# 基本用法

1. 引入包

from sklearn.cluster import KMeans

1. 首先输入得到需要的数据集合

X为[[1 1] [2 3] [3 4]] 里面一个数据为一个向量

X必须为np.array()类型

1. 生成 kmean\_model n\_clusters表示需要聚为几个类 .fix(X)开始聚类

kmeans\_model = KMeans(n\_clusters=tests).fit(X)

1. 得到聚类结果 kmeans\_model.labels\_ 里面存的是没一个数据的类标签

for index,value in enumerate(kmeans\_model.labels\_):

# 多个K值 只用加个for 循环

# 属性

## kmeans\_model.cluster\_centers\_

集群中心的坐标

[[ 8.66666667 2. ]

[ 1.75 2. ]

[ 5.4 6.8 ]

[ 7. 1. ]]

n\_clusters 行 n\_features 列

## kmeans\_model.labels\_ :

Labels of each point 每一类所属的标签 从0开始 [1 1 1 1 2 2 2 2 2 3 0 0 3 0]

# 评价指标

## elbow method

用来选着聚类的类个数K的值

下降最大的是最好的

## silhouette coefficient

值越大越好

引包 from sklearn import metrics

Sc=metrics.silhouette\_score(X,kmeans\_model.labels\_,metric='euclidean')) # Euclidean 表示欧氏距离