

Feature Selection (FS) workflow report

May 18, 2017

Introduction

The report summarizing the Feature Selection pipeline results.

Feature Selection workflow

Univariate canonical correlation (X2) with Recursive Feature Elimination (RFE) wrapped with Random Forest (RF).

Dataset

Transcriptomics analysis of left ventricles of mouse subjected to Isoproterenol challenge (GSE48760).

Summary stats from training phase

Table 1: Best model metrics from 10-folds cross-validation resampling.

Variables	Accuracy	Kappa	AccuracySD	KappaSD
1	0.743	0.4853	0.1325	0.2642
2	0.8432	0.6864	0.09806	0.1959
3	0.8427	0.6856	0.1046	0.2092
4	0.8349	0.6689	0.08152	0.1628
5	0.8416	0.6825	0.08107	0.1622
6	0.8422	0.6842	0.08058	0.161
7	0.8416	0.6825	0.08778	0.1756
8	0.8559	0.711	0.07553	0.1513
9	0.8493	0.6985	0.1085	0.2169
10	0.8564	0.7128	0.1062	0.2123
15	0.8641	0.7284	0.1028	0.2054
20	0.8856	0.7713	0.07581	0.1514
25	0.8713	0.7427	0.09322	0.1863
30	0.8713	0.7427	0.09912	0.1981
35	0.8784	0.757	0.08191	0.1637
40	0.857	0.7142	0.1002	0.2004
45	0.8641	0.7284	0.09708	0.1941
50	0.8713	0.7427	0.09912	0.1981
60	0.8636	0.7271	0.09763	0.1951
70	0.8493	0.6985	0.1085	0.2169
80	0.8636	0.7271	0.09763	0.1951
90	0.8713	0.7427	0.09912	0.1981
100	0.8564	0.7128	0.1007	0.2013
617	0.8713	0.7427	0.09912	0.1981

Summary stats from testing phase

Table 2: Classification metrics from twenty class-balanced and randomized runs.

run	Variables	Accuracy	Kappa	AccuracyPValue
1	20	0.9275	0.8551	4.82e-14
2	25	0.8406	0.6815	7.46e-09
3	8	0.8551	0.7098	1.398e-09
4	6	0.8696	0.7392	2.345e-10
5	9	0.8696	0.7396	2.345e-10
6	15	0.913	0.8261	5.088e-13
7	20	0.9275	0.855	4.82e-14
8	5	0.8406	0.681	7.46e-09
9	9	0.8696	0.7385	2.345e-10
10	25	0.8986	0.7973	4.536e-12
11	50	0.9275	0.8553	4.82e-14
12	90	0.8986	0.7973	4.536e-12
13	70	0.9275	0.8552	4.82e-14
14	8	0.8696	0.7398	2.345e-10
15	80	0.8551	0.7111	1.398e-09
16	30	0.8986	0.797	4.536e-12
17	20	0.8696	0.7396	2.345e-10
18	4	0.8261	0.6521	3.592e-08
19	25	0.913	0.8263	5.088e-13
20	60	0.8986	0.797	4.536e-12

Accuracy_Mean	Accuracy_SD	Accuracy_Max
0.8848	0.0324	0.9275

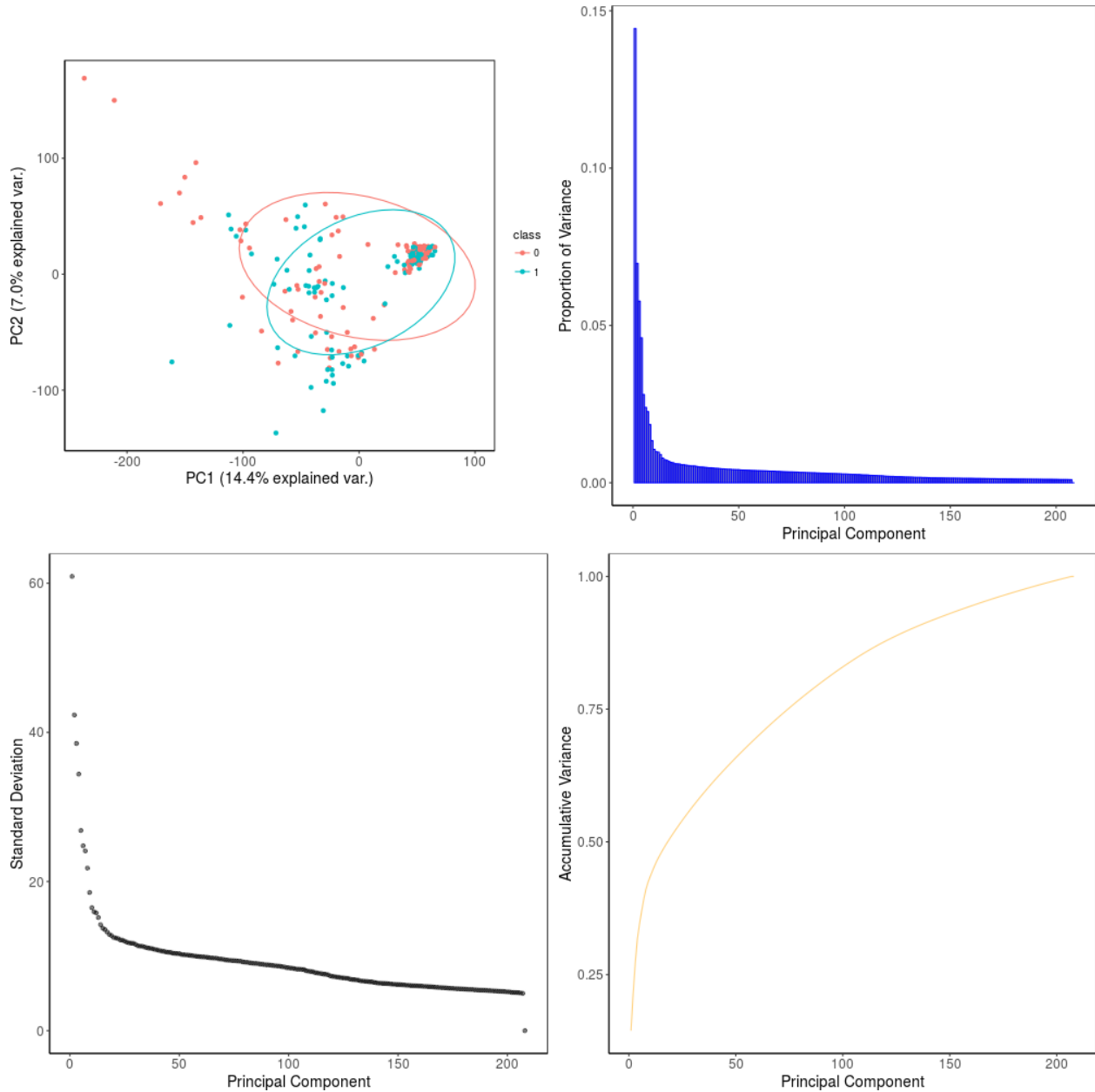
Workflow runtime

9.085 minutes

Plots

Visualization of the classification using PCA

- Groups distribution on the first two Principal Components (PC1 and PC2) from the original data (without apply any FS method).



- Groups distribution on the first two Principal Components (PC1 and PC2) after to apply the FS workflow.

