**scikit-learn——模型评估**

**任务目标**

1.了解各种评估算法的原理

2.熟练掌握sklearn中评估算法对应的函数的使用方法

**相关知识**

在使用机器学习算法的过程中，针对不同场景需要不同的评价指标，在这里对常用的指标进行一个简单的汇总。

**一、分类评价指标**

**1. 精确率与召回率**

精确率与召回率多用于二分类问题。精确率（Precision）指的是模型判为正的所有样本中有多少是真正的正样本；召回率（Recall）指的是所有正样本有多少被模型判为正样本，即召回。设模型输出的正样本集合为A，真正的正样本集合为B，则有：

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/ca50956e-143c-4c94-8bcb-d2a7b44d8bac.png)

有时候我们需要在精确率与召回率间进行权衡，一种选择是画出精确率-召回率曲线（Precision-Recall Curve），曲线下的面积被称为AP分数（Average precision score）；另外一种选择是计算Fβ分数：

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/3d9a7fef-07d2-4002-bac3-376b9cdc7da0.png)

当β=1称为F1-score，是分类与信息检索中最常用的指标之一。

对于数据测试结果有下面4种情况：

真阳性（TP）: 预测为正， 实际也为正

假阳性（FP）: 预测为正， 实际为负

假阴性（FN）: 预测为负，实际为正

真阴性（TN）: 预测为负， 实际也为负

**2. ROC**

设模型输出的正样本集合为A，真正的正样本集合为B，所有样本集合为C，我们称

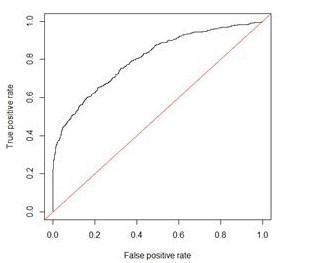
[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/8698532e-5348-4b99-8437-abed6e78b0d2.png)

为真正率（True-positive rate），

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/01.png)

为假正率（False-positive rate）。

ROC曲线适用于二分类问题，以假正率为横坐标，真正率为纵坐标的曲线图，如：

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/2a6253ce-6c29-41a4-9277-1b2b0e42b72d.jpg)

AUC值是曲线下的面积（Area under curve），越大意味着分类器效果越好。

**3. 对数损失**

对数损失（Log loss）亦被称为逻辑回归损失（Logistic regression loss）或交叉熵损失（Cross-entropy loss）。

对于二分类问题，设y∈{0,1}且p=Pr(y=1))，则对每个样本的对数损失为：

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/175750b8-2ef4-4a02-8e58-0842dffd267a.png)

可以很容易地将其扩展到多分类问题上。设Y为指示矩阵，即当样本i的分类为k时yi,k=1；设P为估计的概率矩阵，即pi,k=Pr(ti,k=1)，则对每个样本的对数损失为：

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/d00b345e-bc96-437b-b4cf-74b251e7f7c9.png)

**4. 铰链损失**

铰链损失（Hinge loss）一般用来使“边缘最大化”（maximal margin）。

铰链损失最开始出现在二分类问题中，假设正样本被标记为1，负样本被标记为-1，y是真实值，w是预测值，则铰链损失定义为：

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/86d0130f-4a6c-46bb-9ac1-d96b86a6e4af.png)

然后被扩展到多分类问题，假设yw是对真实分类的预测值，yt是对非真实分类预测中的最大值，则铰链损失定义为：

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/1f474de9-f42f-408c-a8f7-3ba9657c3364.png)

注意，二分类情况下的定义并不是多分类情况下定义的特例。

**5. 混淆矩阵**

混淆矩阵（Confusion Matrix）又被称为错误矩阵，通过它可以直观地观察到算法的效果。它的每一列是样本的预测分类，每一行是样本的真实分类（反过来也可以），顾名思义，它反映了分类结果的混淆程度。混淆矩阵ii行jj列原本是类别ii却被分为类别jj的样本个数，计算完之后还可以对之进行可视化。

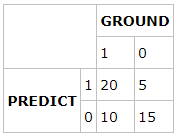
**6. kappa系数**

kappa系数（Cohen's kappa）用来衡量两种标注结果的吻合程度，标注指的是把N个样本标注为C个互斥类别。计算公式为

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/4d7e7ae3-a5e3-4bf0-8ea2-60ed5cced227.png)

其中p0是观察到的符合比例，pe是由于随机性产生的符合比例。当两种标注结果完全相符时，K=1，越不相符其值越小，甚至是负的。

现在举个例子，对于50个测试样本的二分类问题，预测与真实分布情况如下表：



预测与真实值相符共有20+15个，则观察到的符合比例为p0=(20+15)/50=0.7。计算pe比较复杂，PREDICT预测为1的比例为0.5，GROUND中1的比例为0.6，从完全随机的角度来看，PREDICT与GROUND均为1的概率为0.5\*0.6=0.3，PREDICT与GROUND均为0的概率为0.5\*0.4=0.2，则PREDICT与GROUND由于随机性产生的符合比例为0.2+0.3=0.5，即pe=0.5,最后求得：

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/23346dd0-3a12-4f95-a907-04dbc0c094a3.png)

**7. 准确率**

准确率（Accuracy）衡量的是分类正确的比例。设

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/6892b1a2-931c-4bcd-aa23-d457d2ac8361.png)

是第i个样本预测类别，yi是真是类别，在nsample个测试样本上的准确率为

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/db4ab093-6a8c-4c27-ac89-55cccc817948.png)

其中1(x)是indicator function，当预测结果与真实情况完全相符时准确率为1，两者越不相符准确率越低。

虽然准确率适用范围很广，可用于多分类以及多标签等问题上，但在多标签问题上很严格，在有些情况下区分度较差。

**8. 海明距离**

海明距离（Hamming Distance）用于需要对样本多个标签进行分类的场景。对于给定的样本i，

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/5a7c02ea-4189-4df9-8817-43b6900ccbc4.png)

是对第j个标签的预测结果，yi是第j个标签的真实结果，L是标签数量，则

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/5a7c02ea-4189-4df9-8817-43b6900ccbc4.png)

与

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/02.png)

间的海明距离为:

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/f89371f7-676a-4143-ab16-a40b33b553cd.png)

其中l(x)是

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/03.png)

函数。

当预测结果与实际情况完全相符时，距离为0；当预测结果与实际情况完全不符时，距离为1；当预测结果是实际情况的真子集或真超集时，距离介于0到1之间。

我们可以通过对所有样本的预测情况求平均得到算法在测试集上的总体表现情况，当标签数量L为1时，它等于1-Accuracy，当标签数L>1时也有较好的区分度，不像准确率那么严格。

9. 杰卡德相似系数

杰卡德相似系数（ Jaccard similarity coefficients）也是用于需要对样本多个标签进行分类的场景。对于给定的样本i，

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/6892b1a2-931c-4bcd-aa23-d457d2ac8361.png)

是预测结果，yi是真实结果，L是标签数量，则第i个样本的杰卡德相似系数为

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/f3b4bd19-508a-4e19-b2f7-01a4fa493ef4.png)

它与海明距离的不同之处在于分母。当预测结果与实际情况完全相符时，系数为1；当预测结果与实际情况完全不符时，系数为0；当预测结果是实际情况的真子集或真超集时，距离介于0到1之间。

我们可以通过对所有样本的预测情况求平均得到算法在测试集上的总体表现情况，当标签数量LL为1时，它等于Accuracy。

**二、回归评价指标**

回归问题比较简单，所用到的衡量指标也相对直观。假设

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/b5102bbc-0a45-41c9-a8ec-9bf8189068a4.png)

是第i个样本的真实值，

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/6892b1a2-931c-4bcd-aa23-d457d2ac8361.png)

是对第i个样本的预测值。

1、可释方差值（Explained variance score）是根据误差的方差计算得到的

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/0.6222788354242523.png)

2 、平均绝对误差mean\_absolute\_error，又被称为L1范数损失（L1-norm loss）

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/0.24465097751378106.png)

3 、平均平方误差MSE（Mean Squared Error），又被称为L2范数损失（L2-norm loss）

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/0.9018320632797436.png)

4 、中值绝对误差Median absolute error

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/0.18876314615470902.png)

5 、决定系数（Coefficient of determination）又被称为R2 分数：

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/0.5335561842080268.png)

**三、聚类评价指标**

**1 . 兰德指数**

兰德指数（Rand index）需要给定实际类别信息C，假设K是聚类结果，a表示在C与K中都是同类别的元素对数，b表示在C与K中都是不同类别的元素对数，则兰德指数为：

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/099cb269-0f11-41f7-b4db-eb7014828b3d.png)

其中C2nsample数据集中可以组成的总元素对数，RI取值范围为[0,1]，值越大意味着聚类结果与真实情况越吻合。

对于随机结果，RI并不能保证分数接近零。为了实现“在聚类结果随机产生的情况下，指标应该接近零”，调整兰德系数（Adjusted rand index）被提出，它具有更高的区分度：

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/21923130-3205-44aa-bf30-845c5282bfd9.png)

ARI取值范围为[−1,1][−1,1]，值越大意味着聚类结果与真实情况越吻合。从广义的角度来讲，ARI衡量的是两个数据分布的吻合程度。

**2. 互信息**

互信息（Mutual Information）也是用来衡量两个数据分布的吻合程度。假设UU与VV是对NN个样本标签的分配情况，则两种分布的熵（熵表示的是不确定程度）分别为：

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/22f30d49-50ef-4d99-9305-c9057bd28f08.png)

其中

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/05.png)

U与V之间的互信息（MI）定义为：

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/f0d29797-b6bc-49e7-8425-864d32953dc5.png)

其中P(i,j)=|Ui⋂Vj|/N。标准化后的互信息（Normalized mutual information）为：

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/f1cd03da-aaa9-4d3f-b36c-8fda2d866260.png)

与ARI类似，调整互信息（Adjusted mutual information）定义为：

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/3726677b-a697-4032-a0b0-30acda615713.png)

利用基于互信息的方法来衡量聚类效果需要实际类别信息，MI与NMI取值范围为[0,1]，AMI取值范围为[−1,1]，它们都是值越大意味着聚类结果与真实情况越吻合。

**3. 轮廓系数**

轮廓系数（Silhouette coefficient）适用于实际类别信息未知的情况。对于单个样本，设a是与它同类别中其他样本的平均距离，b是与它距离最近不同类别中样本的平均距离，轮廓系数为：

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/3cd9240f-6d3c-406b-8fe2-a3c8d48c9578.png)

对于一个样本集合，它的轮廓系数是所有样本轮廓系数的平均值。

轮廓系数取值范围是[−1,1]，同类别样本越距离相近且不同类别样本距离越远，分数越高。

**4. 同质性Homogeneity 完整性completeness 调和平均V-measure**

同质性homogeneity：每个群集只包含单个类的成员；

完整性completeness：给定类的所有成员都分配给同一个群集。

同质性和完整性分数基于以下公式得出：

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/e603ce2e-f91d-425e-8297-eb0cf14563c3.png)

其中H(C|K)是给定簇赋值的类的条件熵，由以下公式求得：

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/0.9094292639098043.png)

H(C)是类熵，公式为：

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/0.5436059407424343.png)

其中，n是样本总数，nc和nk分别属于类c和类k的样本数，而nc,k是从类c划分到类k的样本数量。

条件熵H(K|C)和类熵H(K)，根据以上公式对称求得。

V-measure是同质性homogeneity和完整性completeness的调和平均数，公式：

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/c437efbc-5dbe-4622-a7d9-fb166c5bf510.png)

优缺点

优点：

* 分数明确：从0到1反应出最差到最优的表现；
* 解释直观：差的调和平均数可以在同质性和完整性方面做定性的分析；
* 对簇结构不作假设：可以比较两种聚类算法如k均值算法和谱聚类算法的结果。

缺点：

以前引入的度量在随机标记方面没有规范化，这意味着，根据样本数，集群和先验知识，完全随机标签并不总是产生相同的完整性和均匀性的值，所得调和平均值V-measure也不相同。特别是，随机标记不会产生零分，特别是当簇的数量很大时。

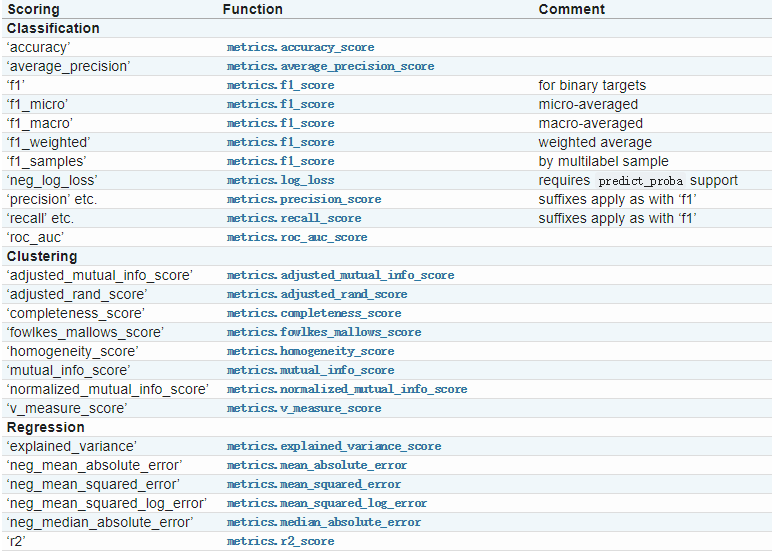
**5.Fowlkes-Mallows scores**

Fowlkes-Mallows指数是针对训练集和验证集数据之间求得的查全率和查准率的几何平均值，其公式为：

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/4b2ae778-ae4f-4fc1-8f57-1e04a83a9862.png)

其中，真阳性（TP）: 预测为正， 实际也为正，假阳性（FP）: 预测为正， 实际为负，假阴性（FN）: 预测为负，实际为正，真阴性（TN）: 预测为负， 实际也为负。

metrics是sklearn用来做模型评估的重要模块，提供了各种评估度量，现在自己整理如下：

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/5f0b5dca-dc6e-4640-8c55-3b707f24374d.png)

**系统环境**

Linux Ubuntu 16.04

Python 3.6.1

Jupyter

**任务内容**

使用sklearn中的metrics模型中的方法，对模型（分类，聚类，回归）进行评估。

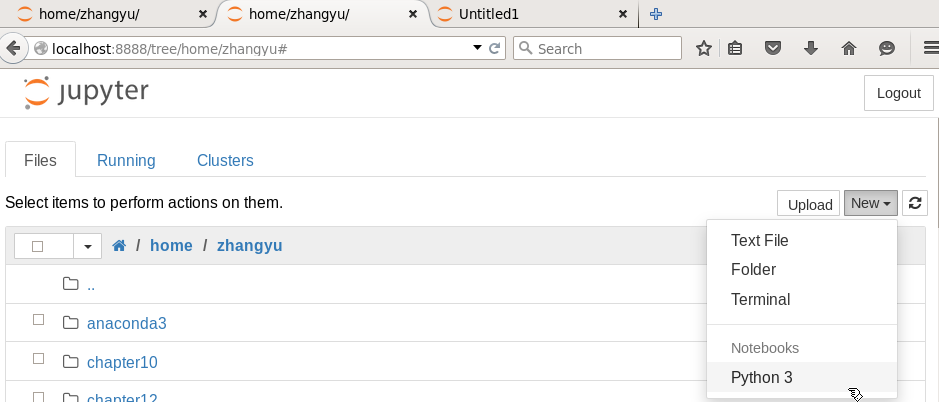
**任务步骤**

1.首先打开终端模拟器，输入下面命令，更新scikit-learn库。

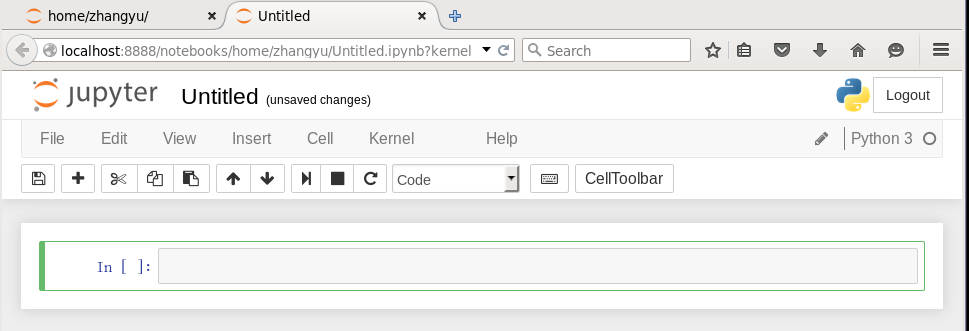
[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. conda update scikit-learn

2.然后在终端模拟器的命令行输入ipython notebook --ip='127.0.0.1'，在浏览器中会打开下面界面，切换到/home/zhangyu目录下，点击New，在其下拉框中选择Python3.

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/205dfc16-99a9-4daf-93e2-970a85696662.png)

3.新建一个ipynb文件，用于编写并执行代码。

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/64175004-973f-479d-8cda-edd07a97f45b.png)

分类模型评估

4.准确率sklearn.metrics.accuracy\_score(y\_true, y\_pred, normalize=True, sample\_weight=None)：预测值与真实值相同的数量比上样本总数。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. **import** numpy as np
2. from sklearn.metrics **import** accuracy\_score
3. y\_pred = [0, 2, 1, 3]
4. y\_true = [0, 1, 2, 3]
5. accuracy\_score(y\_true, y\_pred)

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/68a54026-7891-4579-baf1-10897974a02d.png)

设置accuracy\_score函数的参数normalize=False，返回真实值与预测值相同的个数。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. accuracy\_score(y\_true, y\_pred, normalize=False)

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/e97b73d8-97eb-4de9-a947-77073486dcfb.png)

5.平均准确率average\_precision\_score：针对不平衡数据，对n个类，计算每个类别i的准确率，然后求平均

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. **import** numpy as np
2. from sklearn.metrics **import** average\_precision\_score
3. y\_true = np.array([0, 0, 1, 1])
4. y\_scores = np.array([0.1, 0.4, 0.35, 0.8])
5. average\_precision\_score(y\_true, y\_scores)

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/83610192-1601-4eb5-8c42-bb0151cef958.png)

6.f1\_score值F1\_score：F1-score = 2 \* (precision \* recall) / (precision + recall)，其中，precision（精确率）：所有分正确的正样本比上所有预测为正类的样本数，recall（召回率）：所有分正确的正样本比上所有的正类样本数。函数f1\_score中的参数average=’micro'，表示通过先计算总体的TP，FP和FN的数量，然后计算F1 score。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

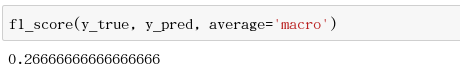
1. from sklearn.metrics **import** f1\_score
2. y\_true = [0, 1, 2, 0, 1, 2]
3. y\_pred = [0, 2, 1, 0, 0, 1]
4. f1\_score(y\_true, y\_pred, average='micro')

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/63197e45-bd4b-42e2-8e46-8bee6aa28a7a.png)

设置参数average=‘ macro'：表示分布计算每个类别的F1，然后做平均（各类别F1的权重相同）

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

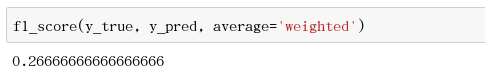
1. f1\_score(y\_true, y\_pred, average='macro')

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/9da67291-a2cc-44bc-8b62-dcb9eff2b80c.png)

设置参数average=‘ weighted'：表示分布计算每个类别的F1，然后做平均(各类别F1的权重为该类别样本数占总样本数的比例)。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

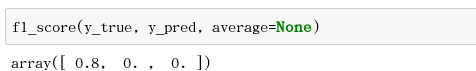
1. f1\_score(y\_true, y\_pred, average='weighted')

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/9efe8ada-ff63-432a-9a5a-769369f8676b.png)

设置参数average=None：表示返回每个类别的F1值。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. f1\_score(y\_true, y\_pred, average=None)

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/91d3f26b-c6a4-416e-bd48-404697f281e6.png)

7.对数损失（Log-loss）：针对分类输出的不是类别而是类别的概率，使用对数损失函数进行评价类别的概率。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. from sklearn.metrics **import** log\_loss
2. log\_loss(["spam", "ham", "ham", "spam"],[[.1, .9], [.9, .1], [.8, .2], [.35, .65]])

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/f8dc04dd-f0fd-42a2-ae7d-7eb1b7ae0d2c.png)

8.精确率(precision\_score)：所有分正确的正样本/所有预测为正类的样本数，即Precision=TP/(TP+FP)。设置参数average=’macro'，表示先计算每个类别的精确率，然后求平均（权重相等）。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. from sklearn.metrics **import** precision\_score
2. y\_true = [0, 1, 2, 0, 1, 2]
3. y\_pred = [0, 2, 1, 0, 0, 1]
4. precision\_score(y\_true, y\_pred, average='macro')

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/ea29f67f-9b63-4814-8c3d-059801ffedbf.png)

设置参数average=‘micro'，表示通过先计算总体的TP，FP和FN的数量，然后计算精确率。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. precision\_score(y\_true, y\_pred, average='micro')

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/30519511-7e7e-43a9-9fe3-25e9f6657648.png)

设置参数average=‘weighted’，表示先计算每个类别的精确率，然后求平均（各个类别精确度的权重为该类别样本数占总样本数的比例）

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. precision\_score(y\_true, y\_pred, average='weighted')

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/9b55f92d-c4cb-4751-aef6-b2d118223b69.png)

设置参数average=None：表示返回每个类别的精确率值。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. precision\_score(y\_true, y\_pred, average=None)

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/1cbaf923-f625-4793-bcd6-e7f33baa7981.png)

9.召回率（recall）：所有分正确的正样本/所有正类的样本数，即Precision=TP/(TP+FN)。设置参数average=’macro'，表示先计算每个类别的召回率，然后求平均（权重相等）。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. from sklearn.metrics **import** recall\_score
2. y\_true = [0, 1, 2, 0, 1, 2]
3. y\_pred = [0, 2, 1, 0, 0, 1]
4. recall\_score(y\_true, y\_pred, average='macro')

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/01bed03e-b181-4288-b48f-47aa64be6ad3.png)

设置参数average=‘micro'，表示通过先计算总体的TP，FP和FN的数量，然后计算召回率。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. recall\_score(y\_true, y\_pred, average='micro')

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/012152ab-9a8b-46c4-8068-80c549f2793b.png)

设置参数average=‘weighted'，表示先计算每个类别的召回率，然后求平均（各个类别精确度的权重为该类别样本数占总样本数的比例）

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. recall\_score(y\_true, y\_pred, average='weighted')

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/6ec3b5ba-dbb3-4ff6-be72-248fa8825257.png)

10.混合矩阵(confusion\_matrix)：被称为错误矩阵，通过它可以直观地观察到算法的效果。它的每一列是样本的预测分类，每一行是样本的真实分类（反过来也可以），顾名思义，它反映了分类结果的混淆程度。混淆矩阵i行j列的原本是类别i却被分为类别j的样本个数，计算完之后还可以对之进行可视化。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

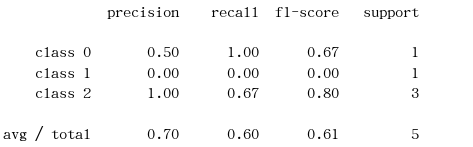
1. from sklearn.metrics **import** confusion\_matrix
2. y\_true=[2,0,2,2,0,1]
3. y\_pred=[0,0,2,2,0,2]
4. confusion\_matrix(y\_true,y\_pred)

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/a99e7ab7-3349-48e1-a948-d8e9667f3fa4.png)

11.分类报告（classification\_report）：返回precision/recall/fi-score/均值/分类个数。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. from sklearn.metrics **import** classification\_report
2. y\_true = [0, 1, 2, 2, 2]
3. y\_pred = [0, 0, 2, 2, 1]
4. target\_names = ['class 0', 'class 1', 'class 2']
5. print(classification\_report(y\_true, y\_pred, target\_names=target\_names))

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/92e15f52-90a2-42f2-a30a-57b34ef1dc83.png)

12.受试者工作特征曲线（roc\_curve）：其实是多个混淆矩阵的结果组合，这里的roc\_curve函数只适合做二分类模型评估。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. **import** numpy as np
2. from sklearn **import** metrics
3. y=np.array([1,1,2,2])
4. scores=np.array([0.1,0.4,0.35,0.8])
5. fpr, tpr, thresholds = metrics.roc\_curve(y, scores, pos\_label=2)
6. print(fpr,tpr,thresholds)

下面返回fpr，tpr，阀值这三个值：

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/b8a83082-43f6-4b66-8da5-d5b5d52de4a0.png)

13.ROC曲线下面的面积(roc\_auc\_score)：面积值越大，模型越精确，在二分类和多分类模型评估中可以用。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. **import** numpy as np
2. from sklearn.metrics **import** roc\_auc\_score
3. y\_true = np.array([0, 0, 1, 1])
4. y\_scores = np.array([0.1, 0.4, 0.35, 0.8])
5. roc\_auc\_score(y\_true, y\_scores)

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/fa6d95fc-bd75-49f5-8c18-c0c5b03dd3b4.png)

14. Cohen’s kappa统计量（cohen\_kappa\_score）：值是介于-1—1之间，当值大于0.8时，普遍认为模型很好，一般可用在二分类和多分类问题中，但不能用在多标签问题。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. from sklearn.metrics **import** cohen\_kappa\_score
2. y\_true = [2, 0, 2, 2, 0, 1]
3. y\_pred = [0, 0, 2, 2, 0, 2]
4. cohen\_kappa\_score(y\_true,y\_pred)

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/08280fdf-21d9-42aa-a984-ccae6b6b3f07.png)

15.Hamming Loss：计算两个样本集合之间的平均汉明损失

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. from sklearn.metrics **import** hamming\_loss
2. y\_pred=[1,2,3,4]
3. y\_true=[2,2,3,4]
4. print(hamming\_loss(y\_true,y\_pred))

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/103768bd-8b25-4d70-b7d6-756481cbc1fd.png)

16.0-1损失（zero\_one\_loss）：该函数计算nsample个样本上的0-1分类损失（L0-1)的和或者平均值。默认情况下，返回的是所以样本上的损失的平均损失，把参数normalize设置为False，就可以返回损失值的和。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

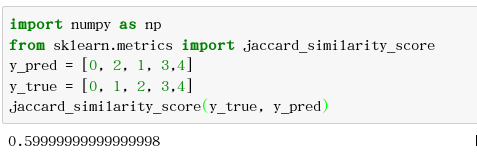
1. from sklearn.metrics **import** zero\_one\_loss
2. **import** numpy as np
3. #二分类问题
4. y\_pred=[1,2,3,4]
5. y\_true=[2,2,3,4]
6. print(zero\_one\_loss(y\_true,y\_pred))
7. print(zero\_one\_loss(y\_true,y\_pred,normalize=False))
8. #多分类标签问题
9. print(zero\_one\_loss(np.array([[0,1],[1,1]]),np.ones((2,2))))
10. print(zero\_one\_loss(np.array([[0,1],[1,1]]),np.ones((2,2)),normalize=False))

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/4028a4fb-88d6-43c0-bb0e-1084e03773e7.png)

17.杰卡德相似系数（Jaccard）:Jaccard相似度分函数jaccard\_similarity\_score，计算平均（默认）或Jaccard相似系数的总和，也称为Jaccard指数，在一对标签集之间。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. **import** numpy as np
2. from sklearn.metrics **import** jaccard\_similarity\_score
3. y\_pred = [0, 2, 1, 3,4]
4. y\_true = [0, 1, 2, 3,4]
5. jaccard\_similarity\_score(y\_true, y\_pred)

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/eaeb9b32-4520-4e7f-abb7-c6ad0dd51233.png)

设置jaccard\_similarity\_score函数的参数normalize=False，表示计算样本集中jaccard相似度的和。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. jaccard\_similarity\_score(y\_true, y\_pred, normalize=False)

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/fc1112ce-0278-41c3-8af5-dfa021a38f05.png)

**回归评价指标**

18.可释方差值（Explained variance score），值取1时回归效果最好，取值越小回归模型效果越差。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. from sklearn.metrics **import** explained\_variance\_score
2. y\_true = [3, -0.5, 2, 7]
3. y\_pred = [2.5, 0.0, 2, 8]
4. explained\_variance\_score(y\_true, y\_pred)

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/6d4b151b-c019-4e77-88c2-0fa3f2c006bf.png)

当输入数据为多维时，设置explained\_variance\_score函数中参数multioutput=‘raw\_values’，表示在多输出多输入的情况下返回一组完整的方差值。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. y\_true = [[0.5, 1], [-1, 1], [7, -6]]
2. y\_pred = [[0, 2], [-1, 2], [8, -5]]
3. explained\_variance\_score(y\_true, y\_pred, multioutput='raw\_values')

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/68f9a88b-7858-4cc1-9f79-05fd561986a7.png)

设置explained\_variance\_score函数中参数multioutput=‘uniform\_average’，表示先计算每个的可解释方差值，然后求平均。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. explained\_variance\_score(y\_true, y\_pred, multioutput='uniform\_average')

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/a7786272-9d9c-4ddf-a2ee-fcd88f5eff3e.png)

设置explained\_variance\_score函数中参数multioutput=‘variance\_weighted’，表示在先计算每个的可解释方差值，然后由每个单独的输出的方差来加权平均。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. explained\_variance\_score(y\_true, y\_pred, multioutput='variance\_weighted')

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/2d32545b-8f28-4db4-9cb6-1fd819d6b86d.png)

19.平均绝对误差（Mean absolute error）：平均绝对误差函数计算的是绝对误差，这是一种与绝对误差损失或1-范数损失的期望值相对应的风险指标。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. from sklearn.metrics **import** mean\_absolute\_error
2. y\_true = [3, -0.5, 2, 7]
3. y\_pred = [2.5, 0.0, 2, 8]
4. mean\_absolute\_error(y\_true, y\_pred)

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/4005d56a-e67b-489a-81d8-414832d946d6.png)

在多输出多输入的情况下，设置mean\_absolute\_error函数中参数multioutput=‘raw\_values’，表示在多输出多输入的情况下返回一组完整的误差值。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. y\_true = [[0.5, 1], [-1, 1], [7, -6]]
2. y\_pred = [[0, 2], [-1, 2], [8, -5]]
3. mean\_absolute\_error(y\_true,y\_pred,multioutput='raw\_values')

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/f1304206-e52e-48f8-a7ce-ca2eac9b9db8.png)

在多输出多输入的情况下，设置mean\_absolute\_error函数中参数multioutput=‘uniform\_average’，表示先计算各自的平均绝对误差值，然后求平均。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. mean\_absolute\_error(y\_true,y\_pred,multioutput='uniform\_average')

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/c77322f8-6473-46cd-b4f3-f979b9733c5a.png)

在多输出多输入的情况下，设置mean\_absolute\_error函数中参数multioutput=[0.3, 0.7]，表示先计算各自的平均绝对误差值，然后求按参数multioutput中的值进行加权求平均。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. mean\_absolute\_error(y\_true,y\_pred,multioutput=[0.3,0.7])

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/3d8ce251-5ca2-4ff4-b5e0-74c661933de9.png)

20.均方误差（Mean squared error）：均方误差函数计算均方误差，这是与平方（二次）误差或损失的期望值相对应的风险度规。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. from sklearn.metrics **import** mean\_squared\_error
2. y\_true = [3, -0.5, 2, 7]
3. y\_pred = [2.5, 0.0, 2, 8]
4. mean\_squared\_error(y\_true, y\_pred)

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/59377415-2ed6-453e-9921-e89dda1d6988.png)

在多输出多输入的情况下，mean\_squared\_error函数中参数multioutput='raw\_values'，表示在多输出多输入的情况下返回一组完整的误差值。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. y\_true = [[0.5, 1],[-1, 1],[7, -6]]
2. y\_pred = [[0, 2],[-1, 2],[8, -5]]
3. mean\_squared\_error(y\_true,y\_pred,multioutput='raw\_values')

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/9dde6b8f-e8eb-40cf-af30-979143111d9e.png)

在多输出多输入的情况下，mean\_squared\_error函数中参数multioutput=[0.3, 0.7]，表示先计算各自的均方误差值，然后求按参数multioutput中的值进行加权求平均。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. mean\_squared\_error(y\_true, y\_pred, multioutput=[0.3, 0.7])

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/16630915-b095-4a6e-8b09-f25c31529066.png)

21.中值绝对误差（Median absolute error）：中值绝对误差特别有趣，因为它对异常值很健壮。损失是通过计算目标和预测之间的所有绝对差异的中值来计算的。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. from sklearn.metrics **import** median\_absolute\_error
2. y\_true = [3, -0.5, 2, 7]
3. y\_pred = [2.5, 0.0, 2, 8]
4. median\_absolute\_error(y\_true, y\_pred)

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/50d8b493-9e31-4656-9612-2dbe3359513e.png)

22.R2方值：r2\_score函数计算R的值，即确定系数。它提供了一种衡量未来样本可能被模型预测的程度的方法。最好的分数是1.0，它可以是负的（因为模型可能会更糟）。一个总是预测y的期望值的常数模型，不管输入的特征是什么，都会得到R2的0.0。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. from sklearn.metrics **import** r2\_score
2. y\_true = [3, -0.5, 2, 7]
3. y\_pred = [2.5, 0.0, 2, 8]
4. r2\_score(y\_true, y\_pred)

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/3cd6f116-21f6-4da1-ae1b-0642ea15585d.png)

聚类指标

23.兰德指数（adjusted\_rand\_score）

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. from sklearn.metrics.cluster **import** adjusted\_rand\_score
2. adjusted\_rand\_score([0, 0, 1, 1], [0, 0, 1, 1])

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/932aec83-4377-468e-a568-1ac1e484c22c.png)

ARI是一个对称的测量方法：adjusted\_rand\_score(a, b) == adjusted\_rand\_score(b, a)

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. adjusted\_rand\_score([0, 0, 1, 1], [1, 1, 0, 0])

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/eeebf93a-11a5-452f-badc-e293e2fe7a8b.png)

如果类成员在不同的集群中完全分开，那么分配是完全不完整的，因此ARI非常低

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. adjusted\_rand\_score([0, 0, 0, 0], [0, 1, 2, 3])

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/7604105b-ae41-4319-8188-bfb4050a4172.png)

24.调整互信息（Adjusted mutual information）

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. from sklearn.metrics.cluster **import** adjusted\_mutual\_info\_score
2. adjusted\_mutual\_info\_score([0, 0, 1, 1], [0, 0, 1, 1])

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/7e8b5364-e875-4a9d-af15-7275a3f7af12.png)

如果类成员在不同的集群中完全分裂，那么分配是完全不完整的，因此AMI非常低。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. adjusted\_mutual\_info\_score([0, 0, 0, 0], [0, 1, 2, 3])

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/c1e9a9da-a271-40e2-a9f8-a8f9b6d0559b.png)

25.轮廓系数（Silhouette Coefficient）：一组样本的轮廓系数作为每个样本的轮廓系数的平均值

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. from sklearn **import** metrics
2. from sklearn.metrics **import** pairwise\_distances
3. from sklearn **import** datasets
4. **import** numpy as np
5. from sklearn.cluster **import** KMeans
7. #导入iris数据
8. dataset = datasets.load\_iris()
9. X = dataset.data
10. y = dataset.target
12. #创建k-means聚类模型
13. kmeans\_model = KMeans(n\_clusters=3, random\_state=1).fit(X)
14. labels = kmeans\_model.labels\_
16. #模型评估：在正常使用中，轮廓系数应用于集群分析的结果。
17. metrics.silhouette\_score(X, labels, metric='euclidean')

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/b1a3c2a9-1302-4608-bf8f-3f6bc426760d.png)

26.Homogeneity, completeness and V-measure

同质性homogeneity：每个群集只包含单个类的成员。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. from sklearn **import** metrics
2. labels\_true = [0, 0, 0, 1, 1, 1]
3. labels\_pred = [0, 0, 1, 1, 2, 2]
5. metrics.homogeneity\_score(labels\_true, labels\_pred)

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/a2534b79-e99b-4302-8844-29cec3f1d0e0.png)

完整性completeness：给定类的所有成员都分配给同一个群集。

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. metrics.completeness\_score(labels\_true, labels\_pred)

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/7599fda9-77cf-4986-bb2a-fa92ca7edaf5.png)

两者的调和平均V-measure：

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. metrics.v\_measure\_score(labels\_true, labels\_pred)

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/e15ab565-b026-42b2-a6d7-f39f5ba91e4d.png)

27.Fowlkes-Mallows指数（fowlkes\_mallows\_score）：是针对训练集和验证集数据之间求得的查全率和查准率的几何平均值

[view plain](https://www.ipieuvre.com/e/411/311/9539) [copy](https://www.ipieuvre.com/e/411/311/9539)

1. from sklearn **import** metrics
2. labels\_true = [0, 0, 0, 1, 1, 1]
3. labels\_pred = [0, 0, 1, 1, 2, 2]
4. metrics.fowlkes\_mallows\_score(labels\_true, labels\_pred)

[](https://www.ipieuvre.com/doc/exper/36a741ee-91ad-11e9-beeb-00215ec892f4/img/02ff5dc1-c491-4539-bcaf-d05ea5d6234b.png)