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 Principal Component Analysis (PCA) follow by 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	0.7835	0.5655	0.07023	0.1423
2	0.8775	0.7548	0.03542	0.07103
3	0.8478	0.6949	0.06799	0.1353
4	0.8918	0.7827	0.03894	0.07873
5	0.906	0.8119	0.0364	0.07313
6	0.8846	0.769	0.0696	0.1393
7	0.8846	0.769	0.05076	0.1017
8	0.8918	0.7833	0.05148	0.1031
9	0.8918	0.7833	0.05148	0.1031
10	0.8989	0.7976	0.03844	0.07715
15	0.8923	0.7847	0.0501	0.1003
20	0.8846	0.7698	0.03798	0.07537
25	0.8846	0.769	0.03798	0.07617
30	0.8769	0.7541	0.06185	0.1232
35	0.8775	0.7548	0.05935	0.1188
40	0.8626	0.7255	0.06501	0.1296
45	0.8918	0.7833	0.05148	0.1031
50	0.8775	0.7548	0.03542	0.07103
60	0.8846	0.769	0.03798	0.07617
70	0.8775	0.7548	0.03542	0.07103
80	0.8846	0.769	0.03798	0.07617
90	0.8846	0.7698	0.03798	0.07537
100	0.8775	0.7548	0.03542	0.07103
208	0.8555	0.7113	0.06945	0.1385

Summary stats from testing phase

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

run	Variables	Accuracy	Kappa	AccuracyPValue
1	35	0.8696	0.7394	2.345e-10
2	25	0.8841	0.7679	3.485e-11
3	30	0.8696	0.7396	2.345e-10
4	50	0.8551	0.7106	1.398e-09
5	25	0.8986	0.7973	4.536e-12
6	5	0.8986	0.7973	4.536e-12
7	8	0.8406	0.6812	7.46e-09
8	30	0.8696	0.7394	2.345e-10
9	9	0.8551	0.7098	1.398e-09
10	9	0.8261	0.6533	3.592e-08
11	**5**	**0.9275**	**0.8551**	**4.82e-14**
12	6	0.8551	0.7103	1.398e-09
13	100	0.913	0.8261	5.088e-13
14	5	0.8551	0.7098	1.398e-09
15	80	0.8406	0.682	7.46e-09
16	9	0.8841	0.7681	3.485e-11
17	50	0.8841	0.7681	3.485e-11
18	7	0.8696	0.7392	2.345e-10
19	3	0.8986	0.7973	4.536e-12
20	2	0.8261	0.653	3.592e-08

Accuracy_Mean	Accuracy_SD	Accuracy_Max
0.871	0.02778	0.9275

Workflow runtime

8.492 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).

```
## PCA plot not available for this FS workflow setting
```

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

```
## PCA plot not available for this FS workflow setting
```