

資料排序

常用資料型態的資料排序：

List
單欄
<code>sorted(list)</code> 指定鍵值(2D)： 1. <code>sorted(list, key = lambda s: s[2])</code> 2. from operator import itemgetter <code>sorted(list, key = itemgetter(2))</code>
<code>list.sort()</code> 指定鍵值(2D)： 1. <code>list.sort(key = lambda s: s[2])</code> 2. from operator import itemgetter <code>list.sort(key = itemgetter(2))</code>
多欄
from operator import itemgetter 1. <code>sorted(list, key = itemgetter(1, 2))</code> 2. <code>list.sort(key = itemgetter(1,2))</code>
#list.sort()是直接原list排序，會改變原來串列內容 sorted(list)是複製一份進行排序，不改變原來串列值
#數字：小→大；英文：A→Z→a→z
#字串(數字)：1→9、1個字元→2字元
#字串(中文)：不一定，但相同字會排一起
#反向排序 (reverse = True)；照鍵值長度排序(key = len)

Dict
單欄
<code>sorted(dict.items(), key=lambda x:x[1])</code>
from operator import itemgetter <code>sorted(dict.items(), key= itemgetter(1))</code>
多欄
<code>sorted(dict.items(), key=lambda x: (x[1],x[0]))</code> #先依value, value相同依key排序
#sorted(dict)#只能依照key值的大小進行排序
#排序順序同list

Numpy
單欄
<code>sorted(ndarray)</code> #同list用法
<code>np.sort(ndarray)</code>
<code>np.argsort(ndarray)</code> #對陣列中的值進行排序並將索引返回
<code>ndarray.sort()</code>
<code>ndarray = np.array([1,2,3,4,5,3,4,5,1])</code> <code>np.searchsorted(ndarray,6, side = 'right')</code> #回傳9，6要插在索引值9的位置，side左至右搜尋或右至左搜尋
多欄
1.使用order參數 需先設置ndarray的dtype名稱 <code>dt = np.dtype([('name', object),('age', int)])</code> <code>ndarray = np.array([("raju",11),("anil",15),("ravi", 17), ("amar",27)], dtype = dt)</code> <code>np.sort(ndarray, order =['age', 'name'])</code>
2. <code>np.lexsort((nd1,nd2))</code> #同argsort()取得排序後的索引值 <code>nd1 = ['raju','anil','ravi','amar']</code> <code>nd2 = ['b.y.', 'a.y.', 's.y.', 'f.y.']</code> <code>ind = np.lexsort((nd1,nd2))</code> #第一個要排序的放後面，ndarray長度、型態需相同，ind=[3 1 0 2] <code>[nd1[i] + ", " + nd2[i] for i in ind]</code> #再使用迴圈取得排序後資料
ndarray.sort()是直接原陣列排序，會改變原來陣列內容 np.sort(ndarray)是複製一份進行排序，不改變原來陣列內容
多維：按列排序axis=0；按行排序axis = 1
kind預設為'quicksort'，還有'mergesort', 'heapsort'跟'stable'
order 一个字串或list，可以設定按照某個屬性進行排序

Series
單欄
<code>Series.sort_index()</code> #依索引值排序
<code>Series.sort_values()</code> #依值排序
預設排序小到大(ascending = True),ascending = False大到小

DataFrame
單欄
<code>df.sort_index()</code> #依索引值排序, axis指定軸向
<code>df.sort_values()</code> #依值排序
<code>df['欄位'].rank()</code> #將數據排名，數值最小的排名最高 有4種method #max：同分則並列排名，但依照最大的排名排，2,3名同分，兩個排名皆為3 #min：同分則並列排名，但依照最小的排名排，2,3名同分，兩個排名皆為2 #average：同分取排名平均，2,3名同分則排名變為(2+3)/2=2.5 #first：同分依照索引值順序排名
多欄
<code>df.sort_values(["欄位1", "欄位2"], ascending = [True, False])</code> #先依欄位1小到大，再依欄位2大到小
預設排序小到大(ascending = True),ascending = False大到小