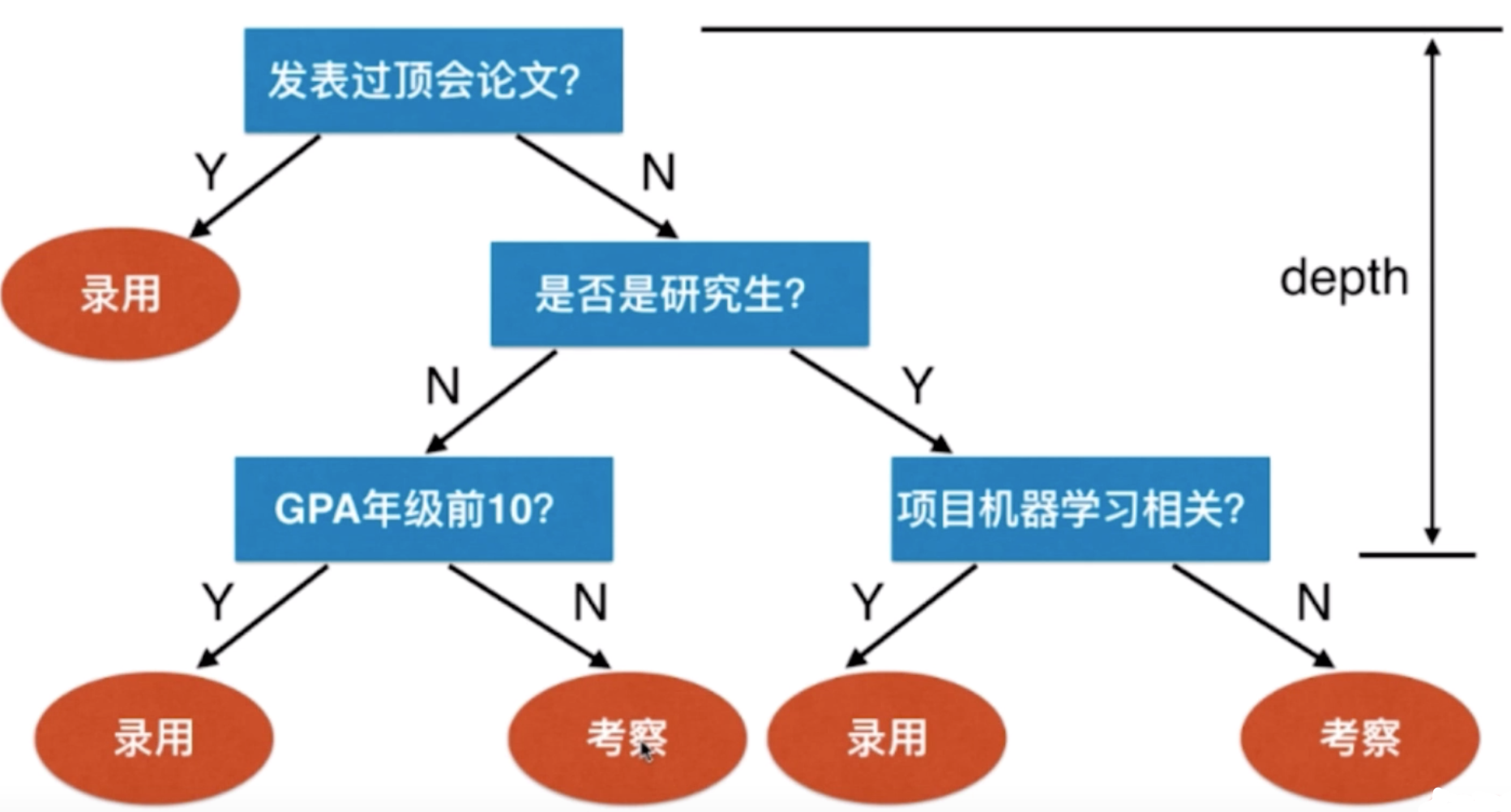
# 决策树

* 概述



**决策树的特点**

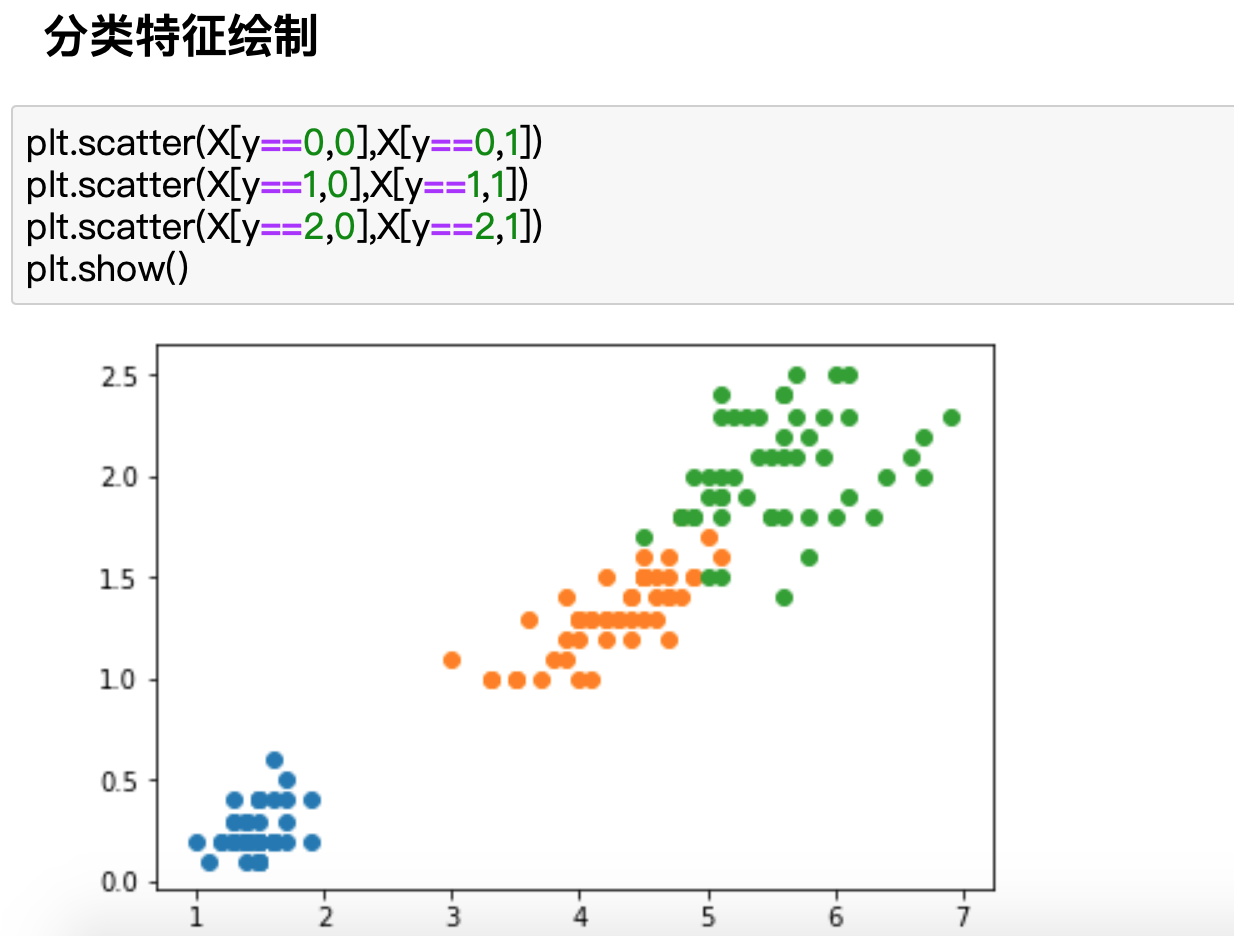
**1.非参数学习方法**

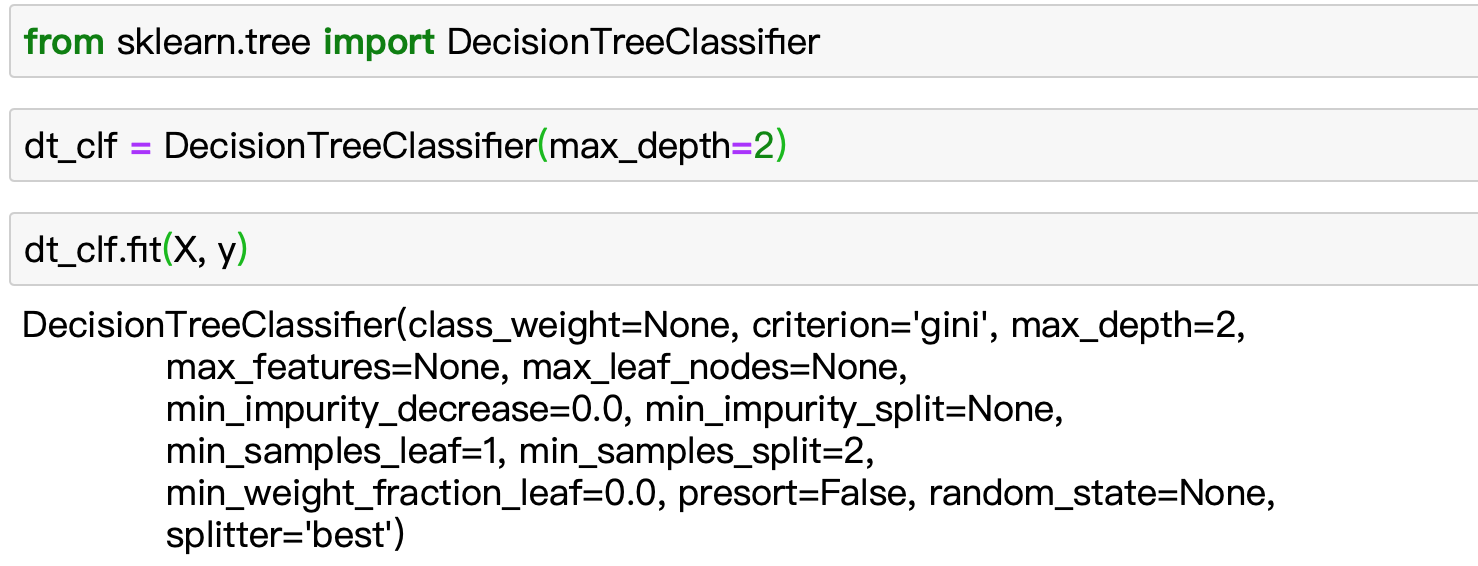
**2.解决分类问题，天然支持多分类问题**

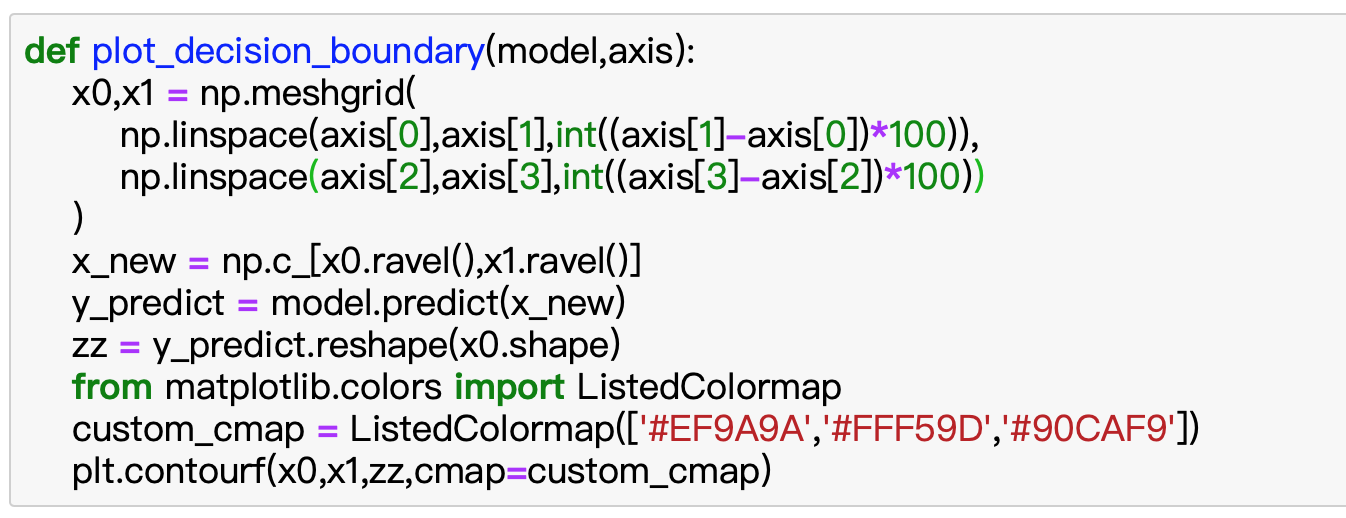
**3.具有很好的解释型**

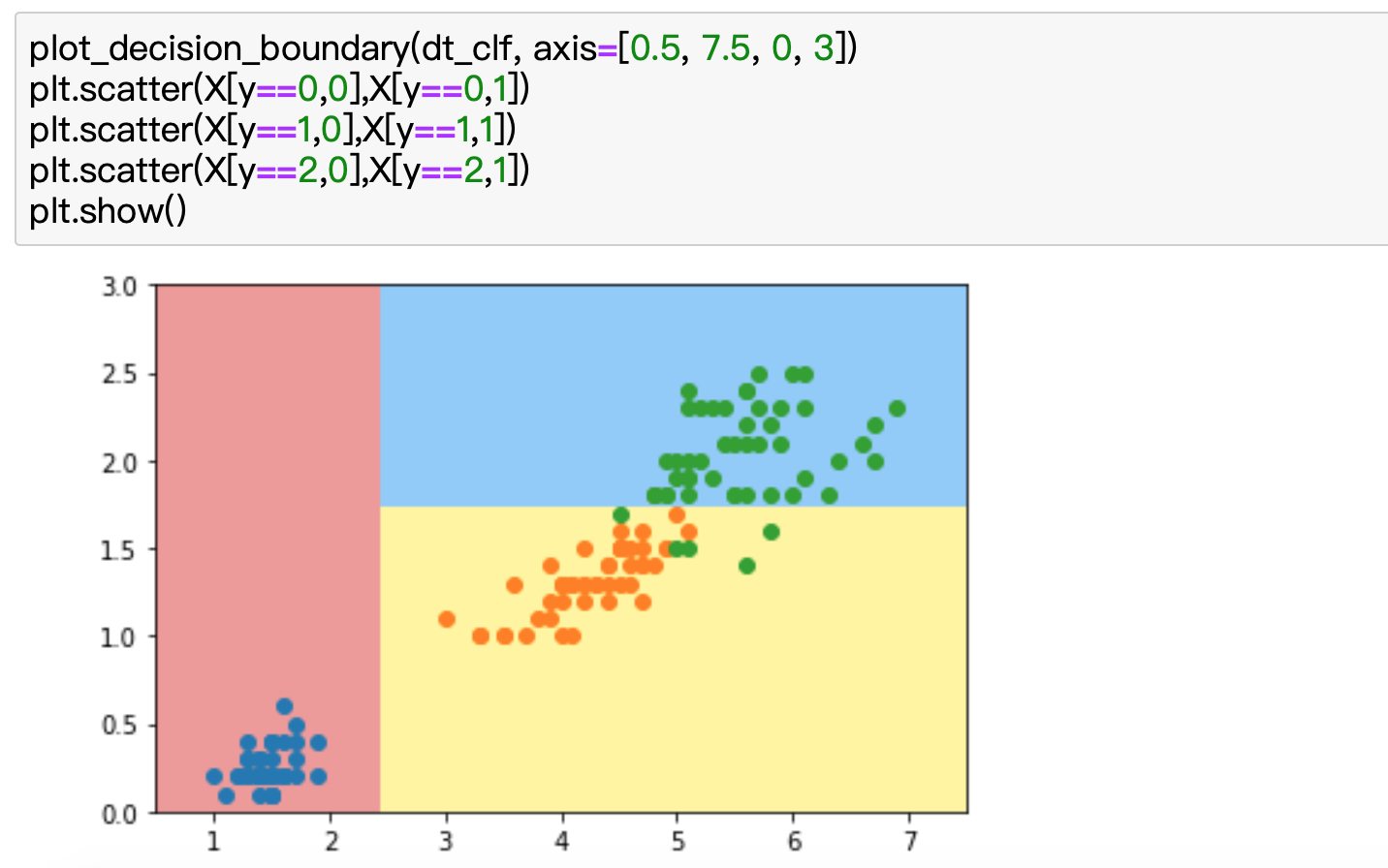
* Sklearn中的决策树实现

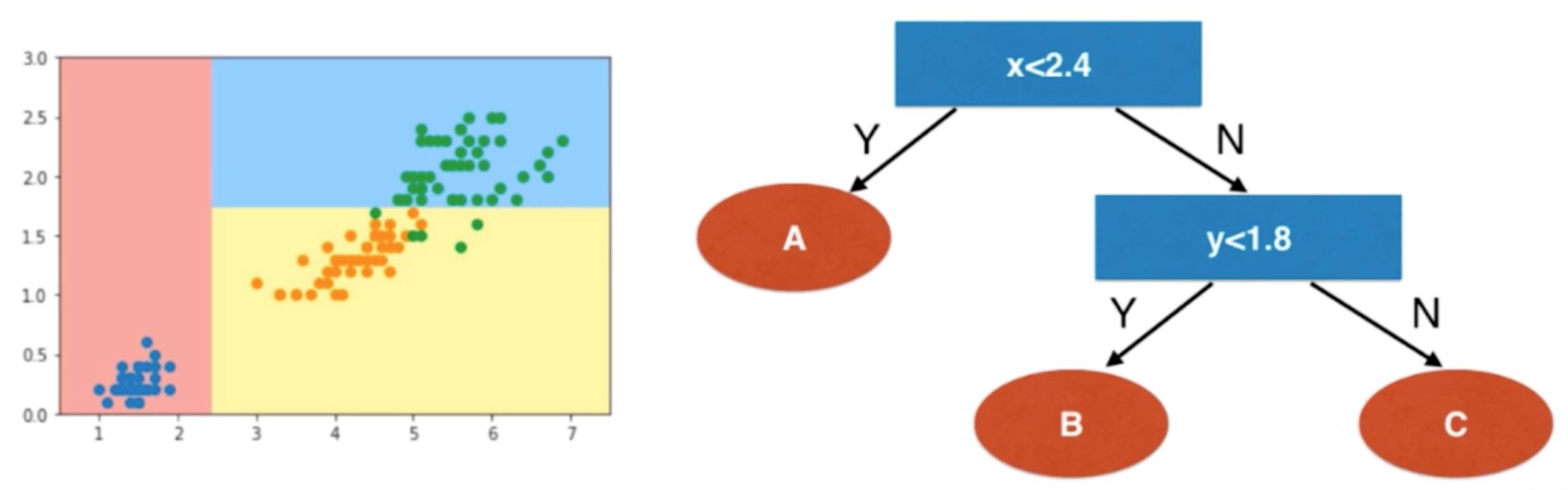












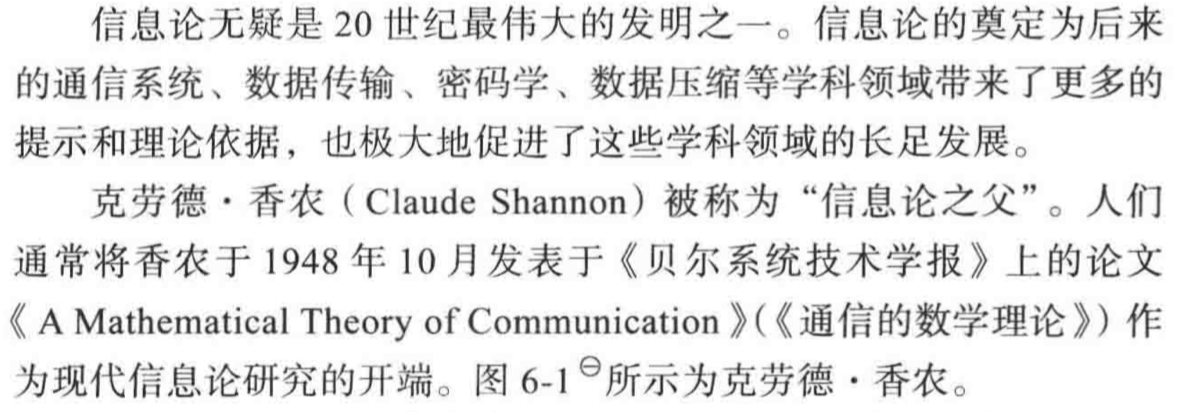
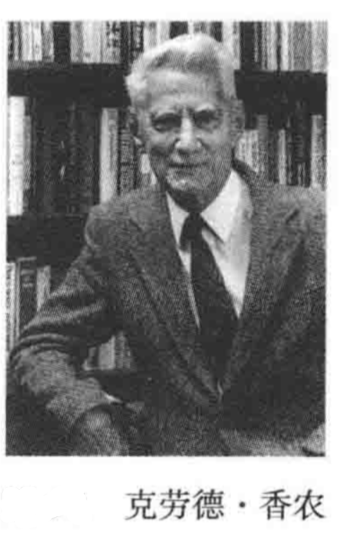
* 信息熵

需要解决的问题：

1.使用哪个特征进行分类的划分

2.在划分时，使用多大的数值进行

**信息论**

**什么是数据**

“承载了信息的东西”才是数据。

不管是刻在石头上的画，还是歪歪扭扭写出的字迹，只要它表达一些确定

的含义，那么它就是数据。

**什么是信息**

1948年，数学家香农在题为《通信的数学理论》的论文中指出：

“信息是用来消除随机不定性的东西”。

例如

两句话向不了解你的人介绍你的年龄

你：我今年33岁

你：我明年34岁

上述两句话，哪句是信息

**信息量的计算**

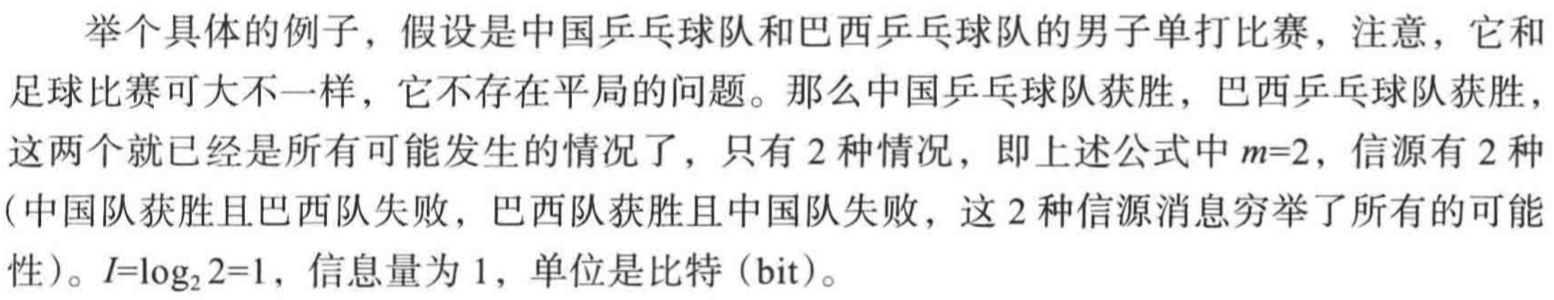
在平时听到一些八卦信息时，有些信息会觉得信息量好大。

在信息论中，对信息量有确定解释并且可量化计算。

1982年，由哈特莱首先提出信息量的计算公式

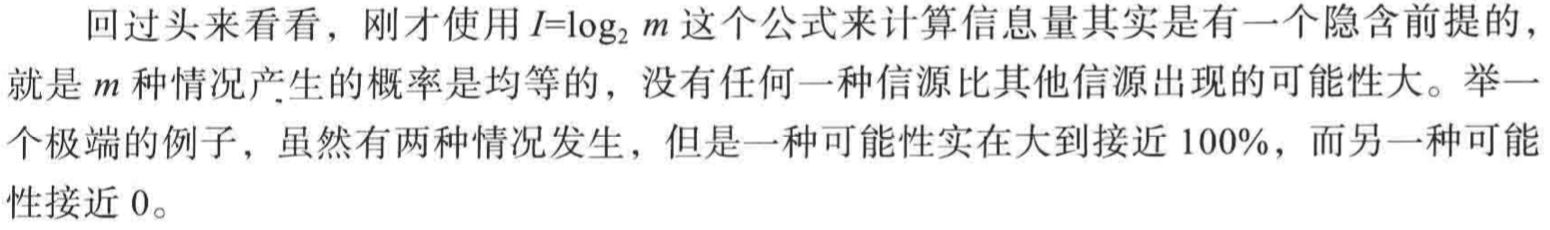
若信源有m种消息，且每个消息以相等可能产生，则该信源的信息量为

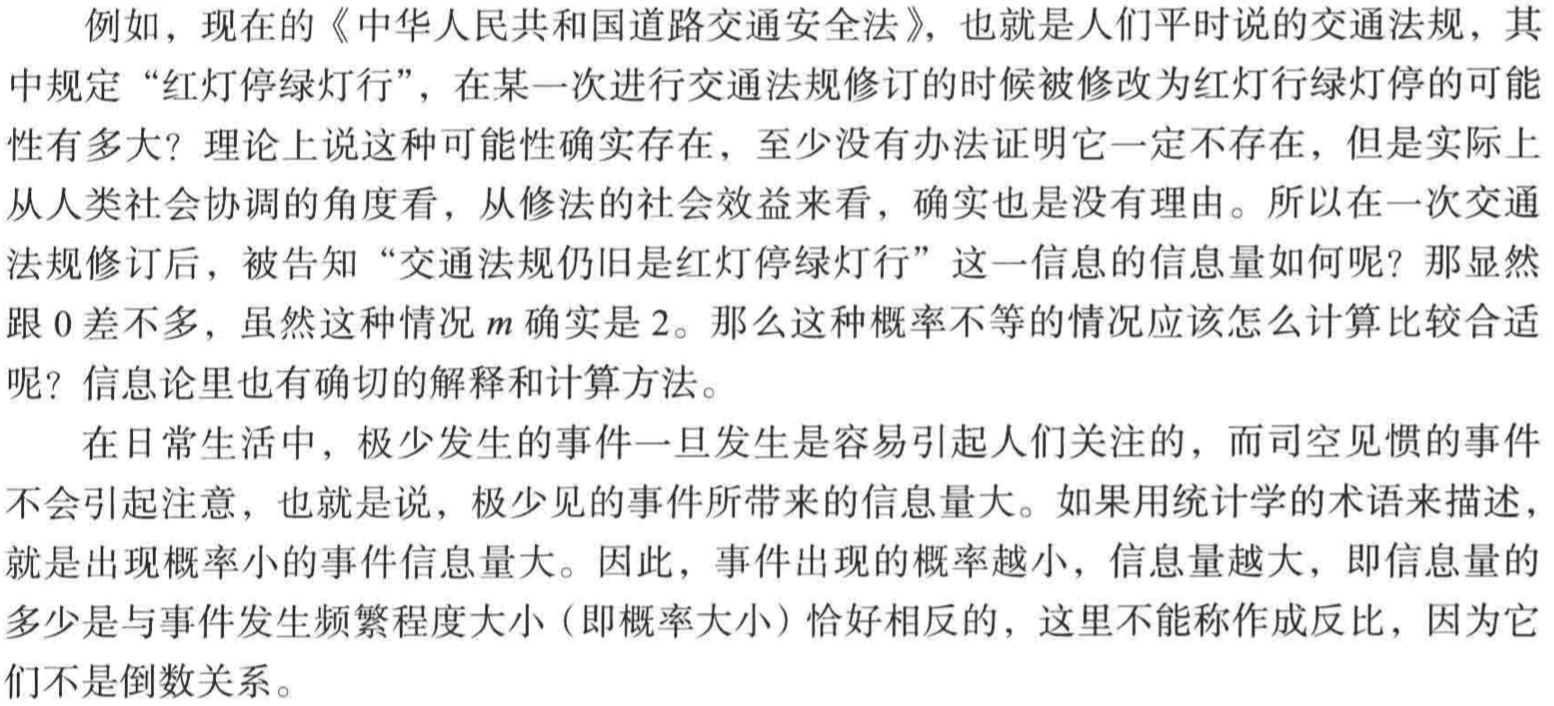




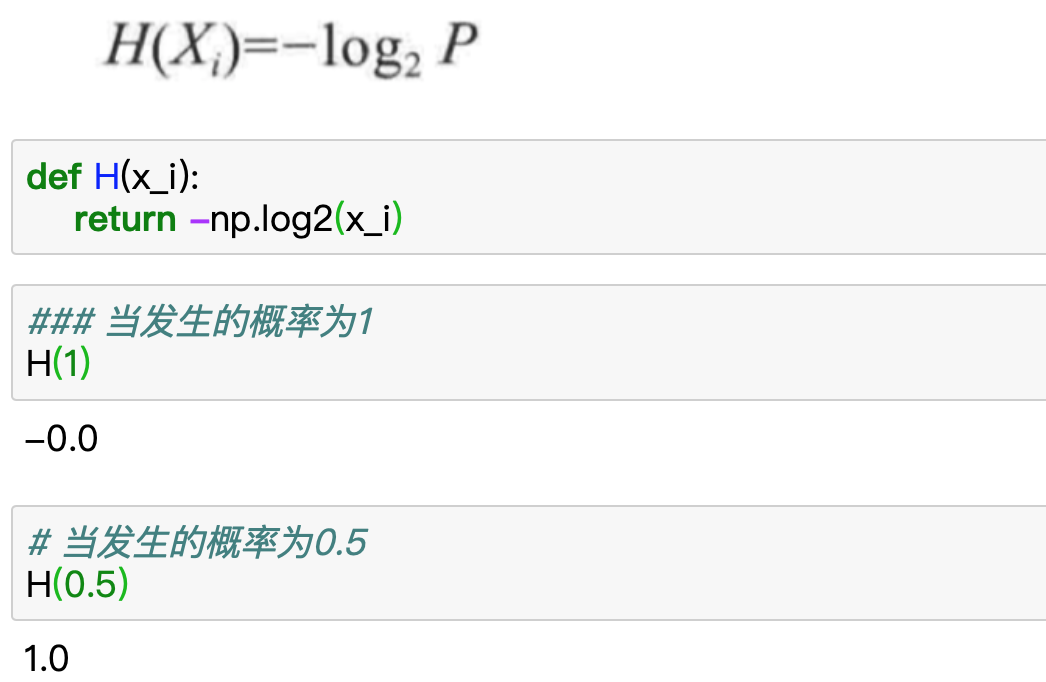
**按照理论，消除不确定程度越高，信息量就越大。消除的不确定程度越低，信息**

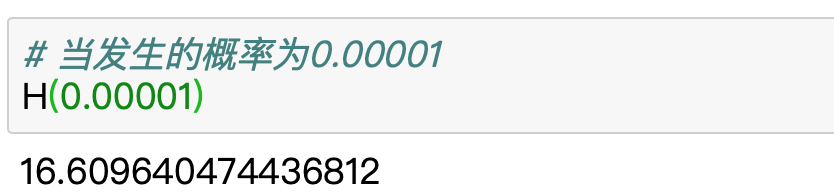
**量越小。**

****

****

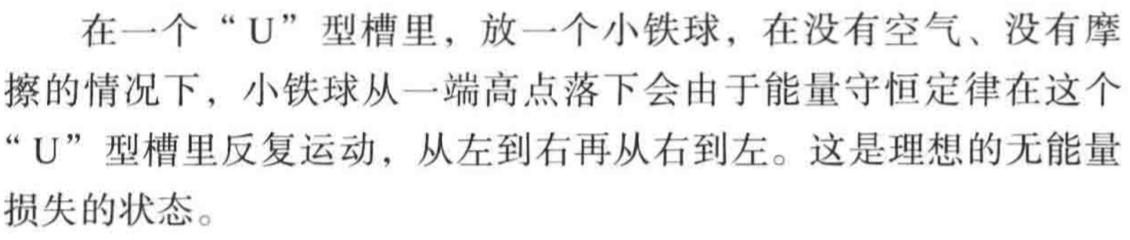
****

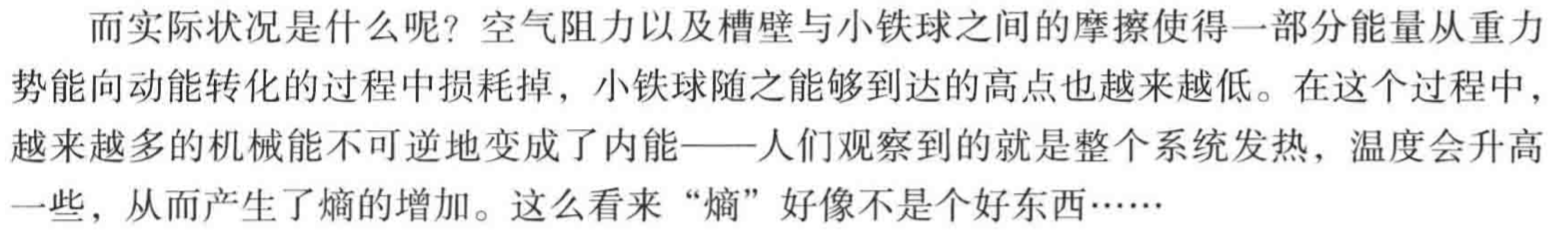
****

****

**熵**

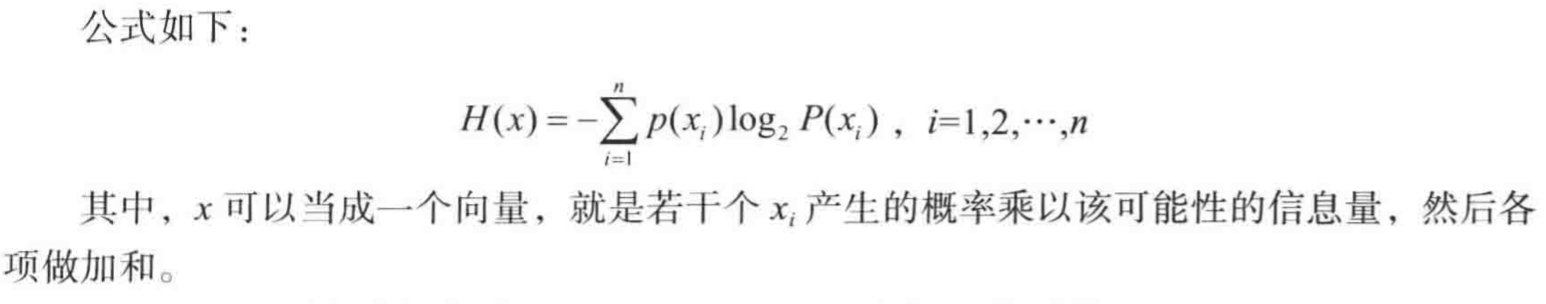
**热力熵**

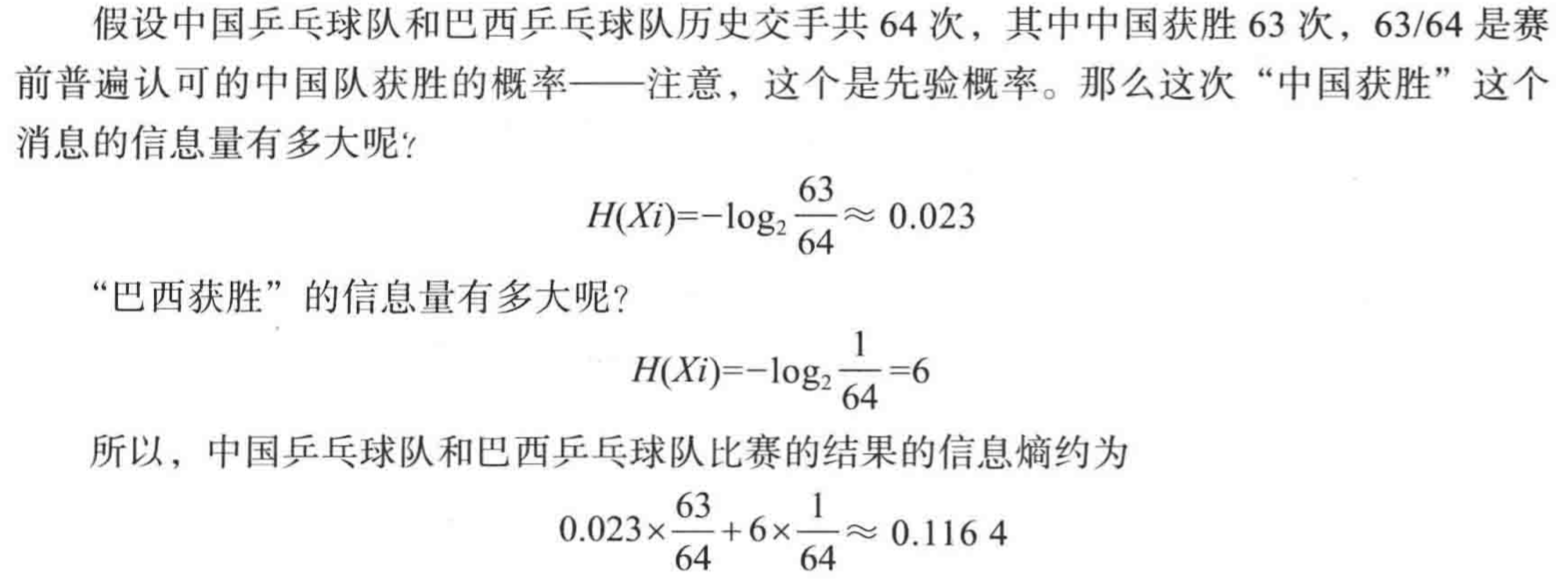
****

****

**信息熵**

信息熵，可以认为是信息杂乱程度的量化。





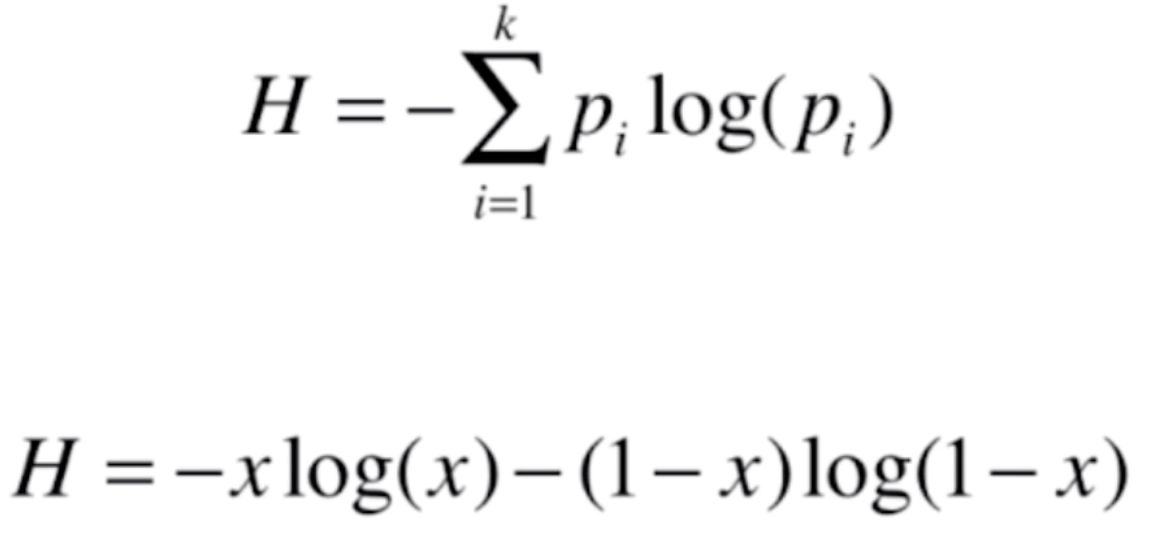
**信息熵的结论**

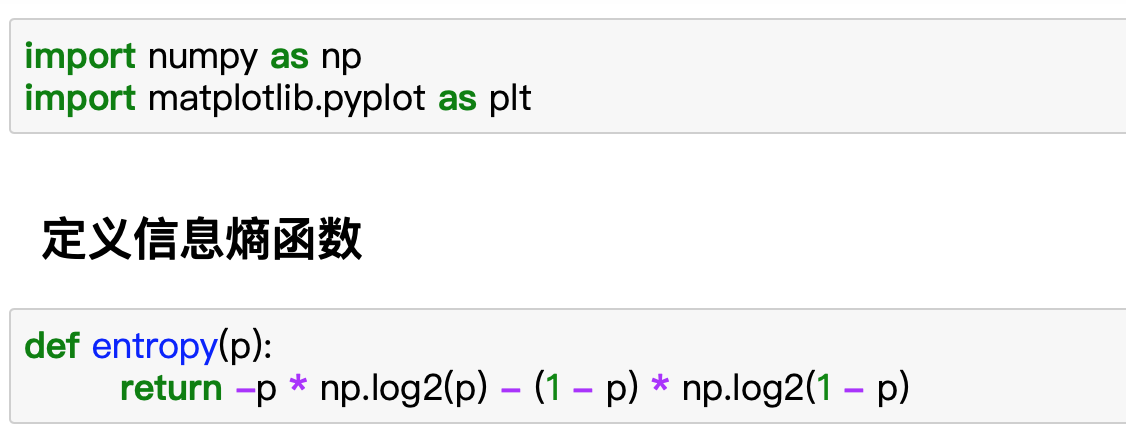
1.信息越确定，越单一，信息熵越小

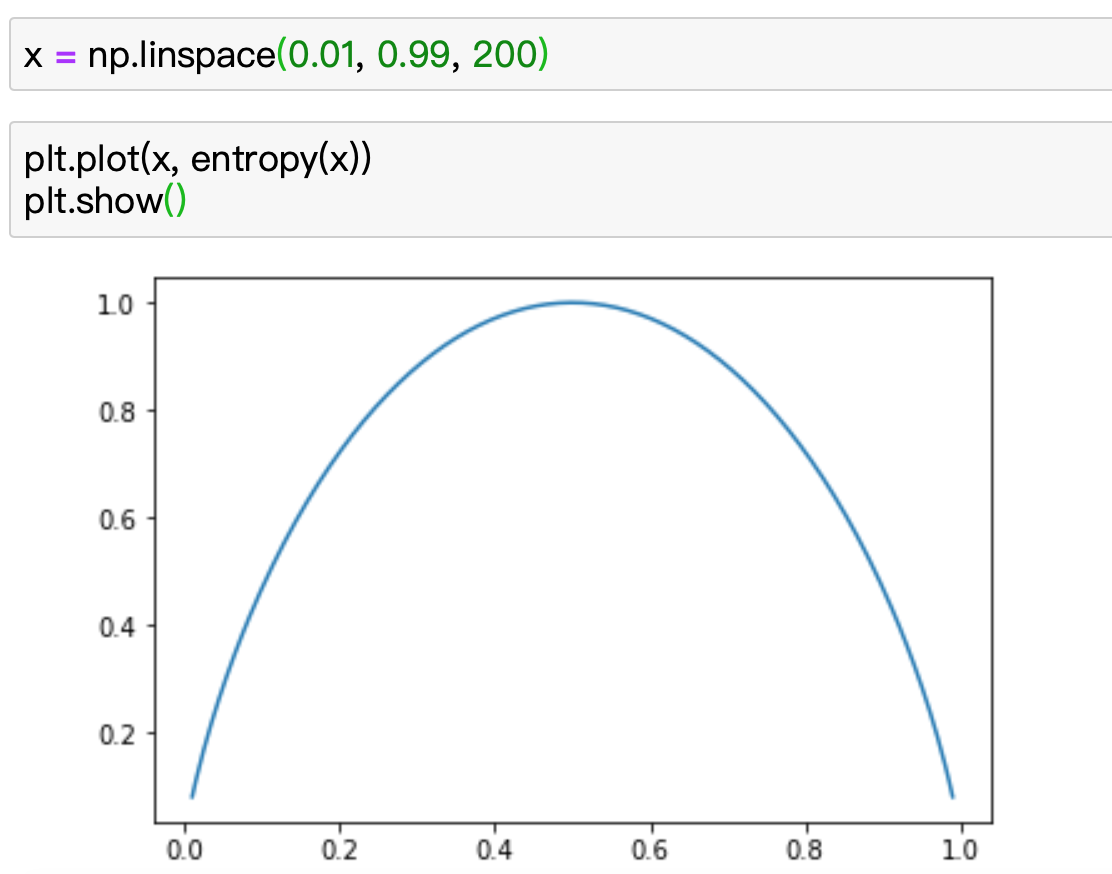
2.信息越不确定，越混乱，信息熵越大

* 信息熵实现

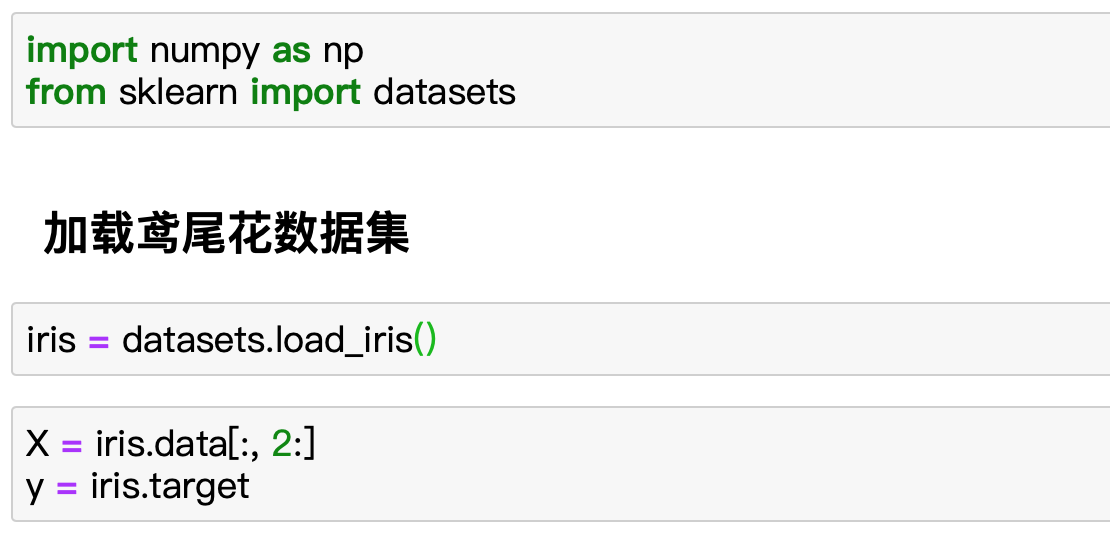
以二分类问题为例

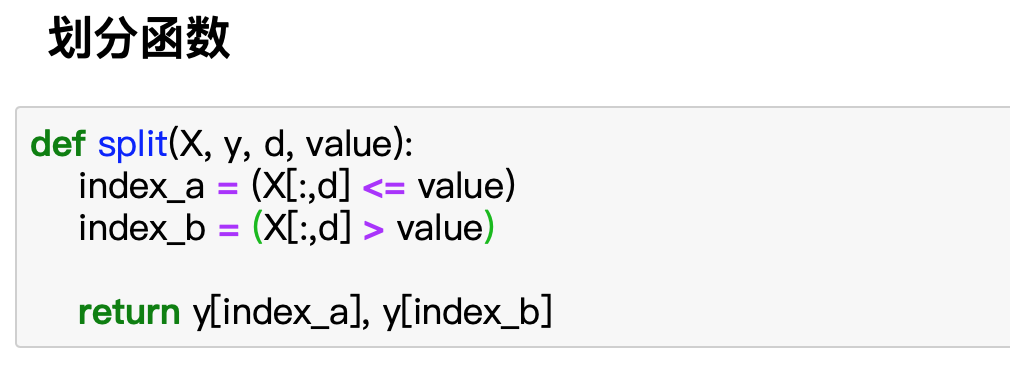






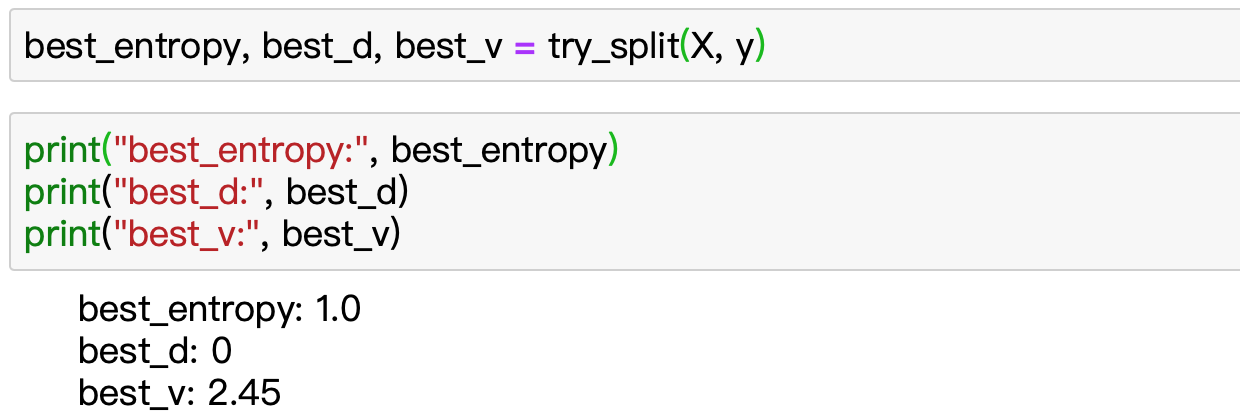
* 使用信息熵寻找最优划分



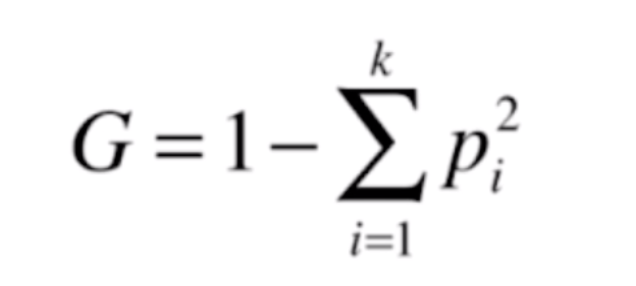




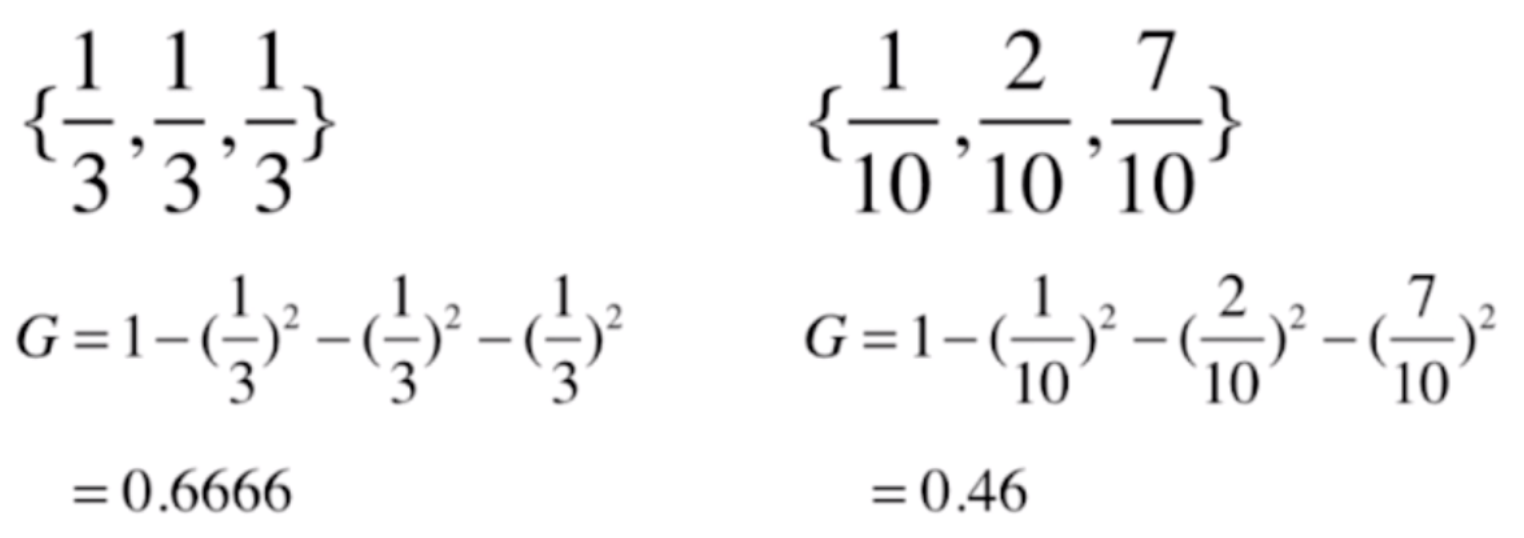


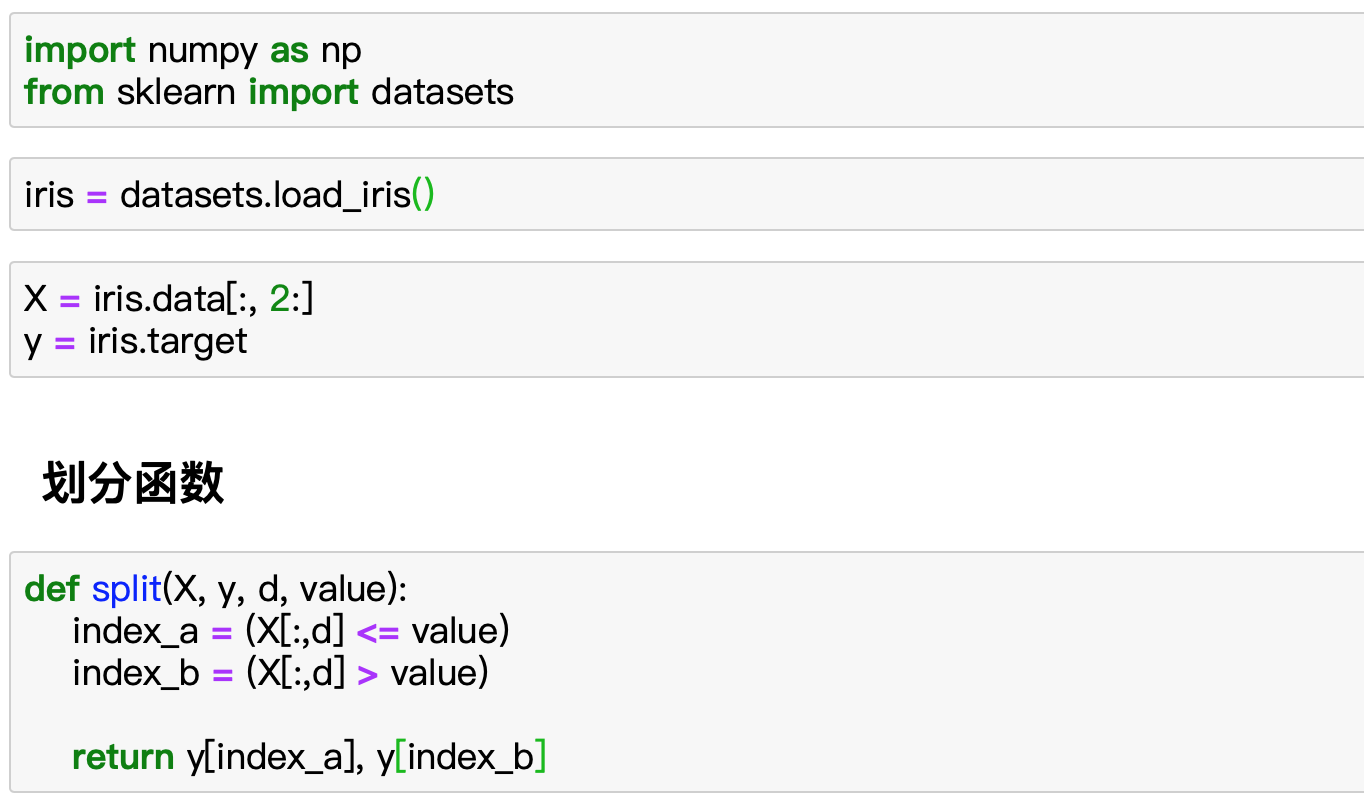


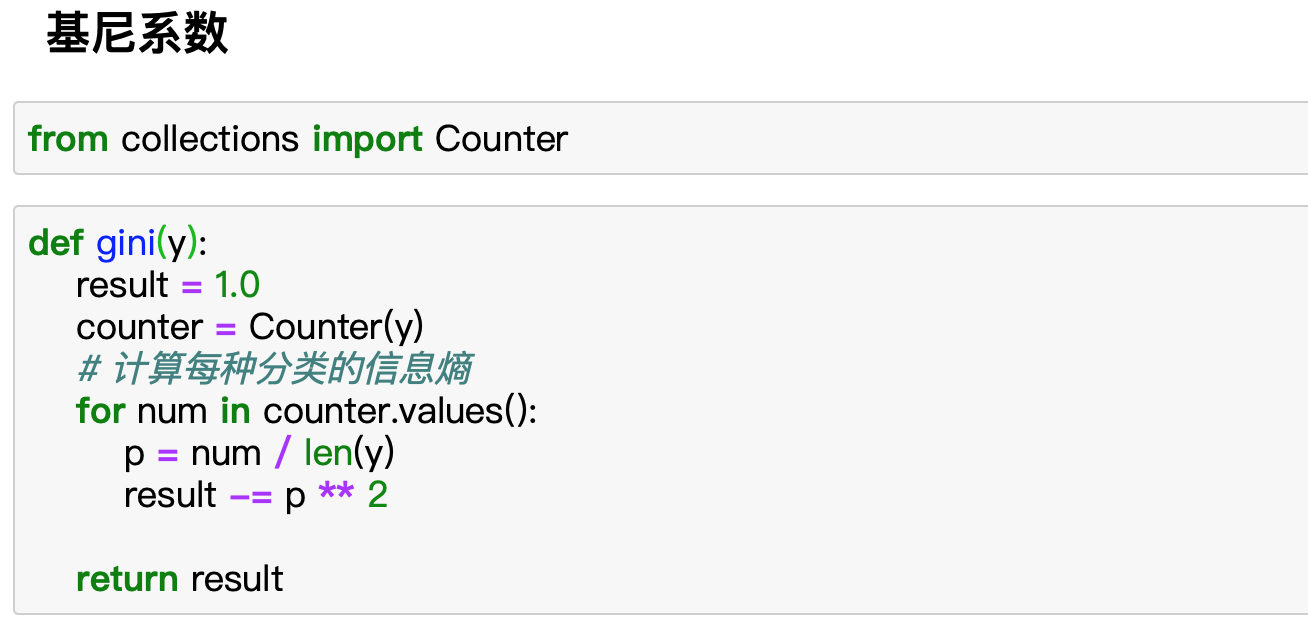
* 基尼系数



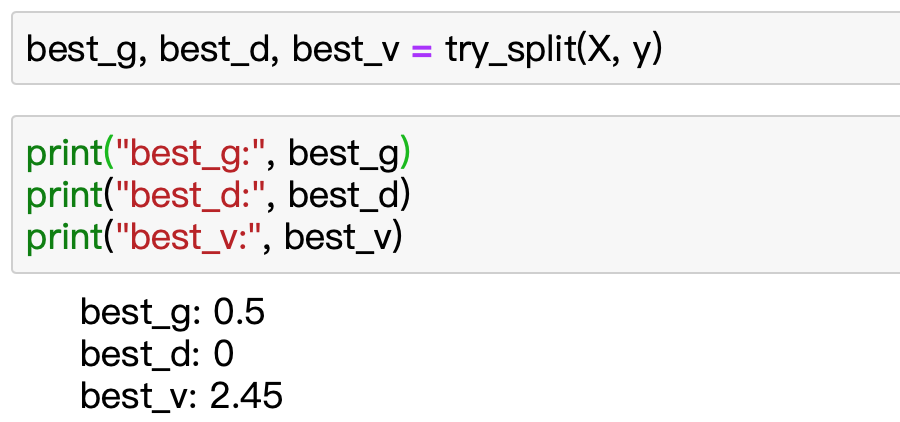
**例**



****

****

****

****

**信息熵 vs 基尼系数**

**1.信息熵的计算比基尼系数稍慢**

**2.sklearn中默认为基尼系数**

**3.大多时候两者没有特别的优劣势**

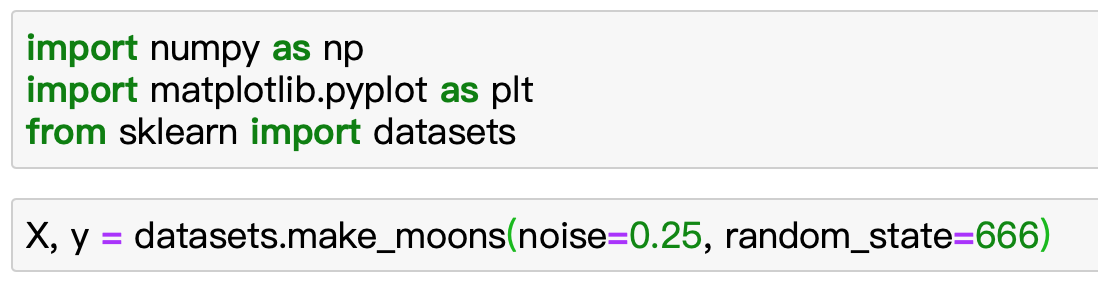
* CART与超参数

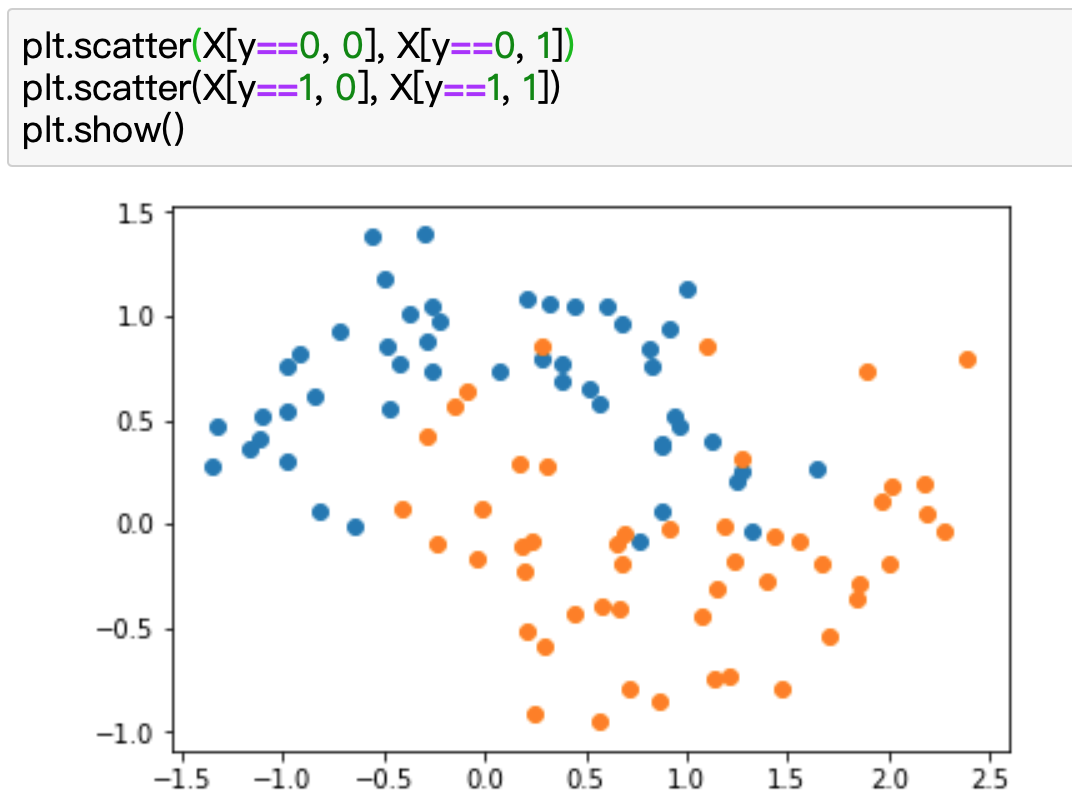
**决策树又有另一个名称叫做CART(Classfication And Regression Tree)**

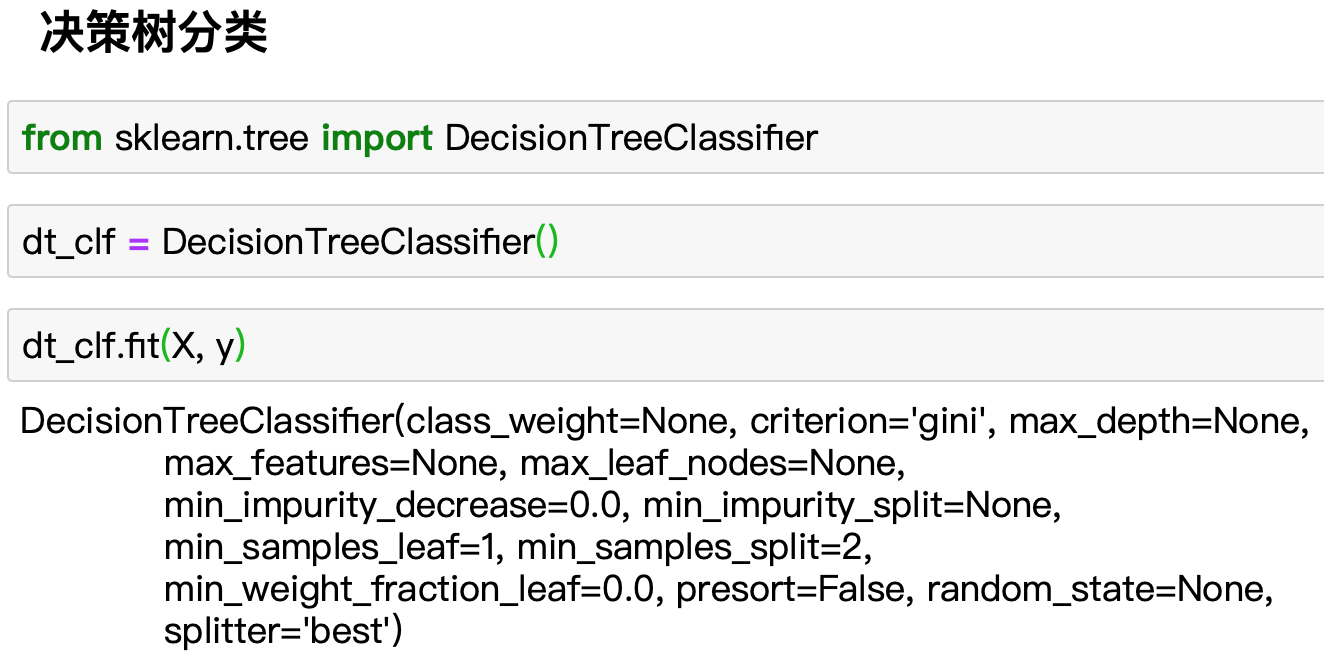
**常用构建决策树的算法：ID3、C4.5、C5.0**

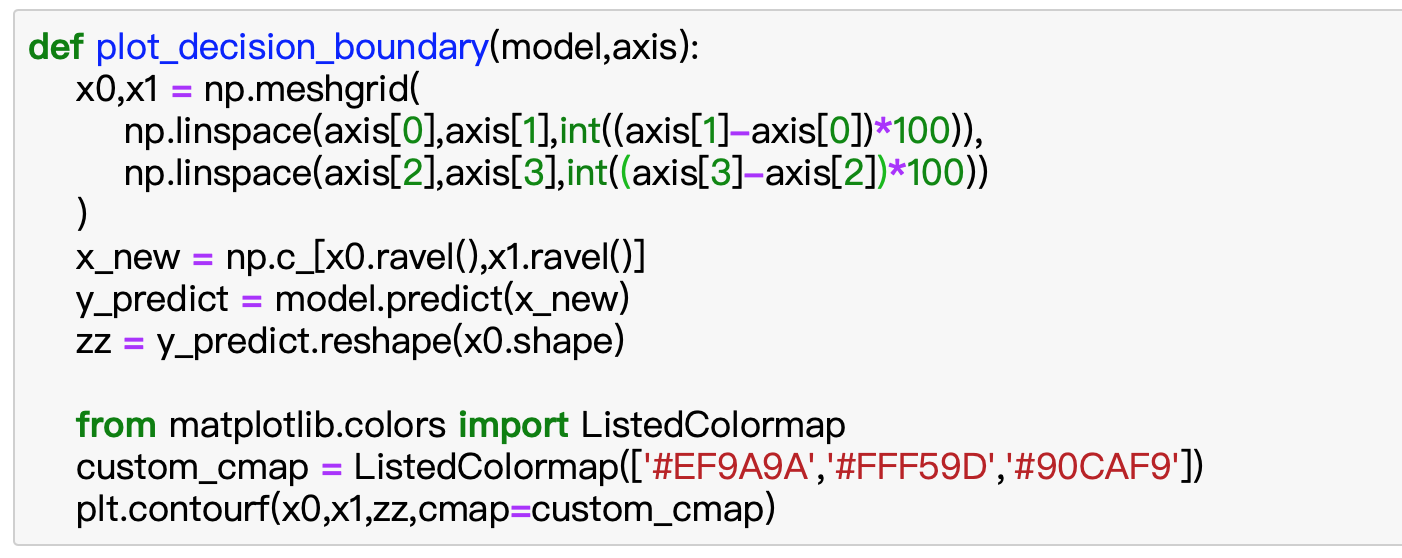
**决策树过拟合解决方案**

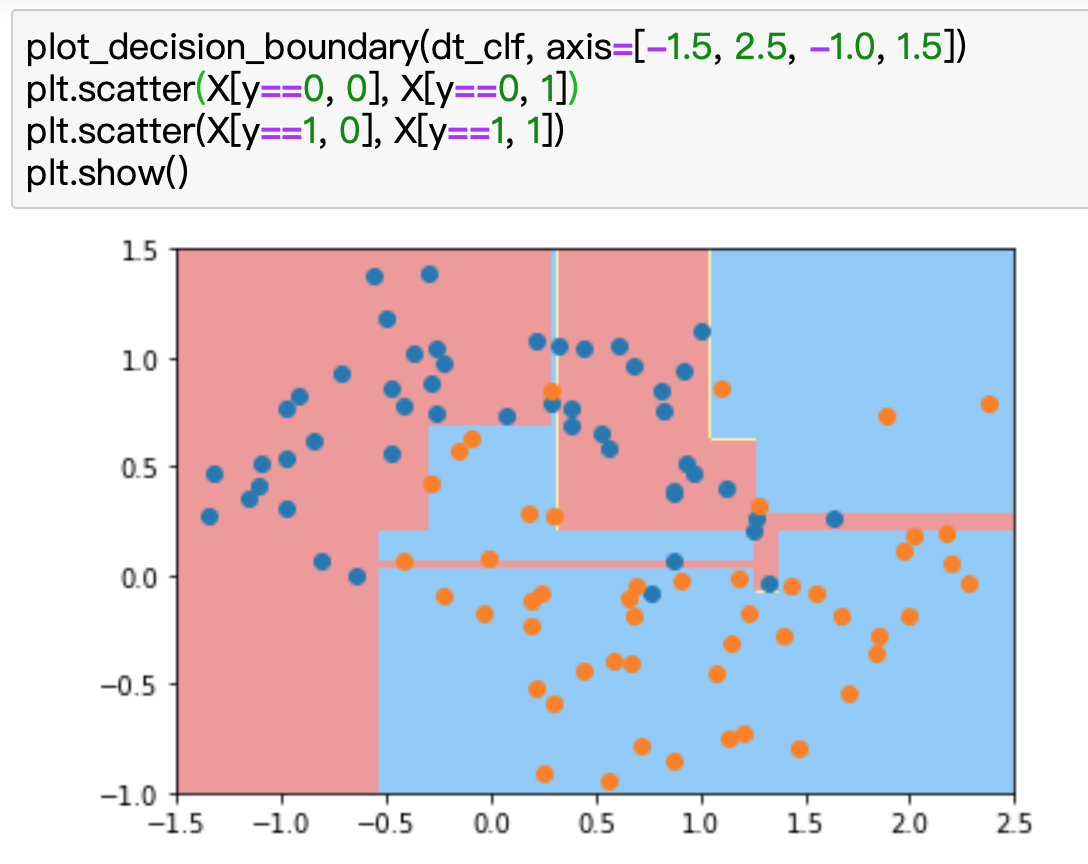
**树剪枝：降低复杂度、解决过拟合**

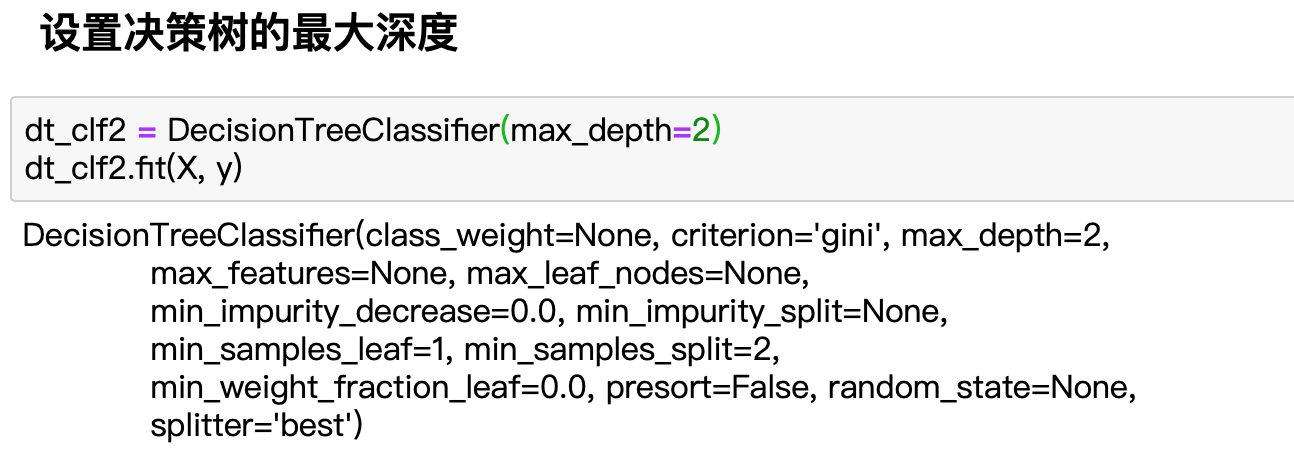
****

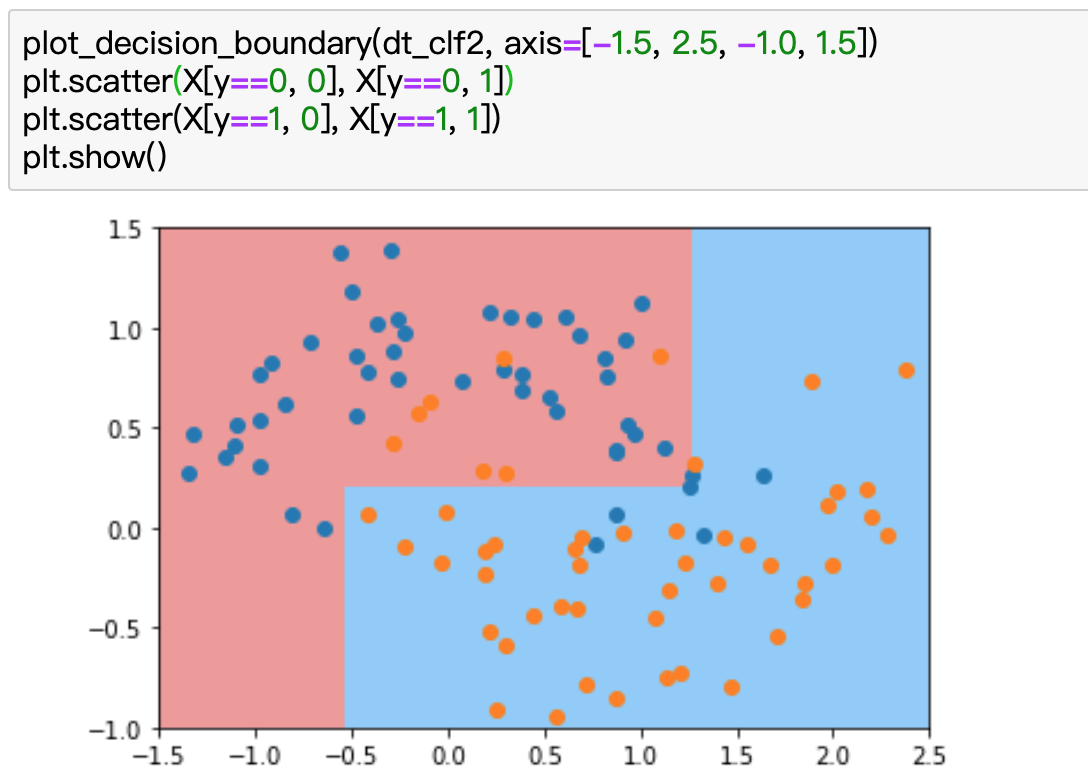
****

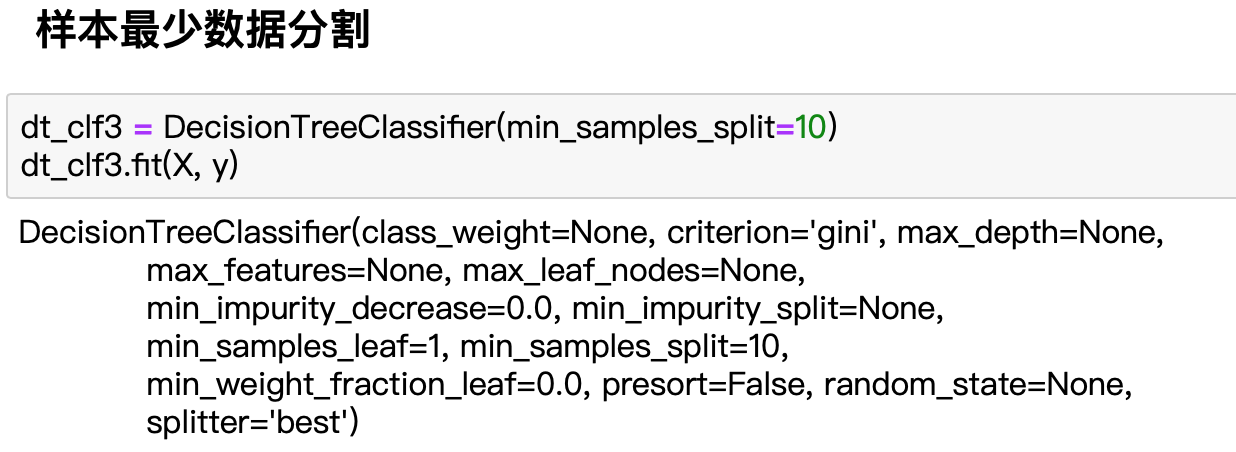
****

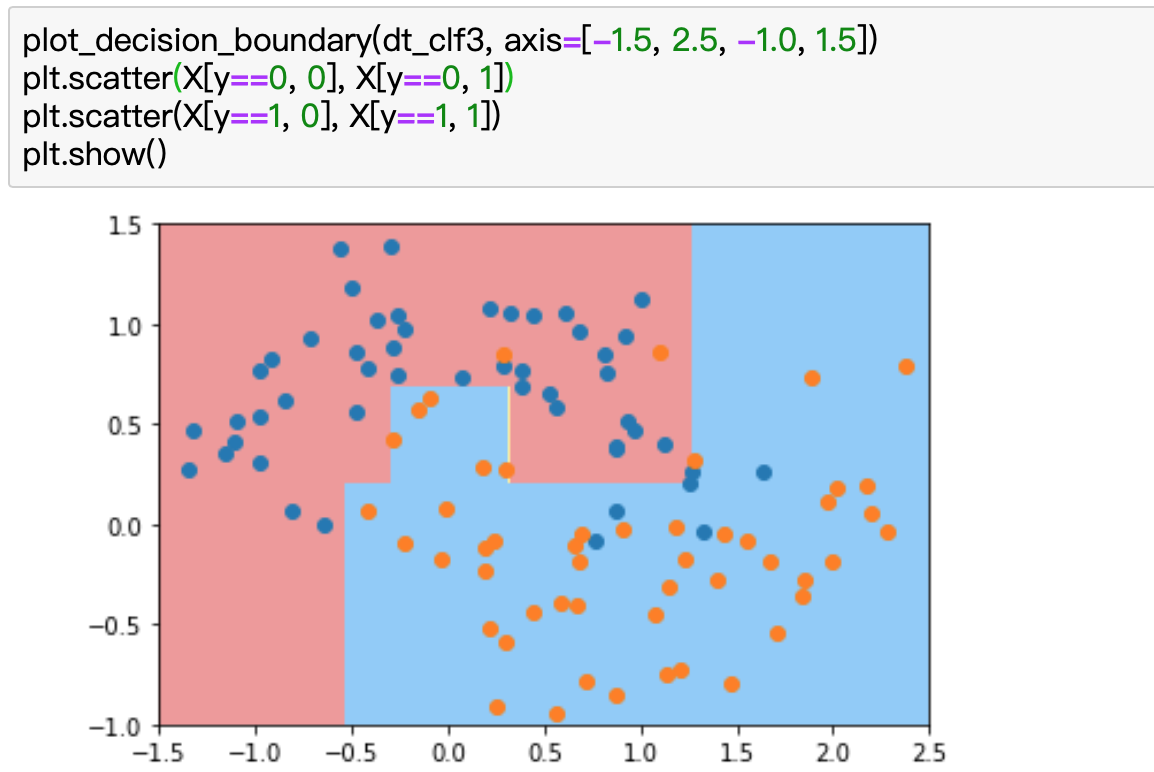
****

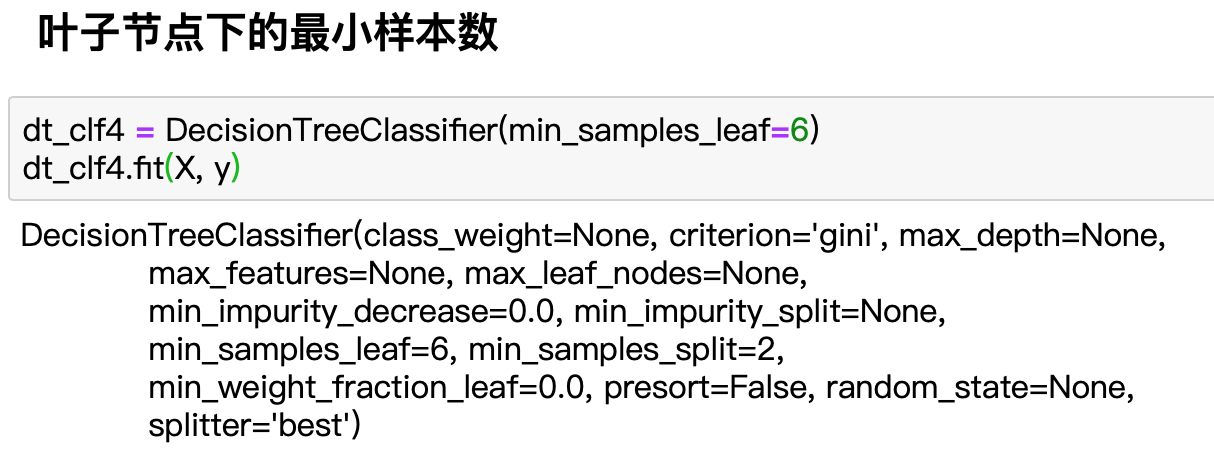
****

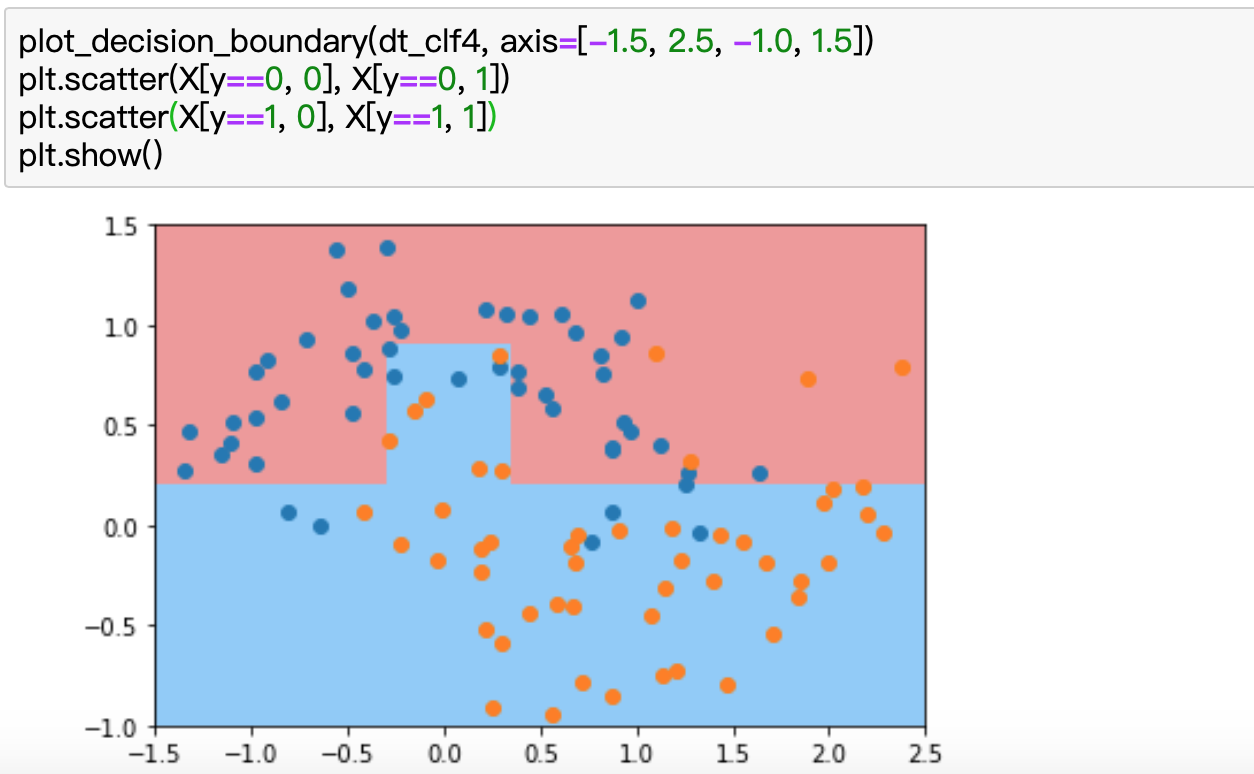
****

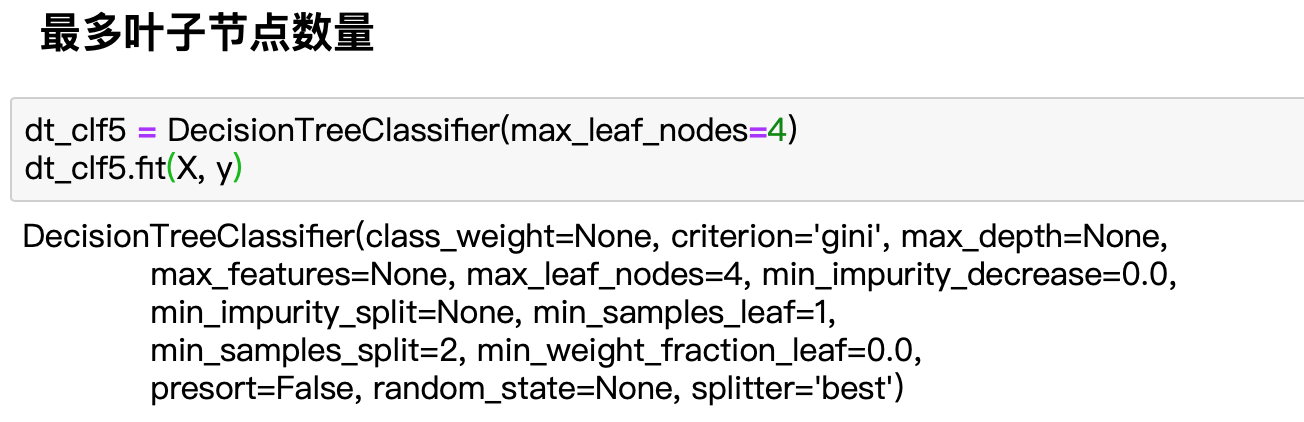
****

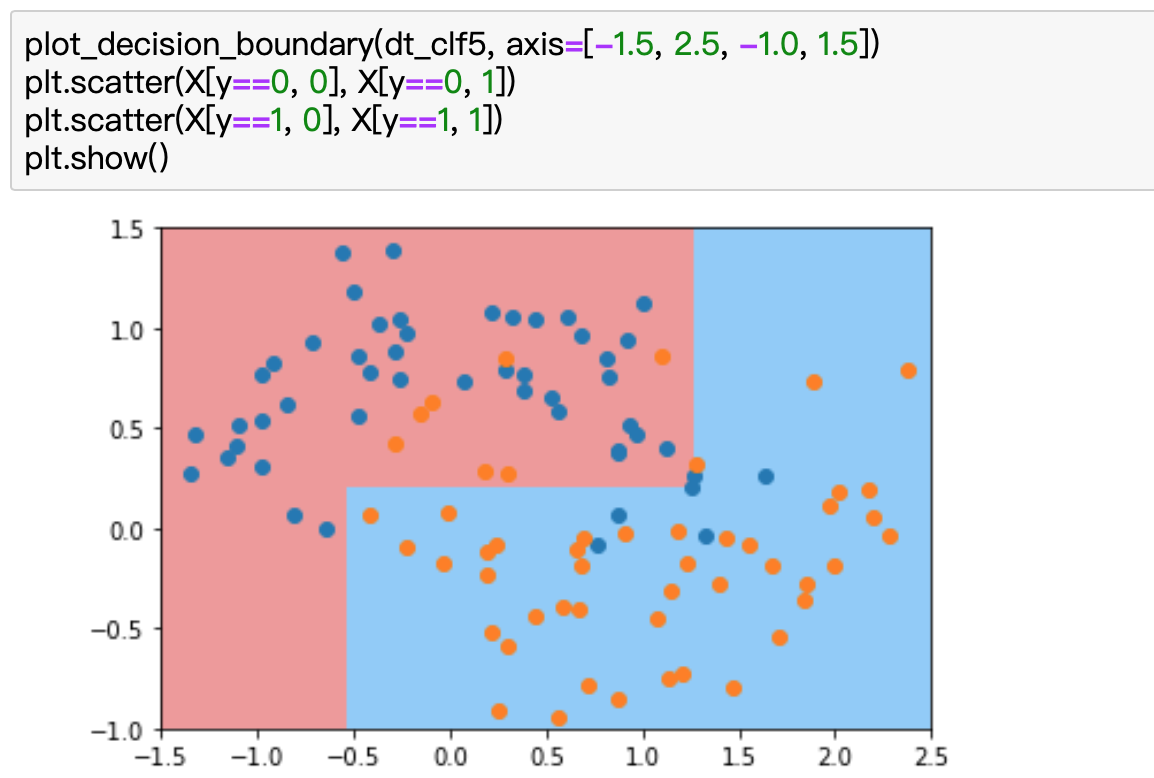
****

****

****

****

****

****