

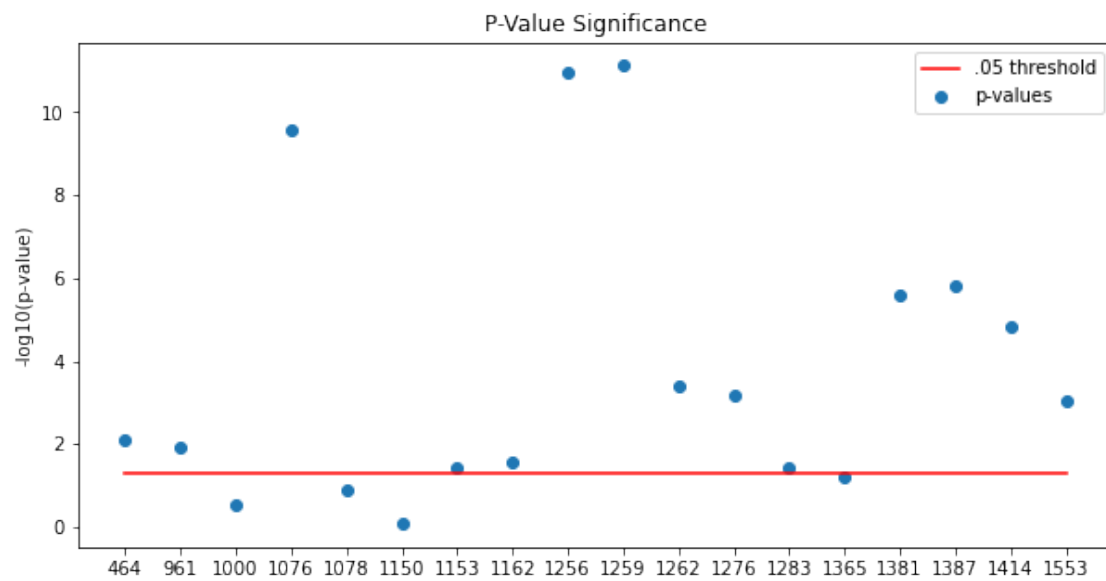
report_executed

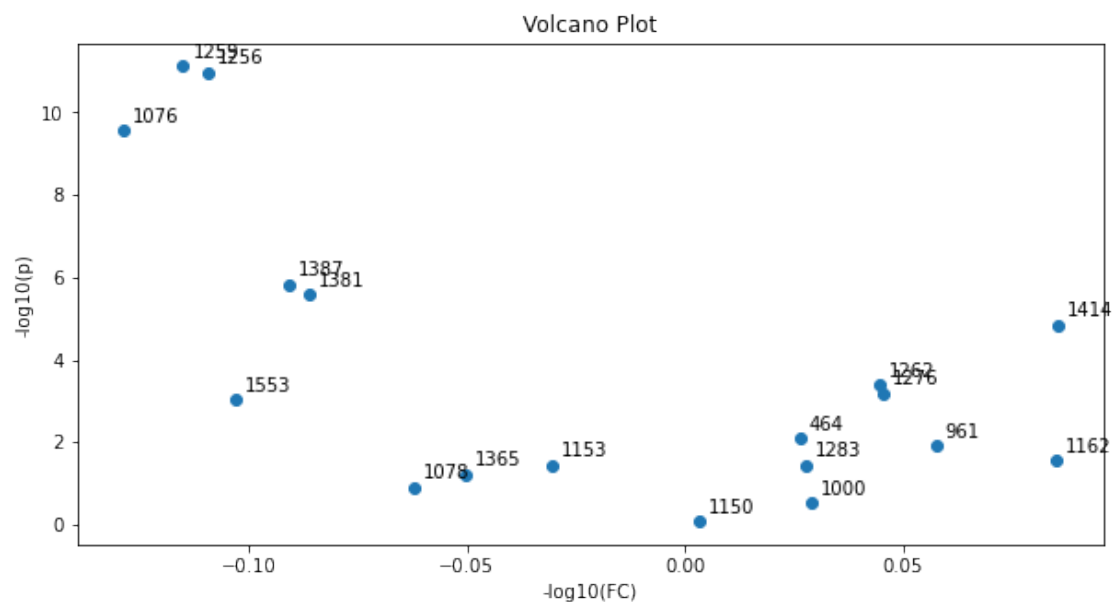
July 14, 2021

1 Statistics

T-test

3.8.8





2 Dimension-Reduction

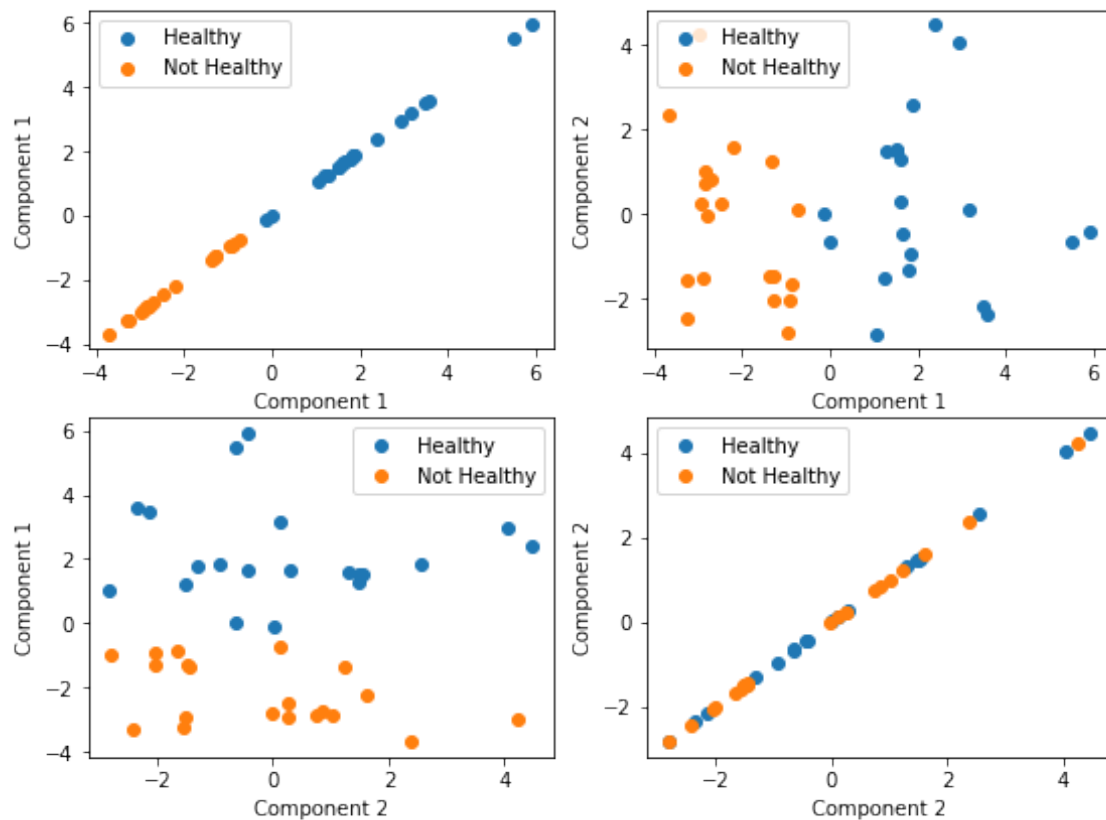
Dimension-reduction methods are used to condense high dimensional data down to dimensions which provide the most information. We have implemented the principal component analysis (PCA). It performs a change of basis and the new basis is chosen, such that the i -th principal component is orthogonal to the first $i-1$ principal components and the direction maximizes the variance of the projected data. We use the Python library sklearn for both PCA and LDA.

LDA

PCA Projections

Projections of data into latent space.

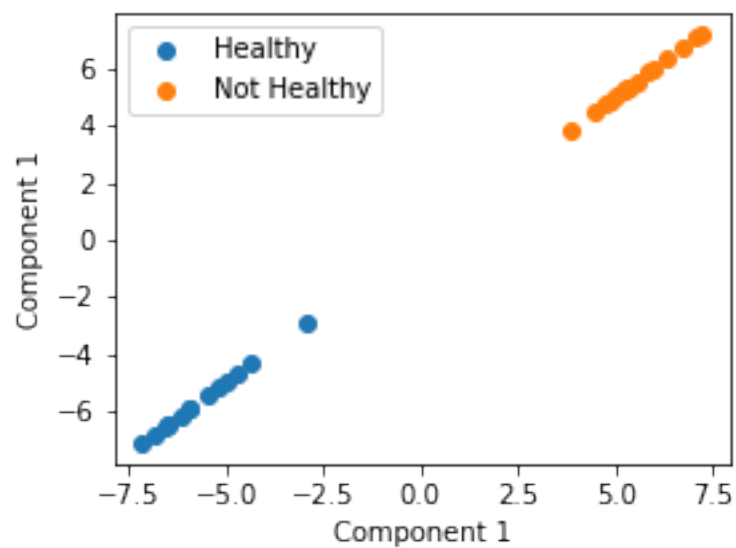
Data is colored by response



LDA Projections

Projections of data into latent space.

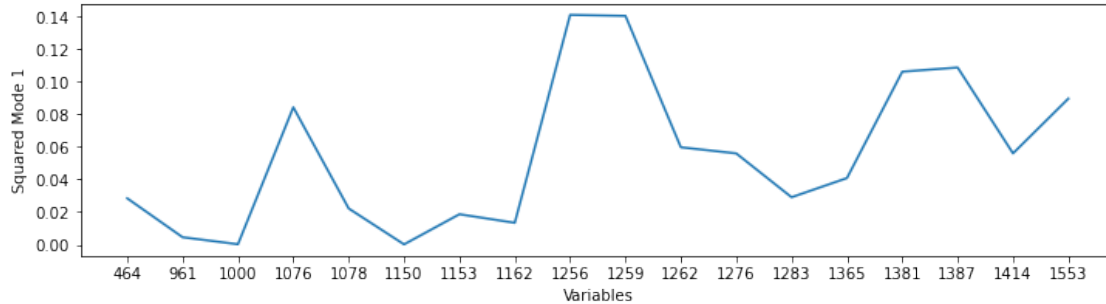
Data is colored by response



PCA Vectors

Plotting the squared loadings of the latent space transformation vectors

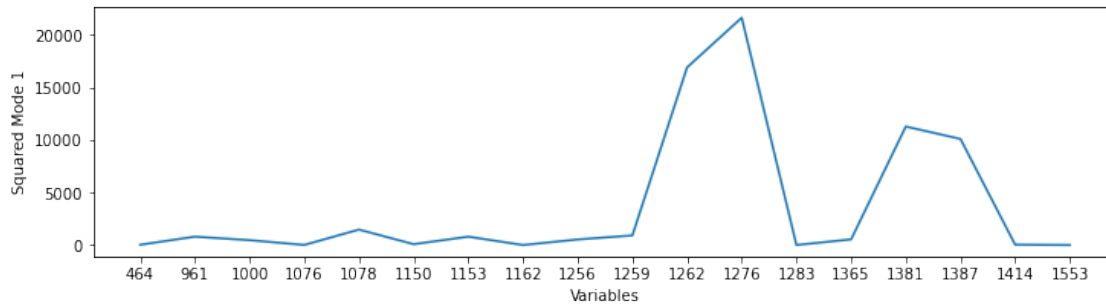
A Larger magnitude indicates larger importance for corresponding feature



LDA Vectors

Plotting the squared loadings of the latent space transformation vectors

A Larger magnitude indicates larger importance for corresponding feature



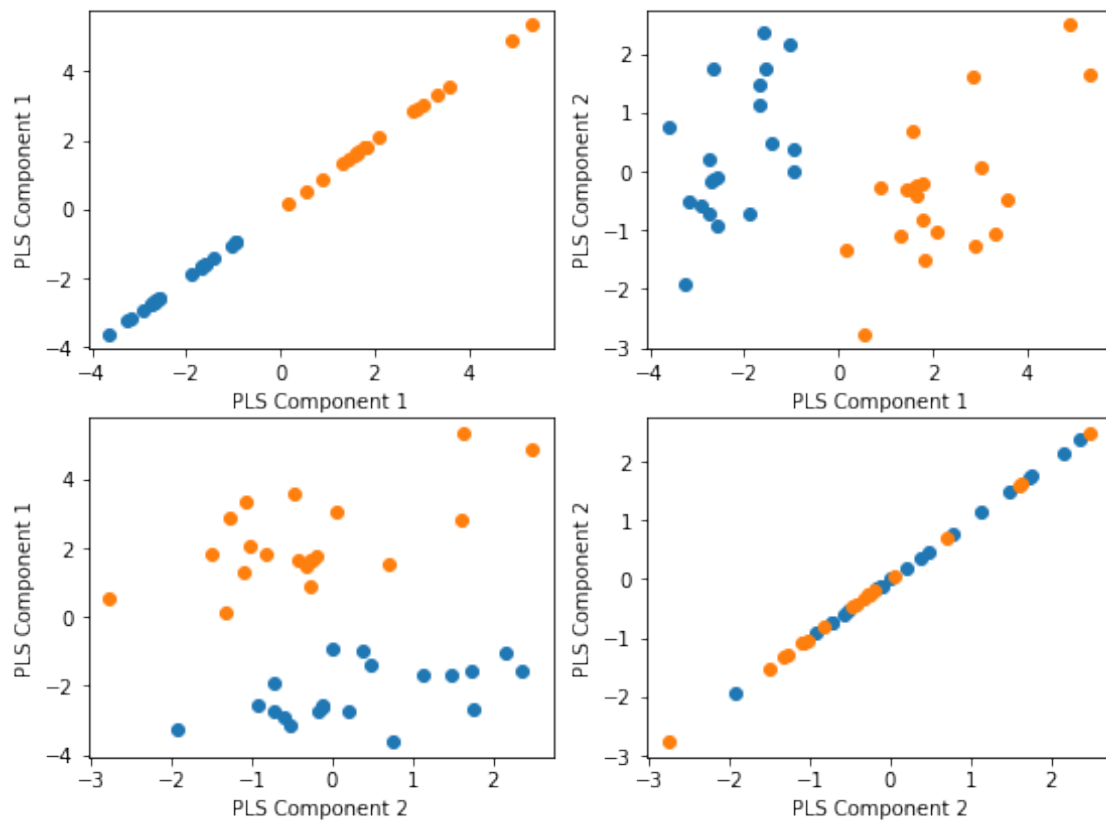
3 Clustering

K-means, hierarchical,

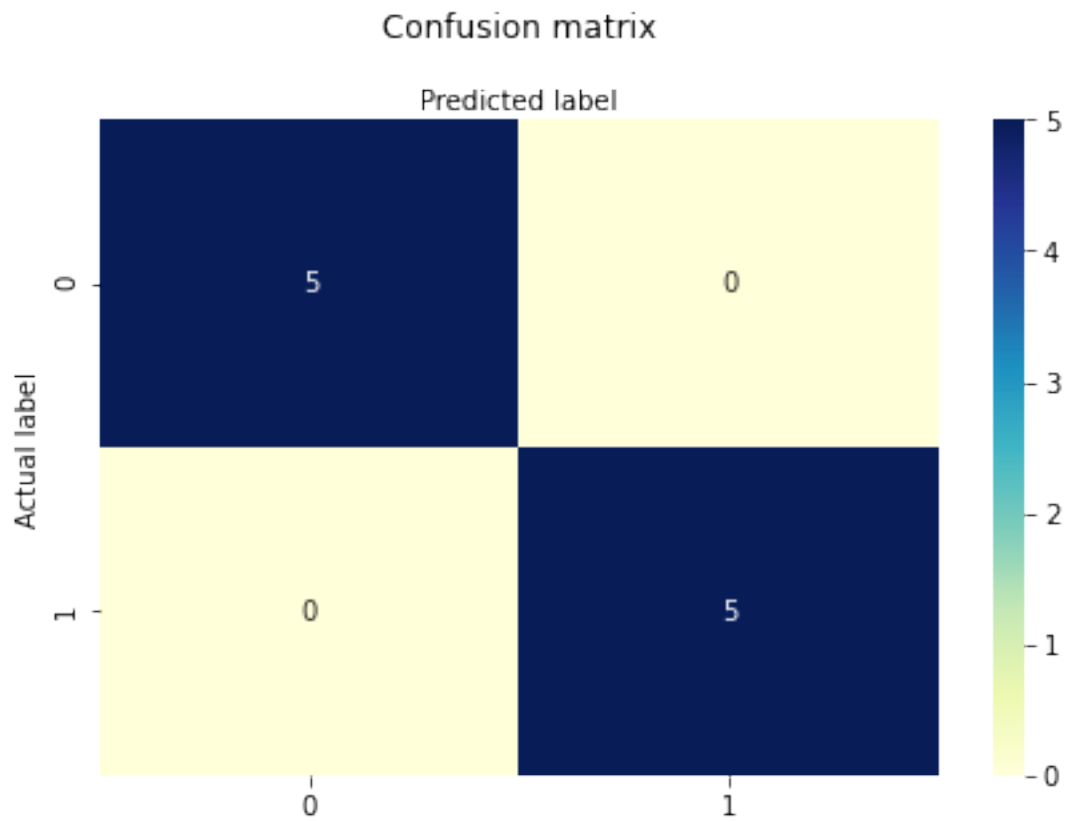
	Cluster 1	Cluster 2	Cluster 3
0	NSCLC_H1437_1	NSCLC_H522_1	NSCLC_A549_1
1	NSCLC_H2228_1	NSCLC_H522_2	NSCLC_H1703_2
2	NSCLC_H2228_2	SCLC_86M1_2	NSCLC_H1703_1
3	NSCLC_H1437_2	SCLC_86M1_1	NSCLC_A549_2
4	NSCLC_H3122_1	SCLC_16HV_1	NSCLC_H322_1
5	NSCLC_H322_2	SCLC_16HV_2	NSCLC_H358_2
6	NSCLC_H3122_2	SCLC_DMS79_1	NSCLC_H358_1
7	NSCLC_HCC4006_1	SCLC_DMS79_2	NSCLC_PC9_1
8	NaN	SCLC_H187_2	NSCLC_PC9_2

9	NaN	SCLC_H187_1	NSCLC_HCC4006_2
10	NaN	SCLC_H209_1	NaN
11	NaN	SCLC_H524_1	NaN
12	NaN	SCLC_H209_2	NaN
13	NaN	SCLC_H524_2	NaN
14	NaN	SCLC_H69_1	NaN
15	NaN	SCLC_H82_1	NaN
16	NaN	SCLC_H82_2	NaN
17	NaN	SCLC_H69_2	NaN
18	NaN	SCLC_N417_2	NaN
19	NaN	SCLC_N417_1	NaN
20	NaN	SCLC_SW210-5_1	NaN
21	NaN	SCLC_SW210_5_2	NaN

	Cluster 1	Cluster 2	Cluster 3
0	SCLC_86M1_2	NSCLC_A549_1	NSCLC_H358_2
1	SCLC_86M1_1	NSCLC_H1703_2	NSCLC_H522_1
2	SCLC_16HV_1	NSCLC_H1703_1	NSCLC_H522_2
3	SCLC_16HV_2	NSCLC_A549_2	NSCLC_H358_1
4	SCLC_DMS79_1	NSCLC_H1437_1	NSCLC_PC9_1
5	SCLC_DMS79_2	NSCLC_H2228_1	NSCLC_PC9_2
6	SCLC_H187_2	NSCLC_H2228_2	NaN
7	SCLC_H187_1	NSCLC_H1437_2	NaN
8	SCLC_H209_1	NSCLC_H3122_1	NaN
9	SCLC_H524_1	NSCLC_H322_2	NaN
10	SCLC_H209_2	NSCLC_H322_1	NaN
11	SCLC_H524_2	NSCLC_H3122_2	NaN
12	SCLC_H69_1	NSCLC_HCC4006_1	NaN
13	SCLC_H82_1	NSCLC_HCC4006_2	NaN
14	SCLC_H82_2	NaN	NaN
15	SCLC_H69_2	NaN	NaN
16	SCLC_N417_2	NaN	NaN
17	SCLC_N417_1	NaN	NaN
18	SCLC_SW210-5_1	NaN	NaN
19	SCLC_SW210_5_2	NaN	NaN



```
SVM Validated Parameters: {'kernel': 'linear', 'shrinking': True}
Random Forest Validated Parameters: {'criterion': 'gini', 'n_estimators': 50}
SVM: R^2=1.0 Q^2=1.0
RF: R^2=1.0 Q^2=1.0
Accuracy: 1.0
<modules.adapml_classification.Classification object at 0x7f8b90b96f40>
```



5 Regression

Linear regression

