Feature Selection (FS) workflow report

May 18, 2017

Introduction

The report summarizing the Feature Selection pipeline results.

Feature Selection workflow

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	0.699	0.3935	0.1004	0.2025
2	0.6893	0.3745	0.1148	0.2315
3	0.764	0.5257	0.1039	0.2073
4	0.7464	0.486	0.09283	0.1992
5	0.7404	0.4773	0.06267	0.1275
6	0.7255	0.4462	0.07077	0.1476
7	0.7475	0.4908	0.1049	0.2149
8	0.7332	0.4622	0.08544	0.1764
9	0.7475	0.4908	0.06484	0.1378
10	0.77	0.5371	0.0541	0.1094
15	0.77	0.5371	0.06372	0.1284
20	0.7552	0.5068	0.06932	0.1392
25	0.7194	0.4344	0.06975	0.1425
30	0.7547	0.5065	0.05282	0.1052
35	0.7629	0.5229	0.09788	0.1981
40	0.7476	0.4934	0.07124	0.1421
45	0.7613	0.5187	0.1054	0.2128
50	0.7619	0.522	0.08481	0.1696
60	0.7558	0.5097	0.08318	0.1669
70	0.7762	0.5502	0.1123	0.2267
80	0.7624	0.5237	0.07679	0.1535
90	0.7773	0.5537	0.08686	0.1741
100	0.7404	0.4788	0.08048	0.1627
208	0.769	0.5359	0.09787	0.1978

Summary stats from testing phase

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

run	Variables	Accuracy	Kappa	AccuracyPValue
1	8	0.7536	0.5073	2.447e-05
2	50	0.7246	0.4508	0.0001931
3	60	0.7971	0.5948	6.282e-07
4	25	0.7246	0.4489	0.0001931
5	10	0.7681	0.5357	7.807e-06
6	60	0.8261	0.6524	3.592e-08
7	70	0.7971	0.5948	6.282e-07
8	9	0.7971	0.5941	6.282e-07
9	25	0.7391	0.4768	7.122e-05
10	15	0.7536	0.5086	2.447e-05
11	50	0.7391	0.4768	7.122e-05
12	35	0.7681	0.5354	7.807e-06
13	35	0.7826	0.5649	2.306e-06
14	25	0.7971	0.5945	6.282e-07
15	45	0.7826	0.5638	2.306e-06
16	80	0.7536	0.5078	2.447e-05
17	25	0.7536	0.5069	2.447e-05
18	70	0.7826	0.5668	2.306e-06
19	90	0.6957	0.3925	0.00116
20	**90**	**0.8696**	**0.7392**	**2.345e-10**

Accuracy_Mean	Accuracy_SD	Accuracy_Max
0.7703	0.03914	0.8696

Workflow runtime

9.221 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
```