Yuchen Duan (yuchend3)

IE598 MLF F18

Module 5 Homework (Dimensionality Reduction)

**Part 1: Exploratory Data Analysis**

Alcohol Malic acid Ash Alcalinity of ash Magnesium \

count 178.000000 178.000000 178.000000 178.000000 178.000000

mean 13.000618 2.336348 2.366517 19.494944 99.741573

std 0.811827 1.117146 0.274344 3.339564 14.282484

min 11.030000 0.740000 1.360000 10.600000 70.000000

25% 12.362500 1.602500 2.210000 17.200000 88.000000

50% 13.050000 1.865000 2.360000 19.500000 98.000000

75% 13.677500 3.082500 2.557500 21.500000 107.000000

max 14.830000 5.800000 3.230000 30.000000 162.000000

Total phenols Flavanoids Nonflavanoid phenols Proanthocyanins \

count 178.000000 178.000000 178.000000 178.000000

mean 2.295112 2.029270 0.361854 1.590899

std 0.625851 0.998859 0.124453 0.572359

min 0.980000 0.340000 0.130000 0.410000

25% 1.742500 1.205000 0.270000 1.250000

50% 2.355000 2.135000 0.340000 1.555000

75% 2.800000 2.875000 0.437500 1.950000

max 3.880000 5.080000 0.660000 3.580000

Color intensity Hue OD280/OD315 of diluted wines Proline \

count 178.000000 178.000000 178.000000 178.000000

mean 5.058090 0.957449 2.611685 746.893258

std 2.318286 0.228572 0.709990 314.907474

min 1.280000 0.480000 1.270000 278.000000

25% 3.220000 0.782500 1.937500 500.500000

50% 4.690000 0.965000 2.780000 673.500000

75% 6.200000 1.120000 3.170000 985.000000

max 13.000000 1.710000 4.000000 1680.000000

Class

count 178.000000

mean 1.938202

std 0.775035

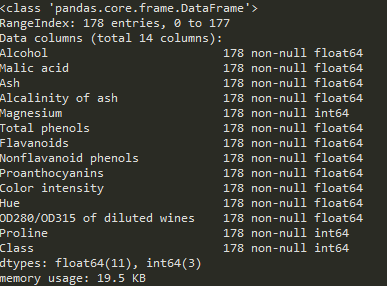
min 1.000000

25% 1.000000

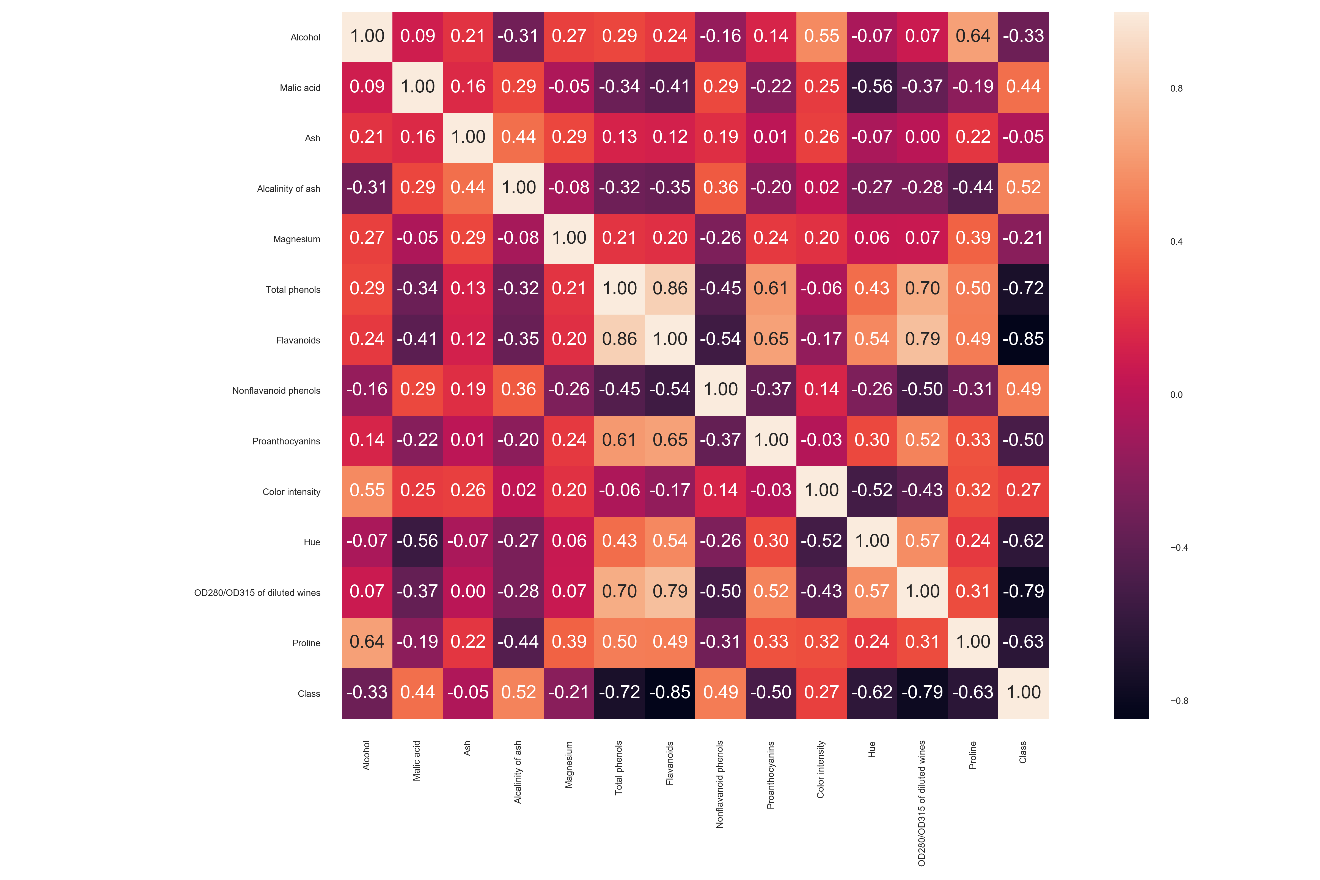
50% 2.000000

75% 3.000000

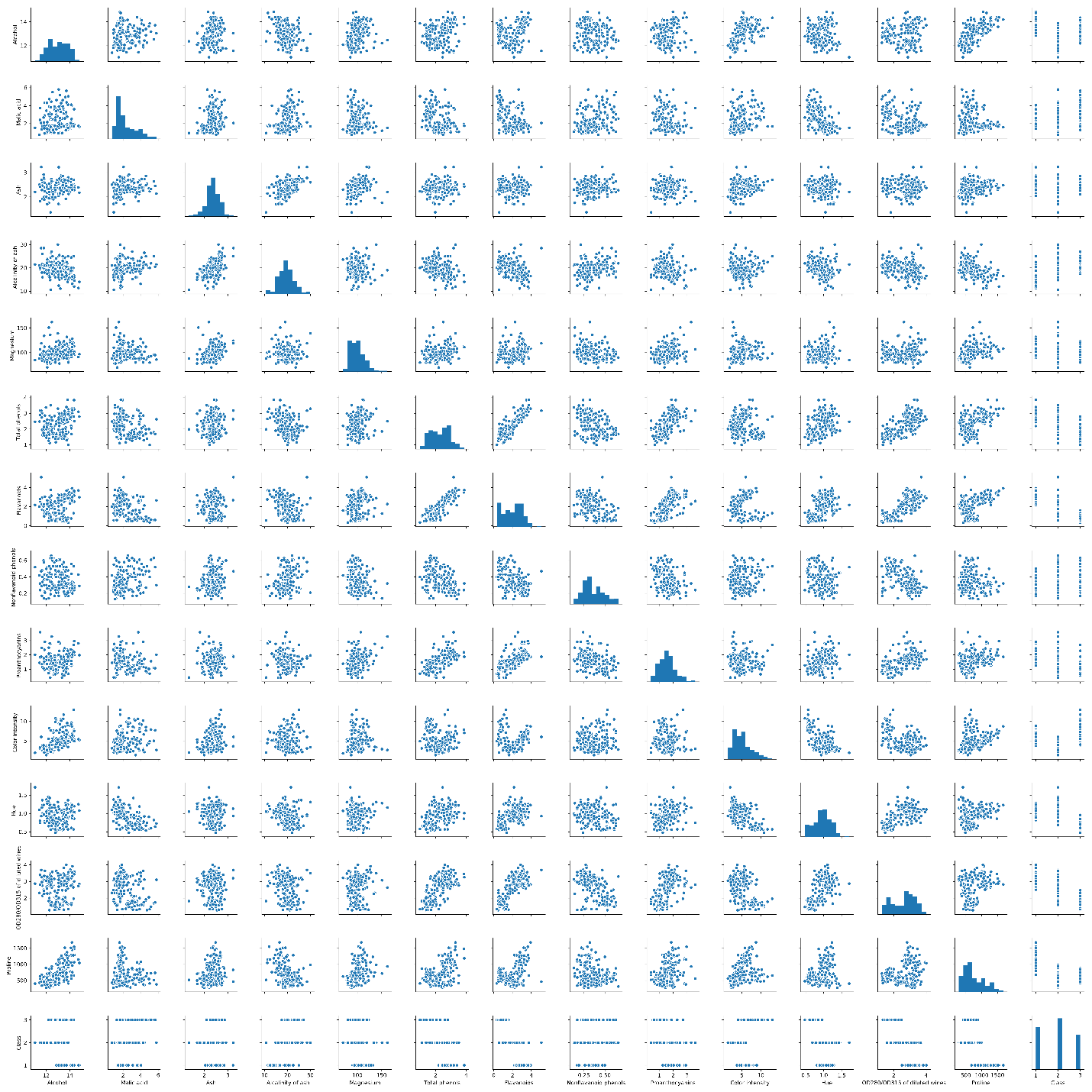
max 3.000000



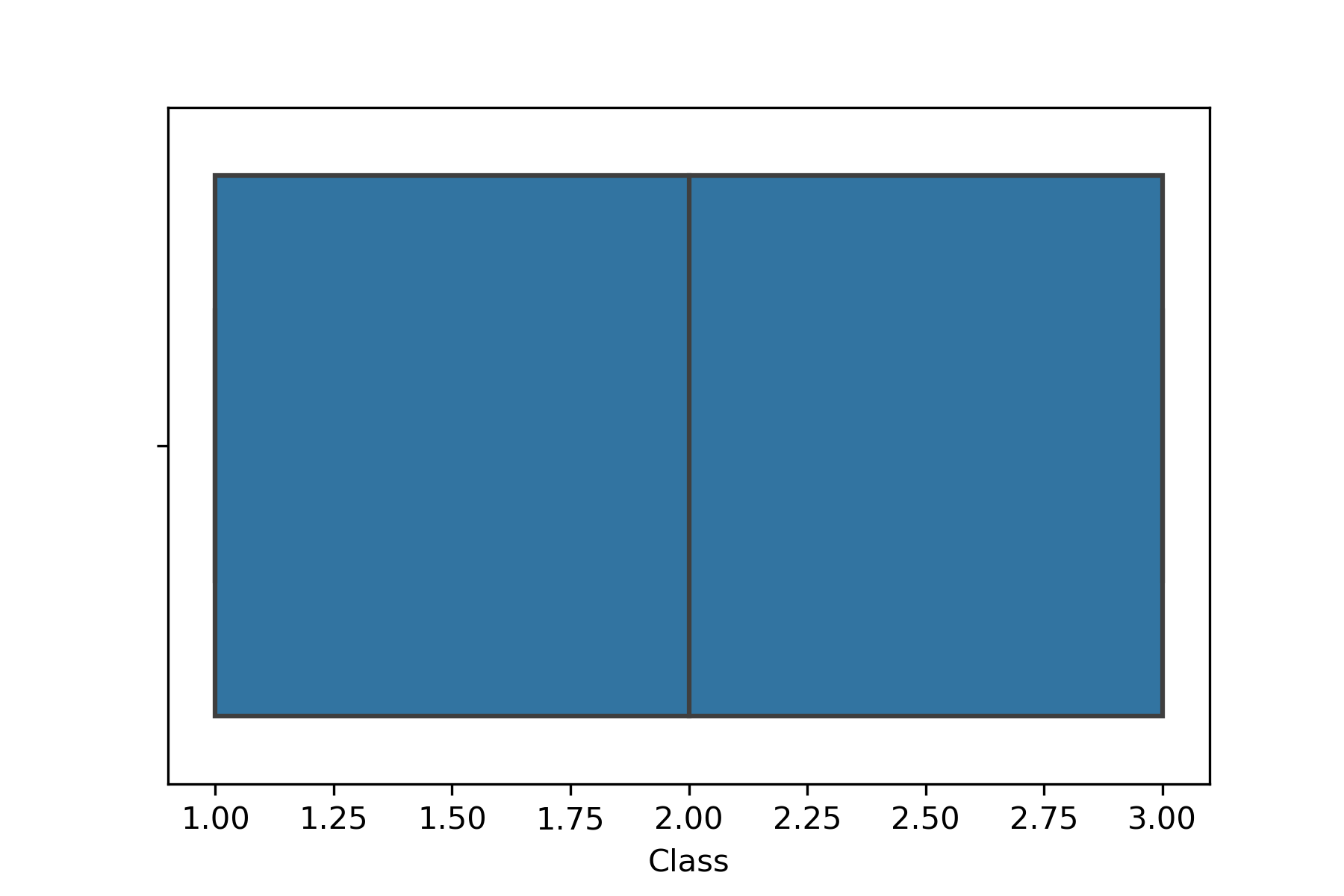
178 Entries; 14 Columns

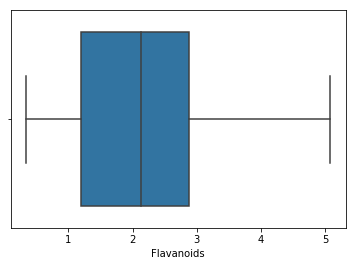


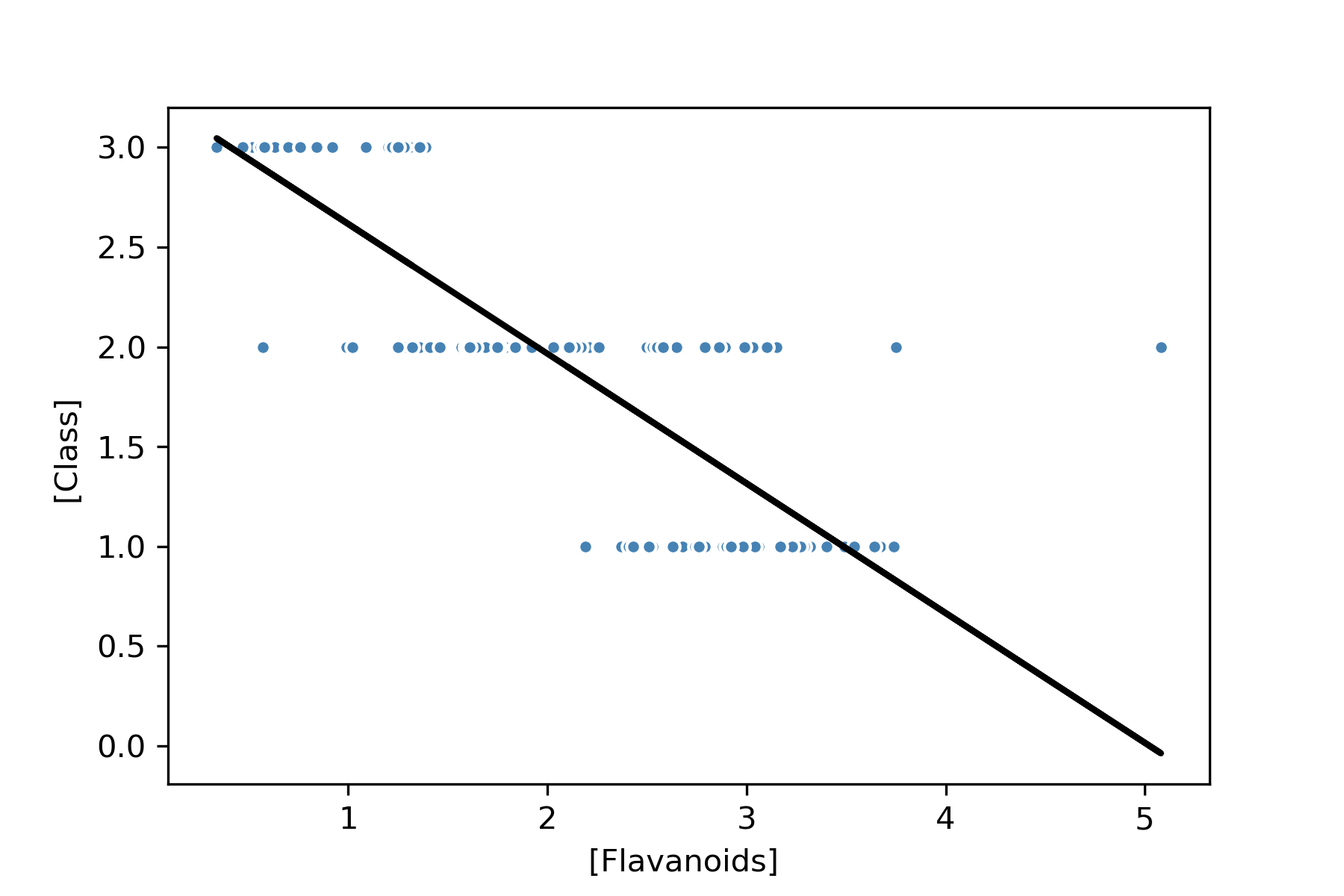
Flavonoids and Class have the strongest relation.



Show all scatter

****



****

Regression between Flavonoids and Class

**Part 2: Logistic regression classifier v. SVM classifier - baseline**

lr train R^2: 1.000

lr test R^2: 1.000

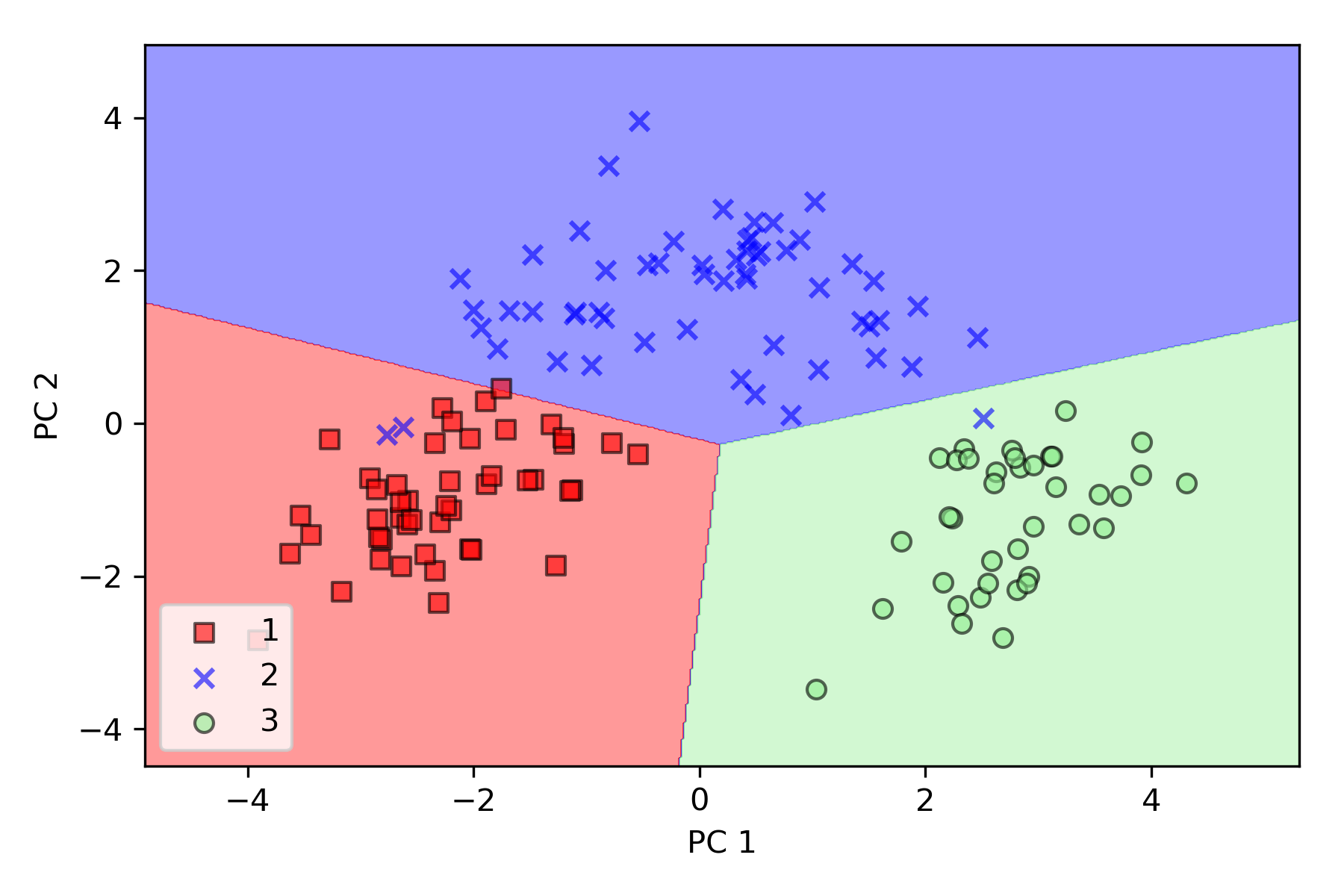
sv train R^2: 0.993

sv test R^2: 0.972

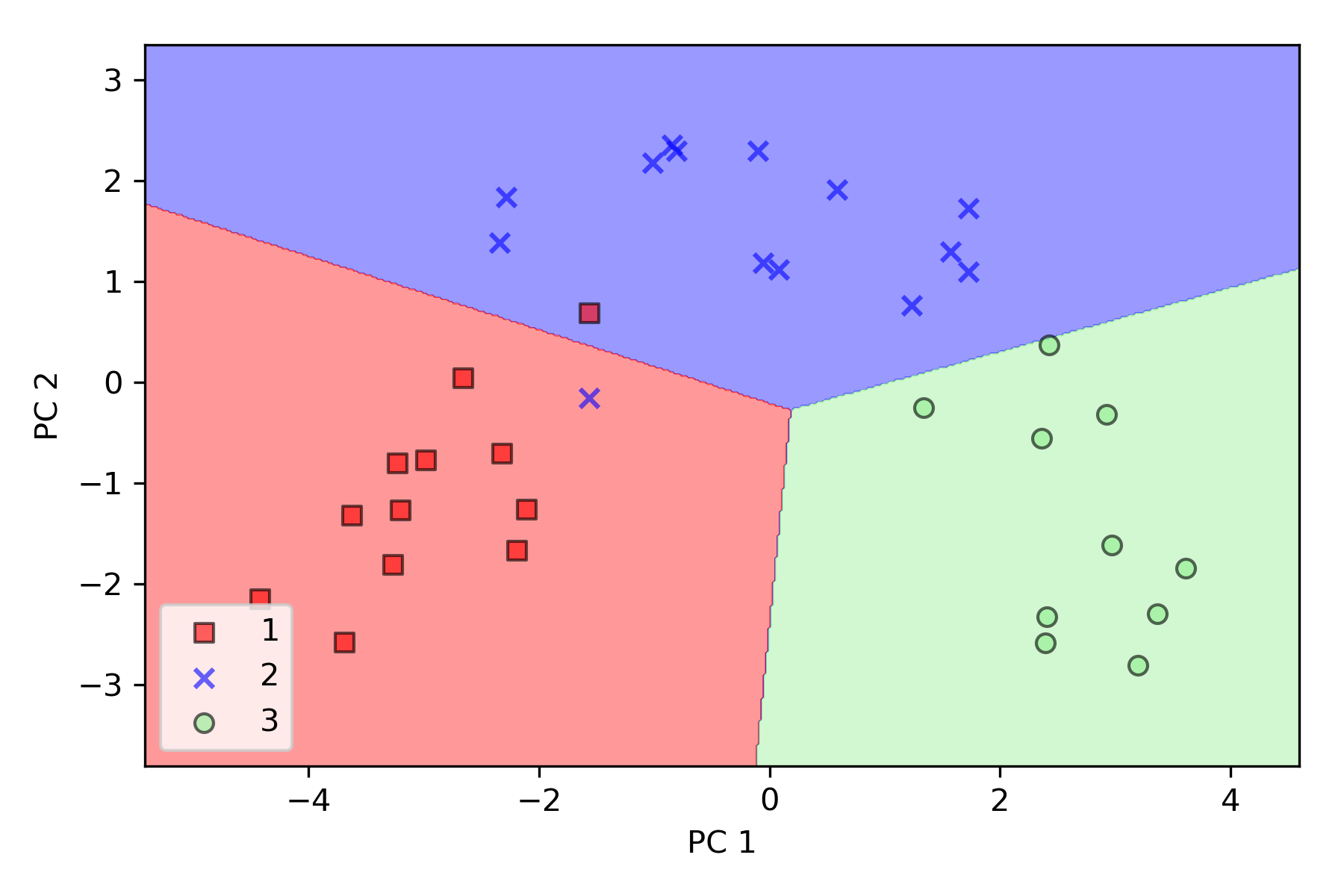
lsv train R^2: 1.000

lsv test R^2: 0.972

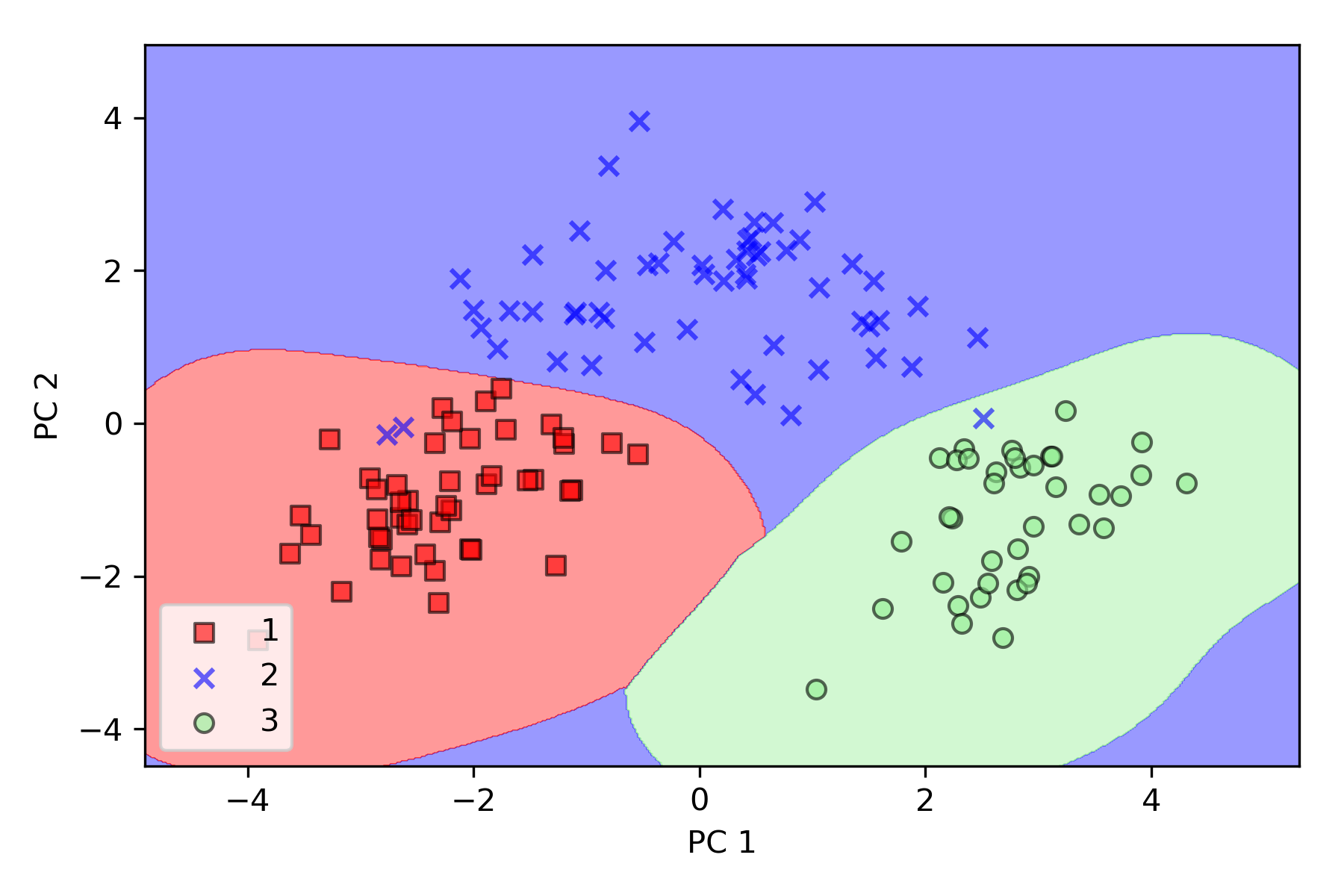
**Part 3: Perform a PCA on both datasets**



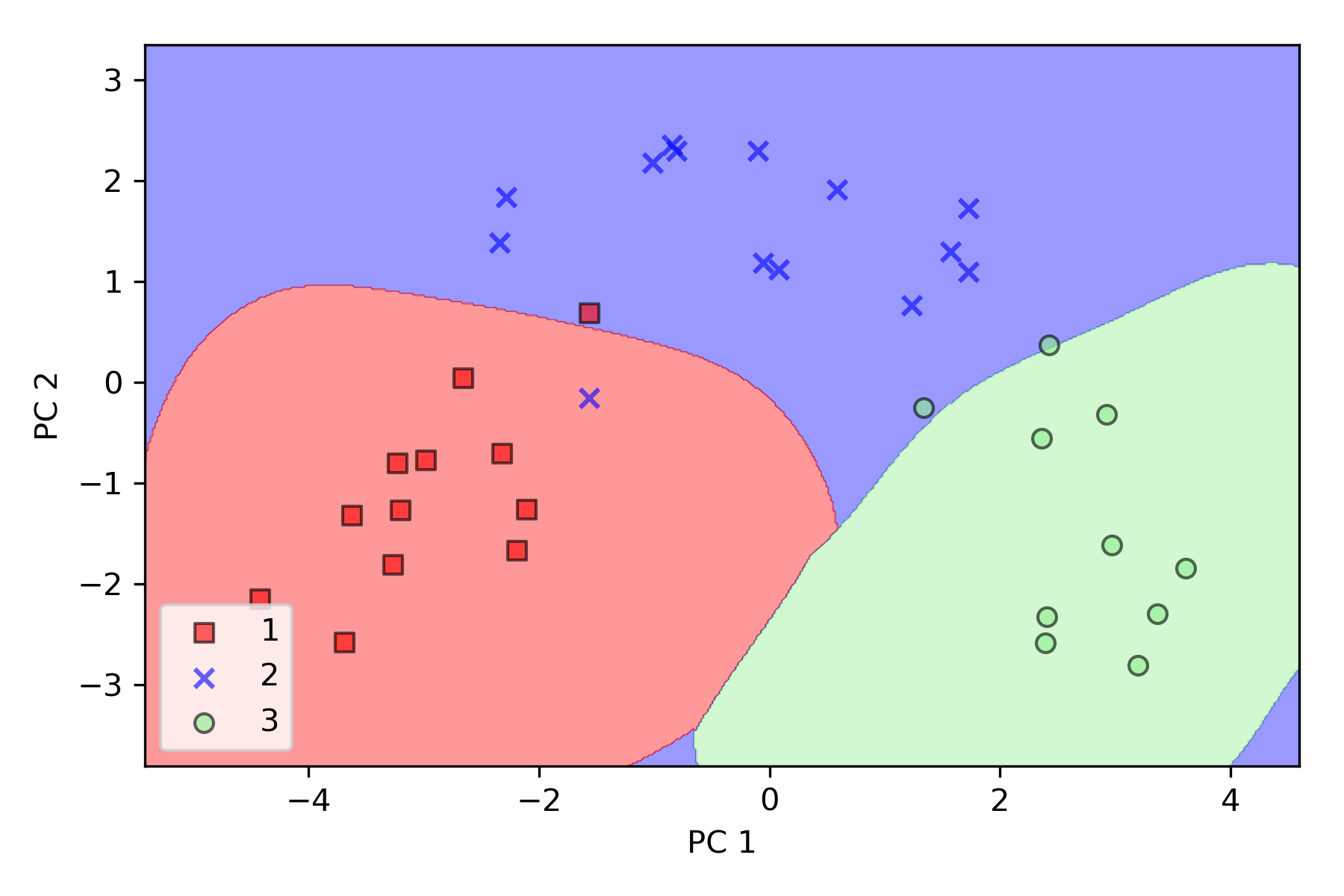
pca lr train R^2: 0.972



pca lr test R^2: 0.944

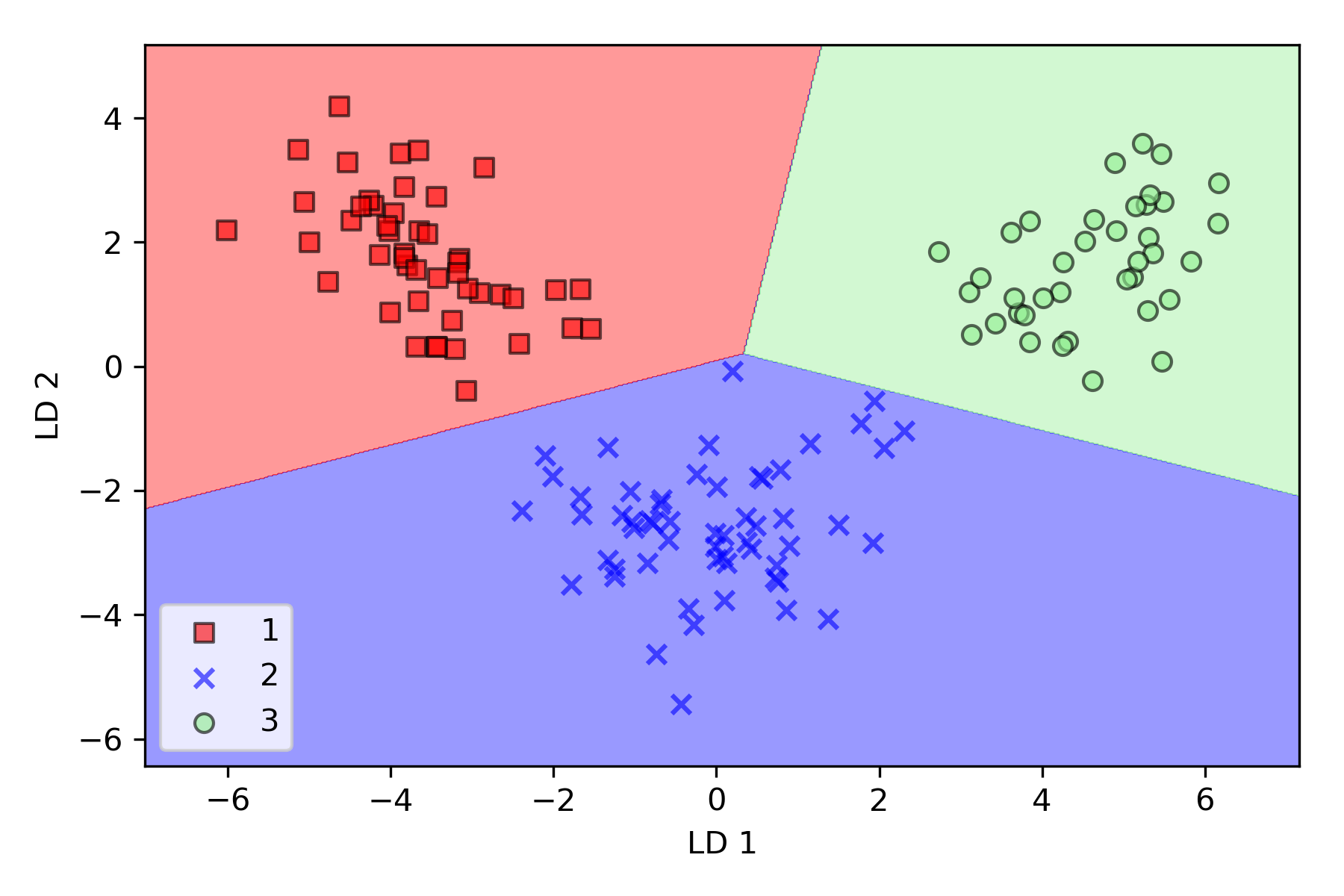


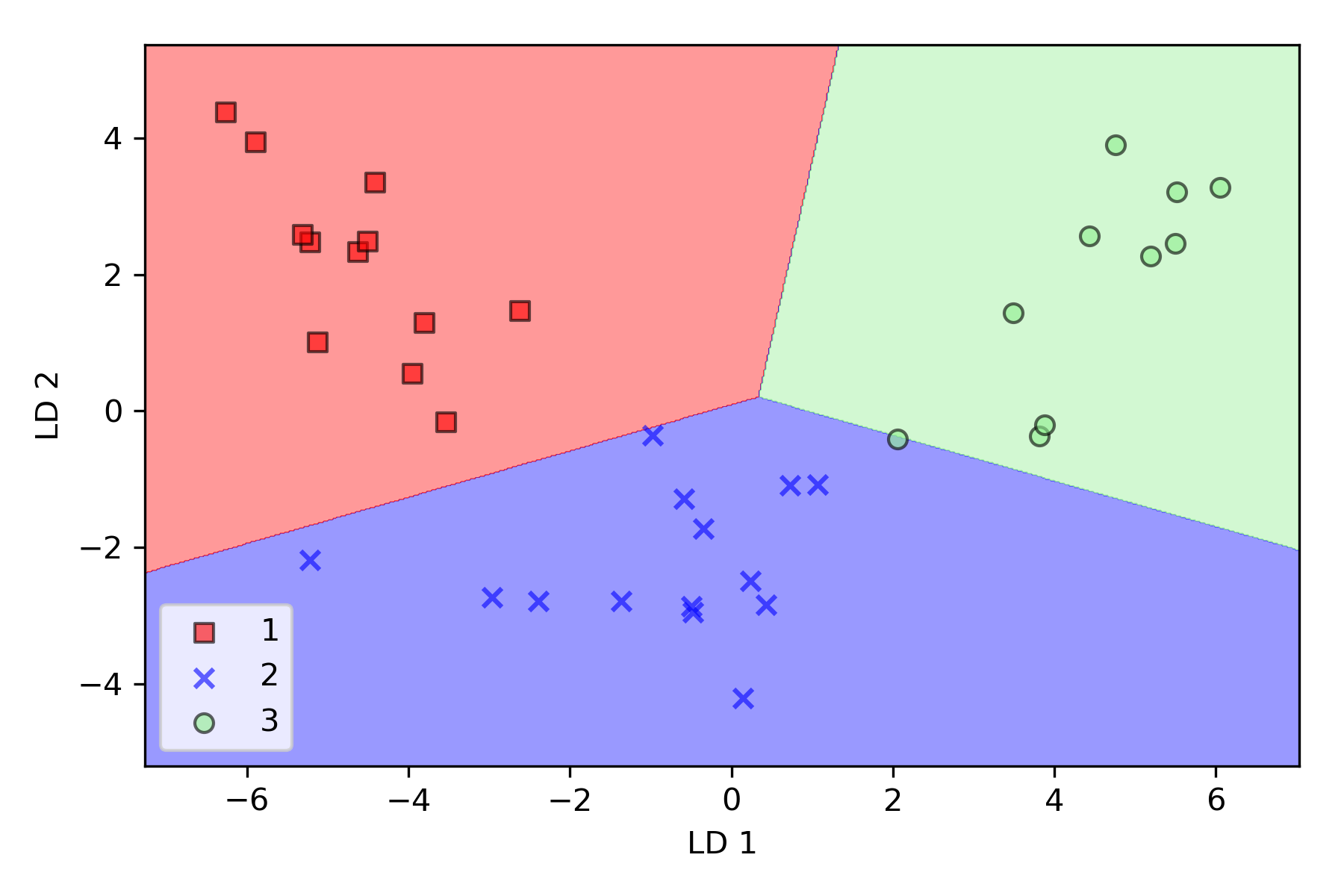
pca sv train R^2: 0.979



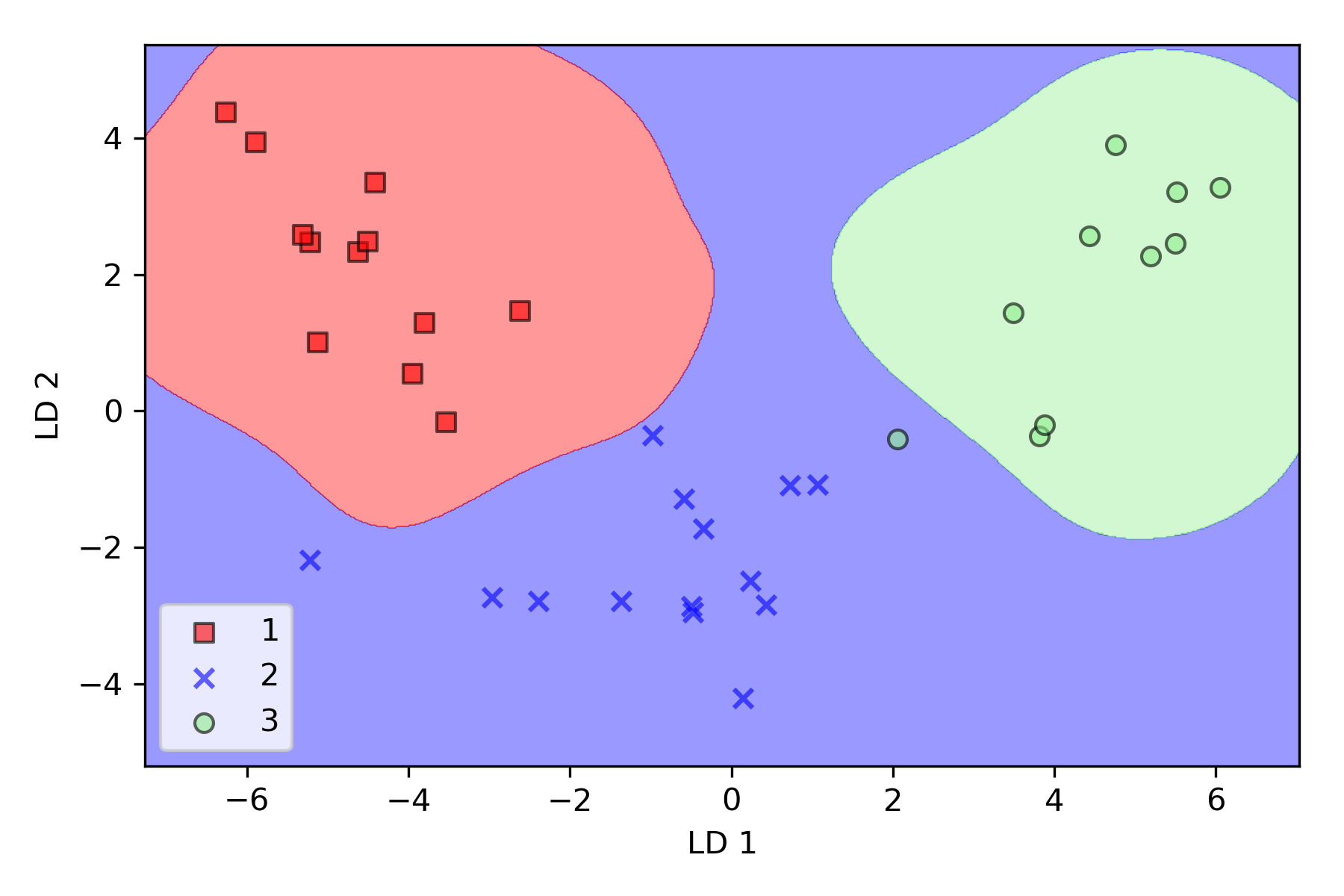
pca sv test R^2: 0.889

**Part 4: Perform and LDA on both datasets**

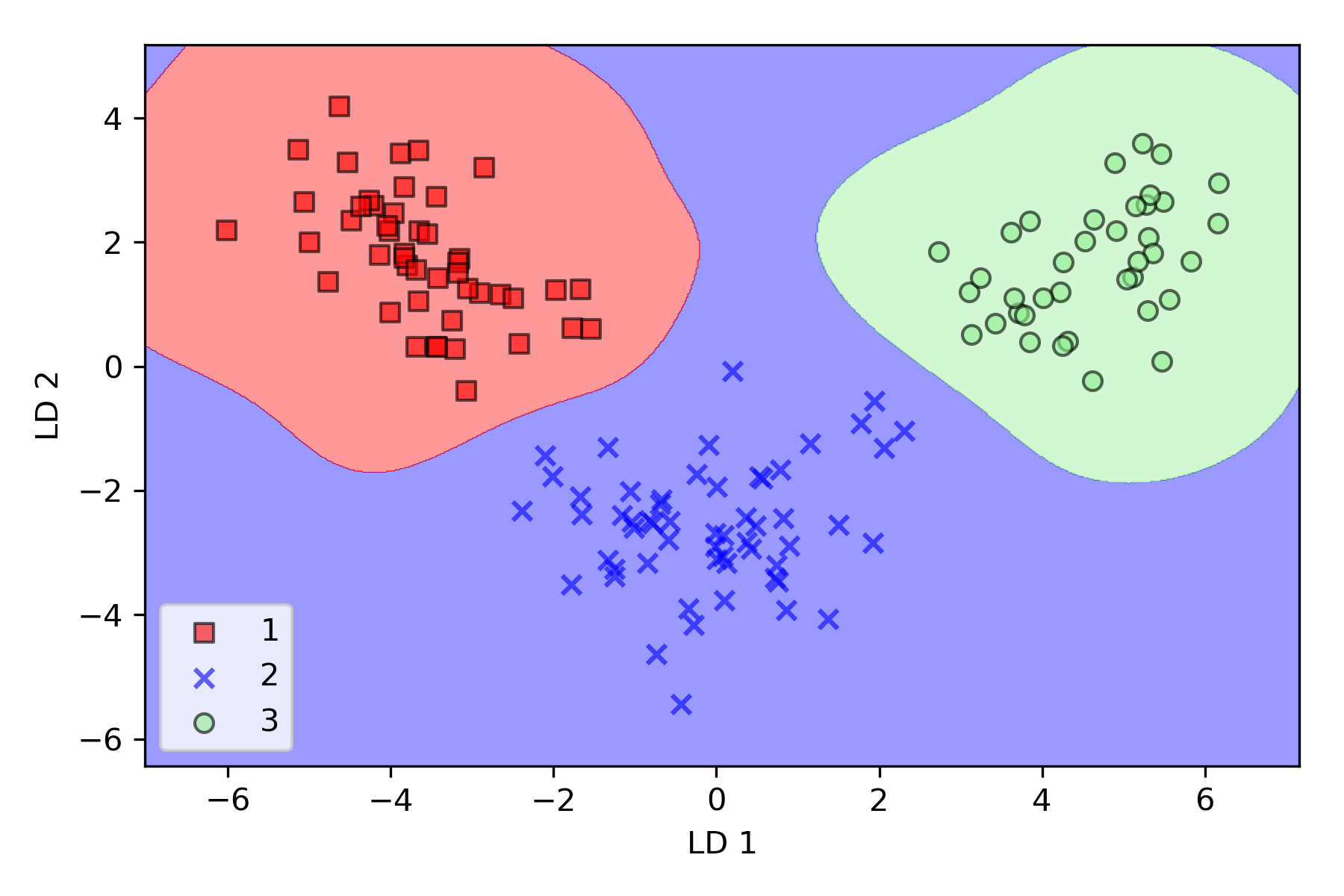
****

lda lr trainR^2: 1.000****

lda lr test R^2: 0.972



lda sv train R^2: 1.000

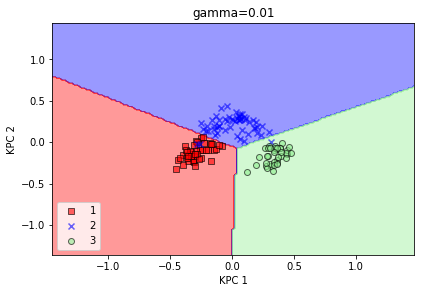


lda sv test R^2: 0.972

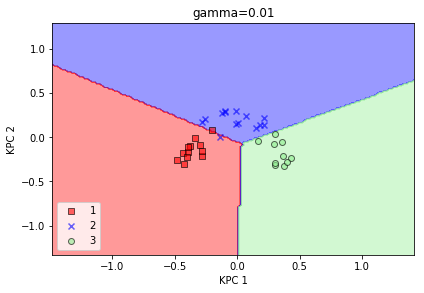
**Part 5: Perform a kPCA on both datasets**

**[See next page]**

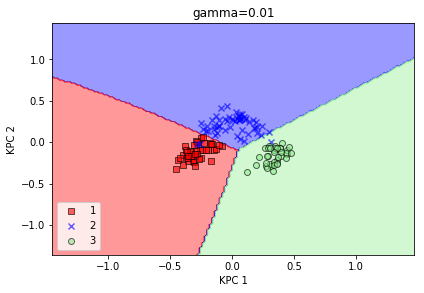
Work best when gamma close to 0.3



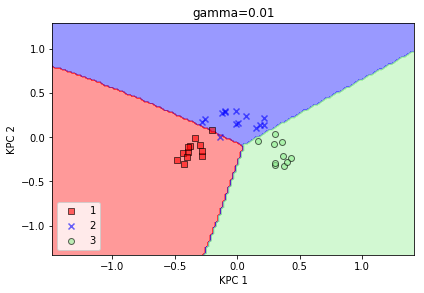
**kpca lr trainR^2: 0.979**



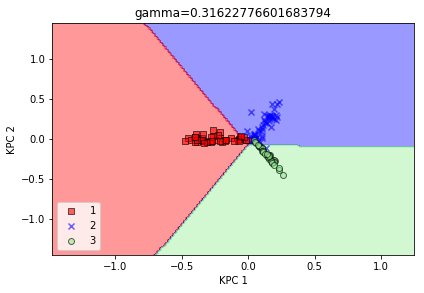
**kpca lr test R^2: 0.944**



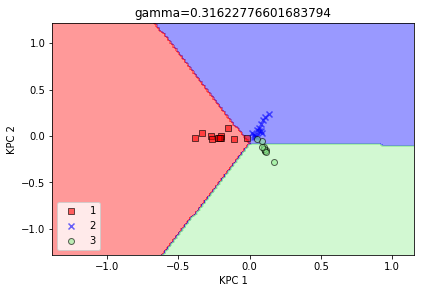
**kpca sv train R^2: 0.979**



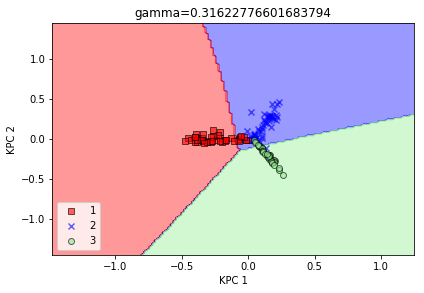
**kpca sv test R^2: 0.944**



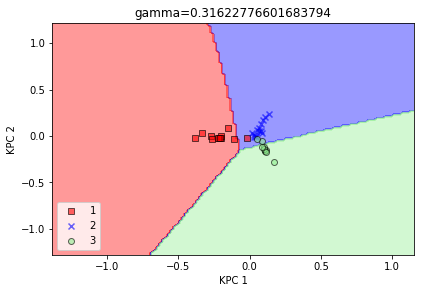
**kpca lr trainR^2: 0.923**



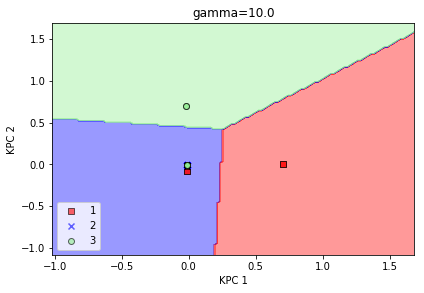
**kpca lr test R^2: 0.917**



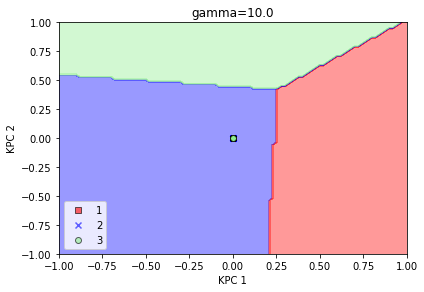
**kpca sv train R^2: 0.873**



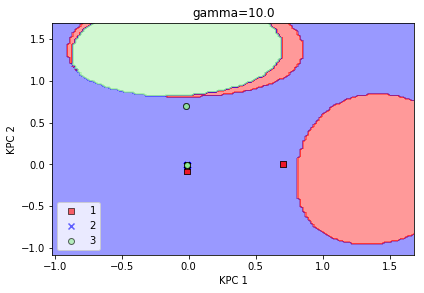
**kpca sv test R^2: 0.917**



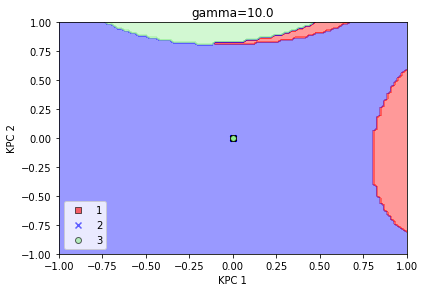
kpca lr trainR^2: 0.430



kpca lr test R^2: 0.389



kpca sv train R^2: 0.401



kpca sv test R^2: 0.389

**Part 6: Conclusions**

The data has a high R^2

Gamma =0.3 have a good graph

Find the different between PCA LDA KPCA

**Part 7: Appendix**

<https://github.com/rainduan/IE598_F18_HW5>

**a**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Experiment 1 (Wine) | | | |  |  |
|  | Logistic | | SVM | | LSVM | |
|  |  |  |  |  |  |  |
| Baseline | Train Acc: | 1 | Train Acc: | 0.993 | Train Acc: | 1 |
| Test Acc: | 1 | Test Acc: | 0.972 | Test Acc: | 0.972 |
|  |  |  |  |  |  |  |
| PCA transform | Train Acc: | 0.972 | Train Acc: | 0.979 | Train Acc: | 0.972 |
| Test Acc: | 0.944 | Test Acc: | 0.889 | Test Acc: | 0.917 |
|  |  |  |  |  |  |  |
| LDA transform | Train Acc: | 1 | Train Acc: | 1 | Train Acc: | 1 |
| Test Acc: | 0.972 | Test Acc: | 0.972 | Test Acc: | 0.972 |
|  |  |  |  |  |  |  |
| kPCA transform | Train Acc: | 0.923 | Train Acc: | 0.873 | Train Acc: | 0.944 |
| Test Acc: | 0.917 | Test Acc: | 0.917 | Test Acc: | 0.944 |