MATLAB聚类有效性评价指标（外部 成对度量）

作者：凯鲁嘎吉 - 博客园 http://www.cnblogs.com/kailugaji/

更多内容，请看： MATLAB: Clustering Algorithms， MATLAB聚类有效性评价指标（外部）

前提：数据的真实标签已知！TP：真阳性，FP：假阳性，FN：假阴性，TN：真阴性

		PREDICTED CLASS	
		Class=Yes	Class=No
ACTUAL CLASS	Class=Yes	a (TP)	b (FN)
	Class=No	c (FP)	d (TN)

1．MATLAB程序

```
function result = Evaluate(real_label,pre_label)
% This fucntion evaluates the performance of a classification model by
% calculating the common performance measures: Accuracy, Sensitivity,
% Specificity, Precision, Recall, F-Measure, G-mean.
% Input: ACTUAL = Column matrix with actual class labels of the training
%                 examples
%        PREDICTED = Column matrix with predicted class labels by the
%                    classification model
% Output: EVAL = Row matrix with all the performance measures
% https://www.mathworks.com/matlabcentral/fileexchange/37758-performance-measures-for-classification

idx = (real_label()==1);
```

```
p = length(real_label(idx));
n = length(real_label(~idx));
N = p+n;

tp = sum(real_label(idx)==pre_label(idx));
tn = sum(real_label(~idx)==pre_label(~idx));
fp = n-tn;
fn = p-tp;

tp_rate = tp/p;
tn_rate = tn/n;

accuracy = (tp+tn)/N; %准确度
sensitivity = tp_rate;  %敏感性：真阳性率
specificity = tn_rate; %特异性：真阴性率
precision = tp/(tp+fp);  %精度
recall = sensitivity;  %召回率
f_measure = 2*((precision*recall)/(precision + recall));  %F-measure
gmean = sqrt(tp_rate*tn_rate);
Jaccard=tp/(tp+fn+fp); %Jaccard系数

result = [accuracy sensitivity specificity precision recall f_measure gmean Jaccard];
fprintf('accuracy=%.4f, sensitivity=%.4f, specificity=%.4f, precision=%.4f, recall=%.4f, f_measure=%.4f, gmean=%.4f, Jaccard=%.4f\n', ...
    accuracy, sensitivity, specificity, precision, recall, f_measure, gmean, Jaccard);
```

2. 结果

```
>> A = [1 1 1 1 1 1 2 2 2 2 2 2 3 3 3 3 3];
>> B = [1 2 1 1 1 1 1 2 2 2 2 3 1 1 3 3 3];
>> result = Evaluate(A,B)
accuracy=0.7059, sensitivity=0.8333, specificity=0.6364, precision=0.5556, recall=0.8333, f_measure=0.6667, gmean=0.7282, Jaccard=0.5000

result =

   0.705882352941177   0.833333333333333   0.636363636363636   0.555555555555556   0.833333333333333   0.666666666666667   0.728219081254419   0.500000000000000
```

3. 参考

[1] MATLAB聚类有效性评价指标（外部）

[2] 相似性度量

[3] Performance Measures for Classification

[4] Gaussian field consensus论文解读及MATLAB实现