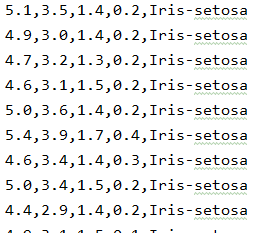
# 机器学习——SVM Python实战

## 数据集准备

本文用的数据集为Iris.data可从UCI数据库中下载，http://archive.ics.uci.edu/ml/datasets/Iris

Iris.data的数据格式如下：共5列，前4列为样本特征，第5列为类别，分别有三种类别Iris-setosa/ Iris-versicolor/Iris-virginica。

NOTE1：第五列的字符串数据最好转化成数字量。



【函数说明】

@set\_module(**'numpy'**)  
**def** loadtxt(fname, dtype=float, comments=**'#'**, delimiter=**None**,  
 converters=**None**, skiprows=0, usecols=**None**, unpack=**False**,  
 ndmin=0, encoding=**'bytes'**, max\_rows=**None**):

函数功能：从text文档中加载数据（数据的每一行必须维度相同）

参数解释：

*delimiter : str, optional（设置分隔号）  
 The string used to separate values. For backwards compatibility, byte strings will be decoded as 'latin1'. The default is whitespace.  
converters : dict, optional转换器（字典型）通过字典将列值映射至一个函数。此函数将会列值解析为期望值。此处使用此函数对数据集中的字符值转化成数值。  
 A dictionary mapping column number to a function that will parse the column string into the desired value. E.g., if column 0 is a date string: ``converters = {0: datestr2num}``. Converters can also be used to provide a default value for missing data (but see also `genfromtxt`): ``converters = {3: lambda s: float(s.strip() or 0)}`` Default: None.*

## 导入SVM模块并划分数据与标签

Python中的sklearn库集成了SVM算法。

【函数说明】

*train\_test\_split(train\_data,train\_label,test\_size=数字, random\_state=0)*

*函数功能：随机划分训练集与测试集。*

*参数解释：*

*train\_data：所要划分的样本特征集*

*train\_label：所要划分的样本类别*

*test\_size：样本占比，如果是整数的话就是样本的数量。*

*test\_size:测试样本占比。默认情况下，该值设置为0.25。只有train\_size没有指定时，它将保持0.25，否则它将补充指定的train\_size，例如train\_size=0.6,则test\_size默认为0.4。*

*train\_size:训练样本占比。*

*random\_state：是随机数的种子。随机数种子：其实就是该组随机数的编号，在需要重复试验的时候，保证得到一组一样的随机数。比如你每次都填1，其他参数一样的情况下你得到的随机数组是一样的。但填0或不填，每次都会不一样。随机数的产生取决于种子，随机数和种子之间的关系遵从以下两个规则：种子不同，产生不同的随机数；种子相同，即使实例不同也产生相同的随机数。*