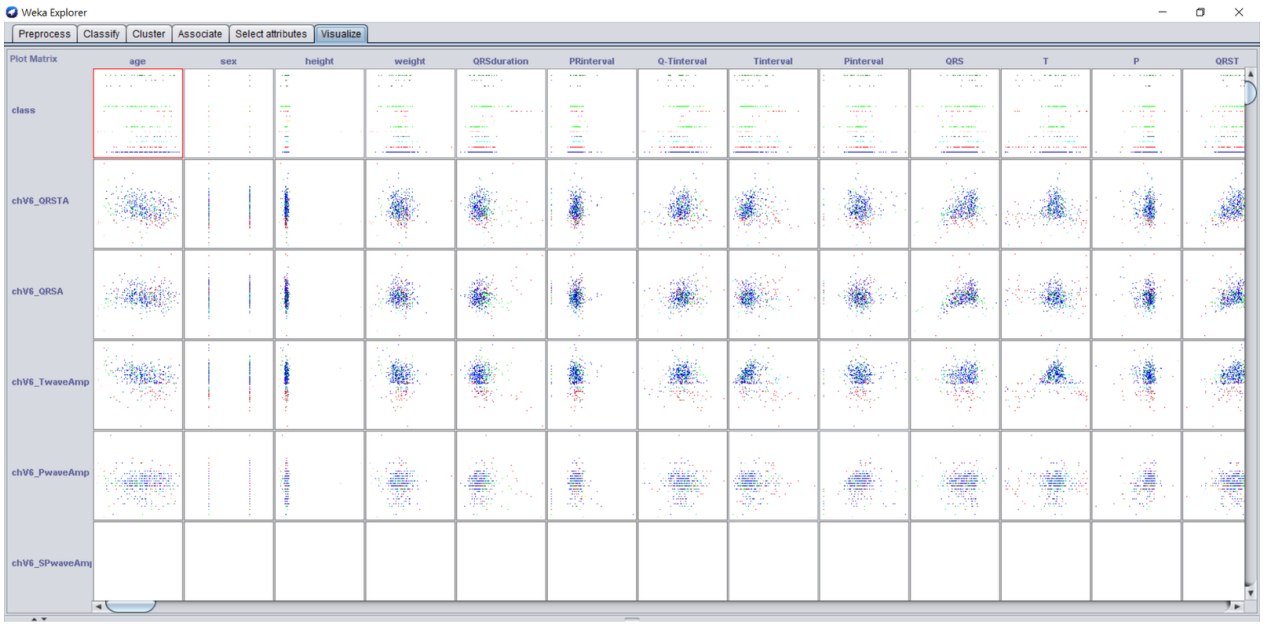
(4) Use WEKA for classification on Arrhythmia dataset and report the results



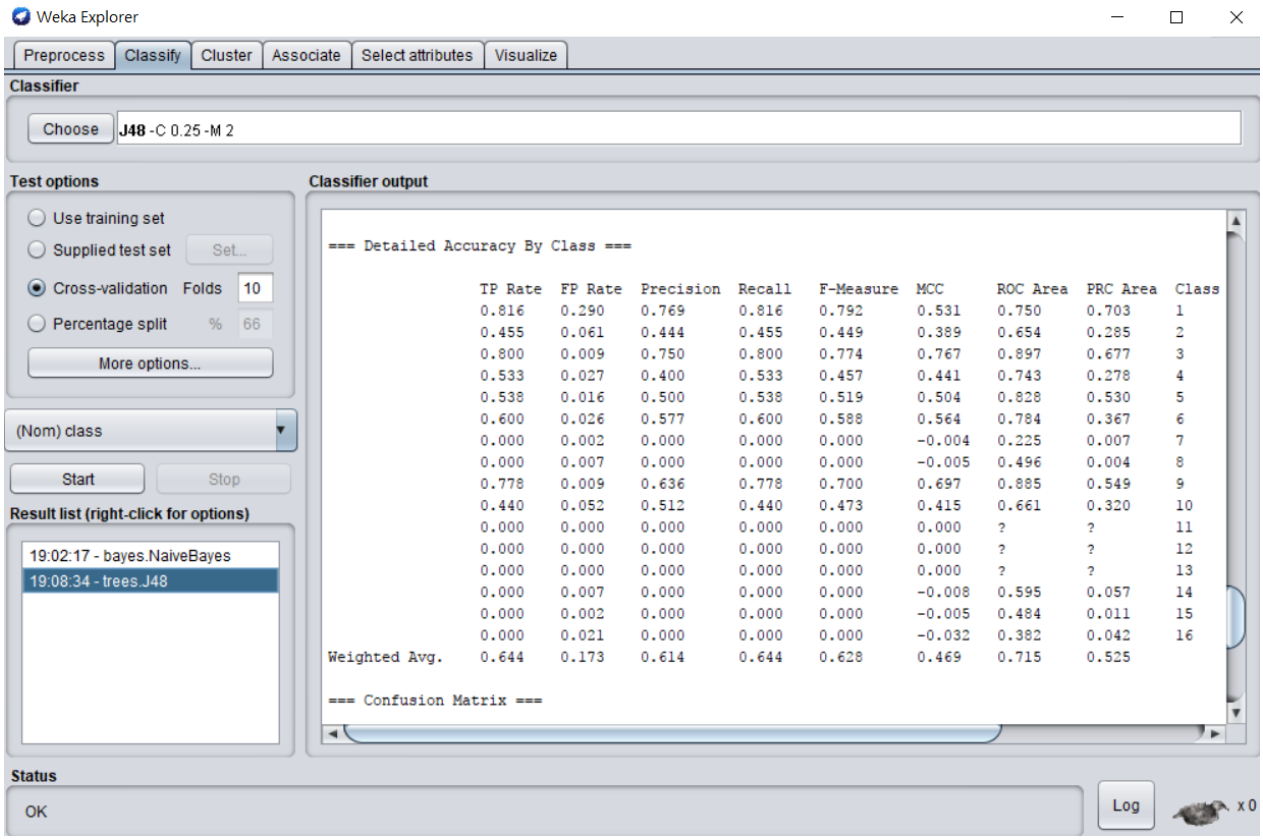
Accuracy of NaiveBayes on Arrhythmia dataset



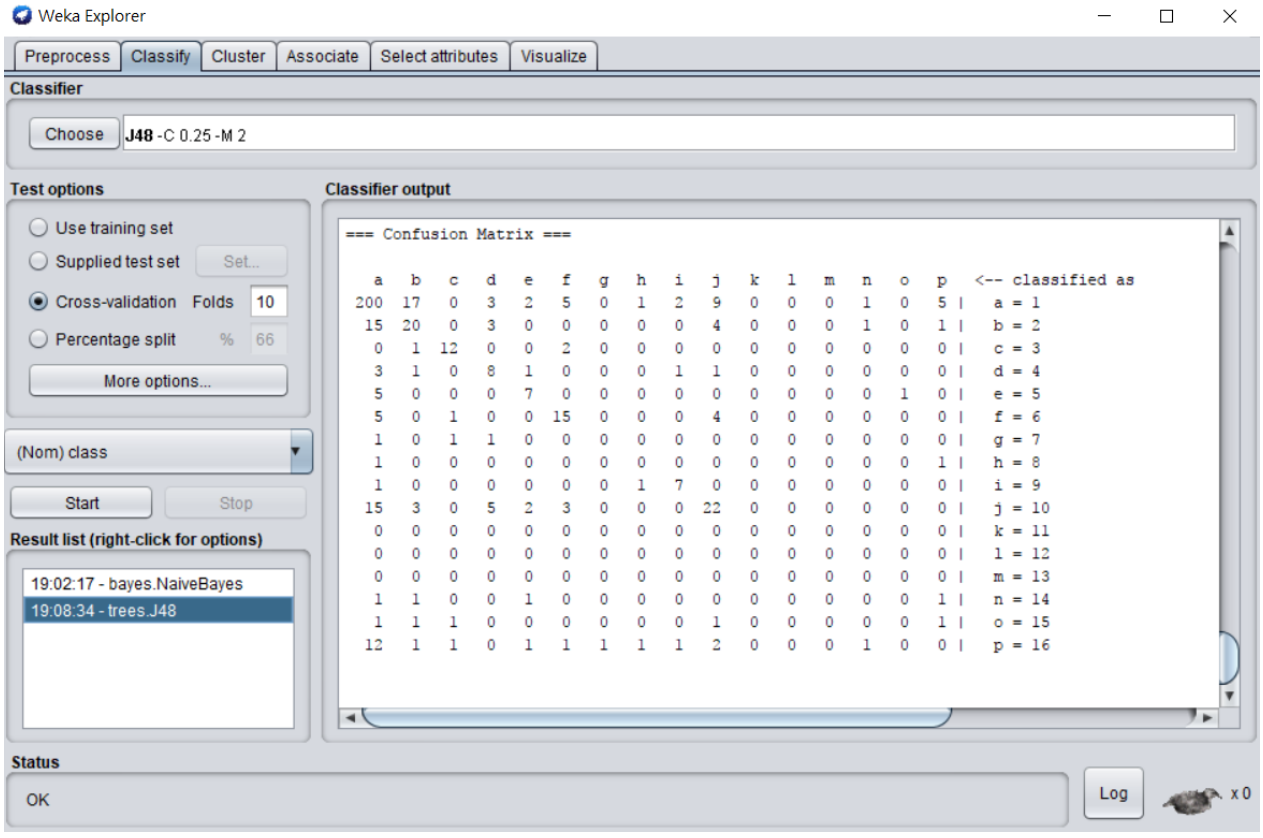
Confusion Matrix of NaiveBayes on Arrhythmia dataset



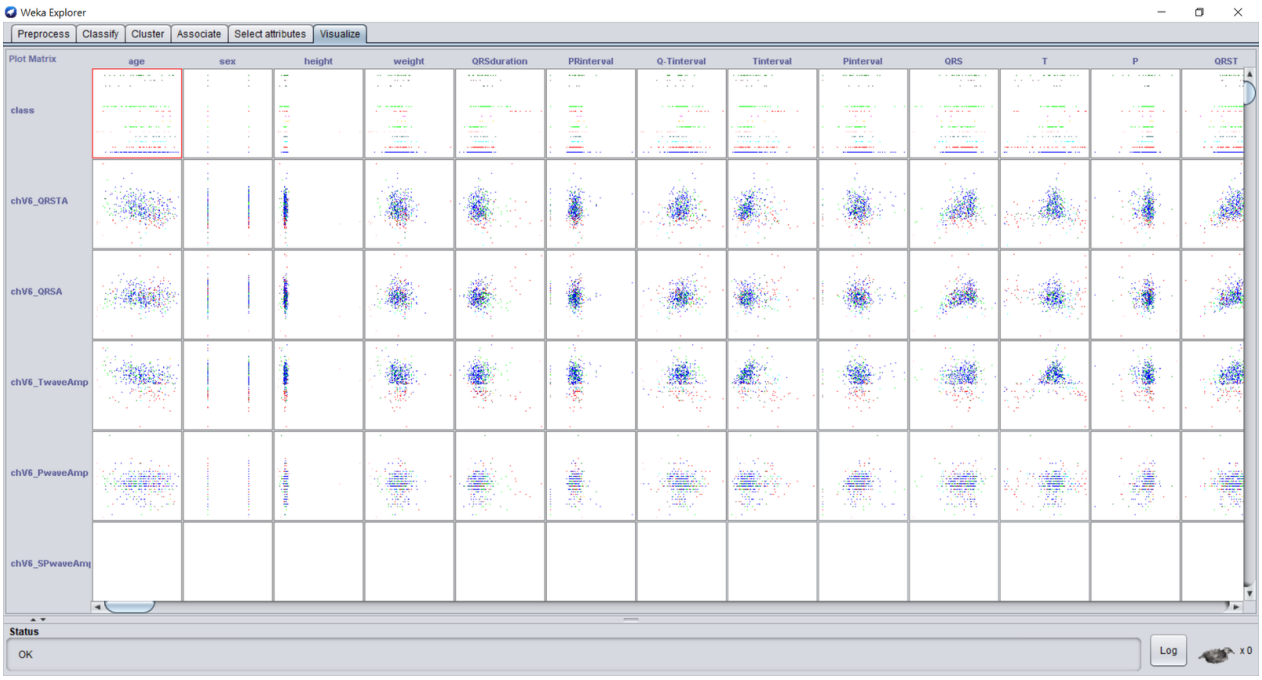
Visualization of NaiveBayes on Arrhythmia dataset



Accuracy of J48 on Arrhythmia dataset



Confusion Matrix of J48 on Arrhythmia dataset



Visualization of J48 on Arrhythmia dataset

(5) Do a literature review on research papers that use Arrhythmia dataset for classification and report the best results on accuracy

I refer to this research paper:
Kemal Polat *, Salih Gu¨nes, Detection of ECG Arrhythmia using a differential expert system approach based on principal component analysis and least square support vector machine, Selcuk University, Electrical and Electronics Engineering Department, 42035 Konya, Turkey, 2007

Below is their experiment results table:

Table 1
The experimental results obtained by PCA-LSSVM classifier for diagnosis of ECG arrhythmia

Number of dataset	Datasets	Measures	PCA-LSSVM
Number of normal ECG	50–50% of training-test partition	Sensitivity (%)	100
Arrhythmia: **245**		Specificity (%)	100
Training: **123**		TP rate (%)	100
Test: **122**			
Number of diseased ECG		FP rate (%)	0
Arrhythmia: **207**		Accuracy (%)	100
Training: **103**		*F*-measure (%)	100
Test: **102**			
Number of normal ECG	70–30% of training-test partition	Sensitivity (%)	100
Arrhythmia: **245**		Specificity (%)	100
Training: **172**		TP rate (%)	100
Test: **73**			
Number of diseased ECG		FP rate (%)	0
Arrhythmia: **207**		Accuracy (%)	100
Training: **145**		*F*-measure (%)	100
Test: **62**			
Number of normal ECG	80–20% of training-test partition	Sensitivity (%)	100
Arrhythmia: **245**		Specificity (%)	100
Training: **196**		TP rate (%)	100
Test: **49**			
Number of diseased ECG		FP rate (%)	0
Arrhythmia: **207**		Accuracy (%)	100
Training: **166**		*F*-measure (%)	100
Test: **41**			