Evaluation of Clustering Algorithms

Derek Chiu

August 17, 2015

Contents

1 Introduction **1**

2 External Evaluation **2**

2.1 Kappa Statistic . . . 3

2.2 Adjusted Rand Index . . . 3

2.3 Mutual Information . . . 3

2.4 Cumulative Distribution Function . . . 4

2.5 Proportion of Ambiguous Clusters . . . 4

3 Internal Evaluation **4**

3.1 Davies-Bouldin Index . . . 4

3.2 Dunn Index . . . 5

3.3 Silhouette Average Width . . . 5

3.4 C-Index . . . 5

3.5 Baker and Hubert Index . . . 6

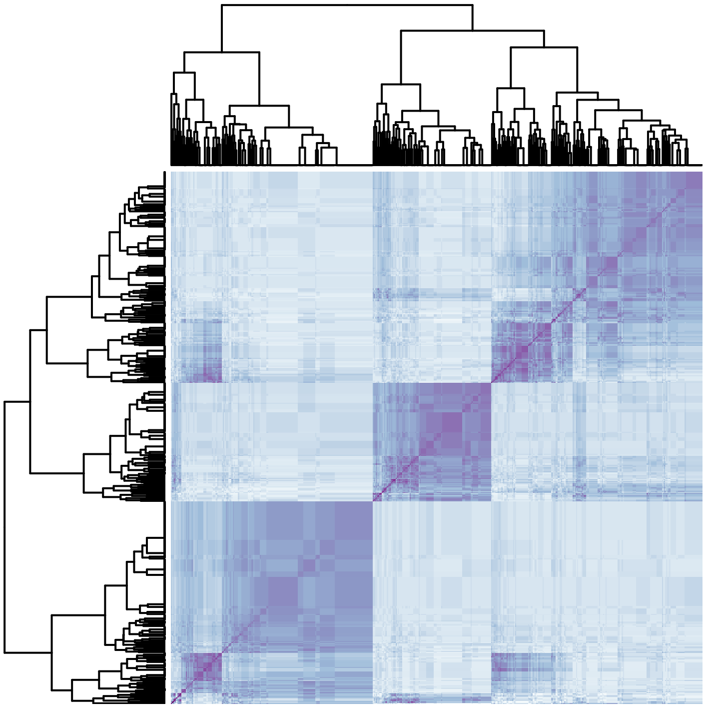
3.6 Calinski-Harabasz Index . . . 6

3.7 Summary . . . 6

4 Ranked Indices **7**

1 Introduction

Cluster analysis is the unsupervised learning method of assigning entities into different groups based on one or more of their attributes. The goal is to place similar objects together and separate dissimilar objects. For example, in genomics studies, we frequently try and cluster patient samples measured on a large number of molecular features. When we get a clustering assignment from an algorithm, we often want to evaluate its performance. Ideally, a good clustering algorithm is able to differentiate entities with no knowledge of the true class labels. In addition, we want the algorithm to arrive at a stable and optimal number of clusters. There are two main categories of clustering evaluation: **external evaluation** and **internal evaluation**.



The proportion of cases with at least 0.6 agreement is 0.2170958.

The confusion matrix is shown below, as well as different metrics for each class.

	C1	C2	C4	C5
C1	104	10	18	7
C2	1	37	10	1
C4	1	60	91	21
C5	3	0	16	109

	Sensitivity	Specificity	Pos Pred Value	Neg Pred Value	Prevalence	Detection Rate	Detection Prevalence	Balanced Accuracy
Class: C1	0.9541	0.9079	0.7482	0.9857	0.2229	0.2127	0.2843	0.931
Class: C2	0.3458	0.9686	0.7551	0.8409	0.2188	0.07566	0.1002	0.6572
Class: C4	0.6741	0.7684	0.526	0.8608	0.2761	0.1861	0.3538	0.7212
Class: C5	0.7899	0.9459	0.8516	0.9197	0.2822	0.2229	0.2618	0.8679

2 External Evaluation

External evaluation usually refers to the case when we compare our clustering assignments to true class labels, or have some gold standard to compare to. In applications, this might be the published clustering result. The downside of using external evaluation is that the reference classes may not be correctly clustered themselves, and we are treating these as the norm. None the less, we can explore a few metrics.

2.1 Kappa Statistic

The unadjusted kappa statistic is 0.5927477 and the weighted kappa statistic is 0.7717546 for the final meta consensus cluster.

2.2 Adjusted Rand Index

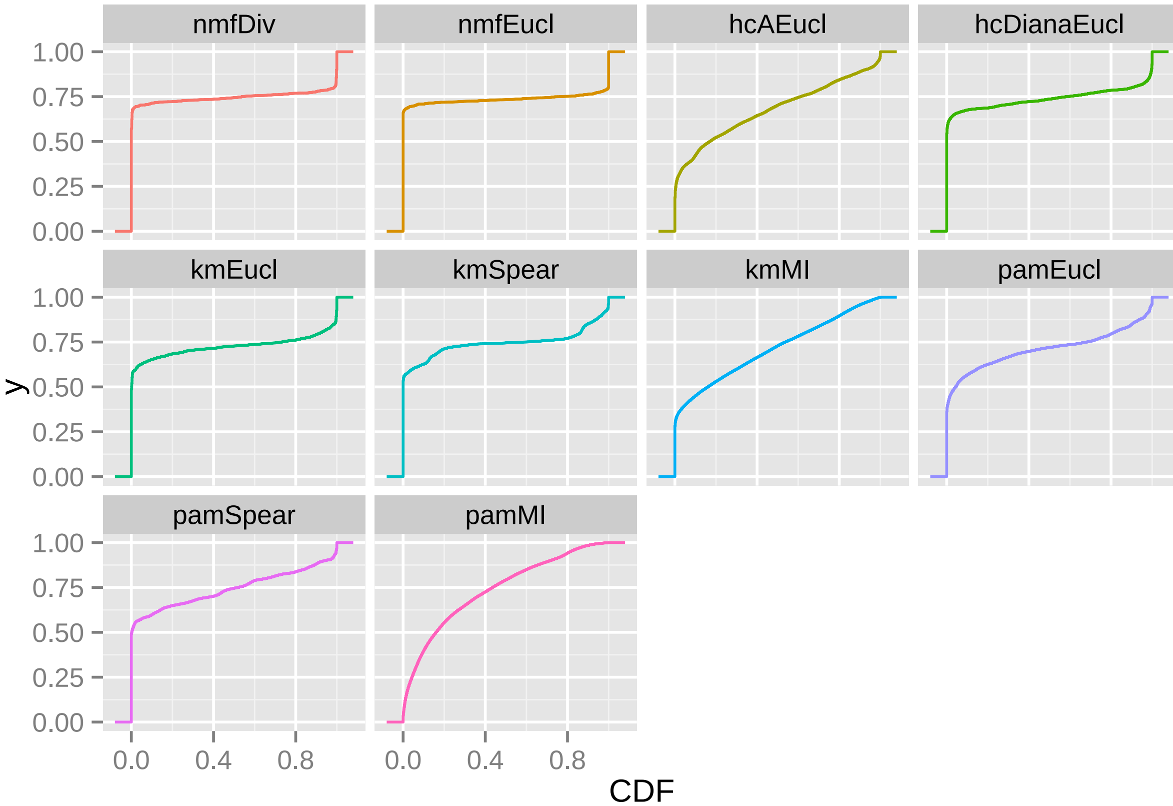
Algorithms	ARI
NMF (Divergence)	0.4799
NMF (Euclidean)	0.4435
PAM (Spearman)	0.427
Hierarchical (Diana)	0.4221
K-Means (Spearman)	0.4049
PAM (Euclidean)	0.3434
Hierarchical (Euclidean)	0.3275
K-Means (Euclidean)	0.2559
PAM (MI)	0.07465
K-Means (MI)	0.07369

2.3 Mutual Information

Algorithms	MI
NMF (Divergence)	0.6723
NMF (Euclidean)	0.6416
PAM (Spearman)	0.621
Hierarchical (Diana)	0.5976
K-Means (Spearman)	0.5889
PAM (Euclidean)	0.546
Hierarchical (Euclidean)	0.481
K-Means (Euclidean)	0.4598
PAM (MI)	0.1652
K-Means (MI)	0.1169

The mutual information for the meta consensus clustering is 0.5987537.

2.4 Cumulative Distribution Function



2.5 Proportion of Ambiguous Clusters

Algorithms	PAC
NMF (Euclidean)	0.1406
NMF (Divergence)	0.3267
Hierarchical (Diana)	0.3829
K-Means (Spearman)	0.425
K-Means (Euclidean)	0.4398
PAM (Spearman)	0.4748
PAM (Euclidean)	0.5984
K-Means (MI)	0.7167
Hierarchical (Euclidean)	0.7836
PAM (MI)	0.9638

The PAC for the meta consensus matrix is 0.7114804.

3 Internal Evaluation

3.1 Davies-Bouldin Index

For DBI, the lower the better.

Algorithms	DBI
Hierarchical (Euclidean)	1.689
NMF (Euclidean)	1.702
NMF (Divergence)	1.702
PAM (Spearman)	1.72
PAM (Euclidean)	1.725
K-Means (Spearman)	1.725
Hierarchical (Diana)	1.731
K-Means (Euclidean)	1.741
PAM (MI)	1.935
K-Means (MI)	1.972

3.2 Dunn Index

For DI, the larger the better.

Algorithms	DI
Hierarchical (Euclidean)	1.04
PAM (Euclidean)	1.028
NMF (Euclidean)	1.003
NMF (Divergence)	0.999
K-Means (Spearman)	0.9926
Hierarchical (Diana)	0.9866
PAM (MI)	0.9821
PAM (Spearman)	0.9492
K-Means (MI)	0.9262
K-Means (Euclidean)	0.9238

3.3 Silhouette Average Width

For SAW, the larger the better.

Algorithms	SAW
Hierarchical (Euclidean)	0.1226
PAM (Euclidean)	0.1171
NMF (Divergence)	0.1111
NMF (Euclidean)	0.1084
PAM (Spearman)	0.1069
K-Means (Spearman)	0.1059
Hierarchical (Diana)	0.09062
K-Means (Euclidean)	0.08707
PAM (MI)	-0.0061
K-Means (MI)	-0.007504

3.4 C-Index

For CI, the lower the better.

Algorithms	CI
K-Means (Euclidean)	0.2816
NMF (Divergence)	0.3023
PAM (Euclidean)	0.31
Hierarchical (Euclidean)	0.3129
Hierarchical (Diana)	0.3212
NMF (Euclidean)	0.3369
PAM (MI)	0.3376
PAM (Spearman)	0.3489
K-Means (MI)	0.3529
K-Means (Spearman)	0.356

3.5 Baker and Hubert Index

For BHI, the larger the better.

Algorithms	BHI
Hierarchical (Euclidean)	2.051
PAM (Euclidean)	1.754
Hierarchical (Diana)	1.746
PAM (Spearman)	1.715
K-Means (Euclidean)	1.676
NMF (Euclidean)	1.666
K-Means (Spearman)	1.627
NMF (Divergence)	1.617
K-Means (MI)	-3.142
PAM (MI)	-3.358

3.6 Calinski-Harabasz Index

For CHI, the larger the better.

Algorithms	CHI
NMF (Divergence)	80.36
NMF (Euclidean)	79.26
K-Means (Spearman)	78.71
PAM (Spearman)	77.19
PAM (Euclidean)	75.71
K-Means (Euclidean)	75.38
Hierarchical (Diana)	74.95
Hierarchical (Euclidean)	66.49
PAM (MI)	13.23
K-Means (MI)	7.322

3.7 Summary

Here is a summary of all the indices for each algorithm, in unsorted order.

Algorithms	DBI	DI	SAW	CI	BHI	CHI
Hierarchical (Diana)	1.731	0.9866	0.09062	0.3212	1.746	74.95
Hierarchical (Euclidean)	1.689	1.04	0.1226	0.3129	2.051	66.49
K-Means (Euclidean)	1.741	0.9238	0.08707	0.2816	1.676	75.38
K-Means (MI)	1.972	0.9262	-0.007504	0.3529	-3.142	7.322
K-Means (Spearman)	1.725	0.9926	0.1059	0.356	1.627	78.71
NMF (Divergence)	1.702	0.999	0.1111	0.3023	1.617	80.36
NMF (Euclidean)	1.702	1.003	0.1084	0.3369	1.666	79.26
PAM (Euclidean)	1.725	1.028	0.1171	0.31	1.754	75.71
PAM (MI)	1.935	0.9821	-0.0061	0.3376	-3.358	13.23
PAM (Spearman)	1.72	0.9492	0.1069	0.3489	1.715	77.19

4 Ranked Indices

The table below shows the ranking of algorithms for performance on a clustering index, for each index. There is an additional column that shows the propoportion of indices where an algorithm was ranked **first or second**.

	DBI	DI	SAW	CI	BHI	CHI	Top
Hierarchical (Diana)	7	6	7	5	3	7	0
Hierarchical (Euclidean)	1	1	1	4	1	8	0.6667
K-Means (Euclidean)	8	10	8	1	5	6	0.1667
K-Means (MI)	10	9	10	9	9	10	0
K-Means (Spearman)	6	5	6	10	7	3	0
NMF (Divergence)	3	4	3	2	8	1	0.3333
NMF (Euclidean)	2	3	4	6	6	2	0.3333
PAM (Euclidean)	5	2	2	3	2	5	0.5
PAM (MI)	9	7	9	7	10	9	0
PAM (Spearman)	4	8	5	8	4	4	0