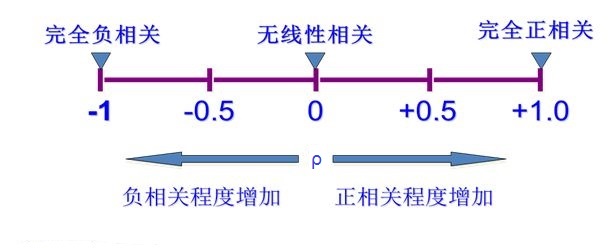
一、皮尔逊相关系数计算公式



编程实现公式：

sum\_xy = 0

sum\_x = 0

sum\_y = 0

sum\_xx = 0

sum\_yy = 0

n = 0

for (x, y) in some\_condition:

n += 1

sum\_xy += x \* y

sum\_x += x

sum\_y += y

sum\_xx += pow(x, 2)

sum\_yy += pow(y, 2)

if n == 0:

return 0

#皮尔逊相关系数计算公式

denominator = sqrt(sum\_xx - pow(sum\_x, 2) / n) \* sqrt(sum\_yy - pow(sum\_y, 2) / n)

if denominator == 0:

return 0

else:

numerator = sum\_xy - (sum\_x \* sum\_y) / n

return numerator / denominator

二、数据准备

|  |  |  |
| --- | --- | --- |
| username | score | bookid |
| Liu Yi | 3 | 1001 |
| Chen Er | 4 | 1001 |
| Zhang San | 3 | 1001 |
| Li Si | 3 | 1001 |
| Liu Yi | 3 | 1002 |
| Li Si | 4 | 1002 |
| Liu Yi | 4 | 1003 |
| Zhang San | 5 | 1003 |
| Li Si | 5 | 1003 |

三、程序

三、程序分析

1 读取user\_book.csv中的数据，放到list中

[

[‘Liu Yi’, ‘3’, ‘1001’]

[‘Chen Er’, ‘4’, ‘1001’]

[‘Zhang San’, ‘3, ‘1001’]

[‘Li Si’, ‘3’, ‘1001’]

[‘Liu Yi’, ‘3’, ‘1002’]

[‘Li Si’, ‘4’, ‘1002’]

[‘Liu Yi’, ‘4’, ‘1003’]

[‘Zhang San’, ‘5, ‘1003’]

[‘Li Si’, ‘5’, ‘1003’]

]

把list中的数据，转换成dict形式。

注意，外面的dict是以用户为key，以bookid和评分构成的字典为value；里面的字典是以bookid为key，以评分为value

{

‘Liu Yi’ : {

‘1001’ : 3.0,

‘1002’ : 3.0,

‘1003’ : 4.0

},

‘Chen Er’ : {

‘1001’ : 4.0

},

‘Zhang San’ : {

‘1001’ : 3.0,

‘1003’ : 5.0

},

‘Li Si’ : {

‘1001’ : 3.0,

‘1002’ : 4.0,

‘1003’ : 5.0

}

}

distances:

[

('Zhang San', 0.9999999999999998),

('Liu Yi', 0.8660254037844402),

('Chen Er', 0)

]