**CountVectorizer-计数词袋模型**

[**https://blog.csdn.net/liuerin/article/details/91492708**](https://blog.csdn.net/liuerin/article/details/91492708)

from sklearn.feature\_extraction.text import CountVectorizer

CV=CountVectorizer()

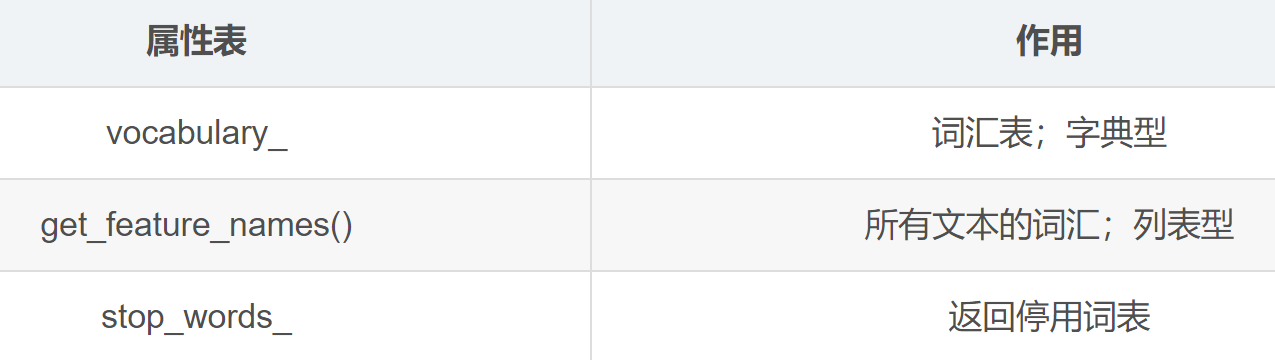
wordmatrix=CV.fit\_transform(text\_data) #此处返回一个密集矩阵，需要通过toarray()转化为稀疏矩阵

**#获取单词矩阵列名**

CV.get\_feature\_names()

**#转化为 pandas.DataFrame**

wordmatrix=pd.DataFrame(wordmatrix.toarray(),columns=CV.get\_feature\_names())



**TfIdf词袋模型**

from sklearn.feature\_extraction.text import TfidfVectorizer

TV= TfidfVectorizer()

wordmatrix=TV.fit\_transform(text\_data) #此处返回一个密集矩阵，需要通过toarray()转化为稀疏矩阵

**#获取单词矩阵列名**

TV.get\_feature\_names()

**#转化为 pandas.DataFrame**

wordmatrix=pd.DataFrame(wordmatrix.toarray(),columns=TV.get\_feature\_names())