一、数据准备

为了便于理解程序，这里准备了包含四行二列的简单数据

1.658985 4.285136

-3.453687 3.424321

4.838138 -1.151539

-5.379713 -3.362104

二、初始化形心

在initCentroids方法中，dataSet就是上面准备的数据，k取2

则index会取两次随机数，随机为0~4之间（不包含4）的某个整数

i = 1时，我这里取到的随机数是1，

centroids[1, :] = dataSet[1, :] = [-3.453687, 3.424321]

i = 2时，我这里取到的随机数是0

centroids[2, :] = dataSet[2, :] = [1.658985, 4.285136]

所以，centroids =

[[ 0. 0. ]

[-3.453687 3.424321]

[ 1.658985 4.285136]]

三、求dataSet四个点与形心两个点的距离

形心点会随着每次的调整而不断发生变化

clusterAssment的初始值：

[[-1. 0.]

[-1. 0.]

[-1. 0.]

[-1. 0.]]

（一）第一轮

i = 0, j = 0, distance = sqrt((1.658985 + 3.453687) ^ 2 + (4.285136 - 3.424321) ^ 2) = 5.184633

i = 0, j = 1, distance = sqrt((1.658985 - 1.658985) ^ 2 + (4.285136 - 4.285136) ^ 2) = 0

i = 1, j = 0, distance = sqrt((-3.453687 + 3.453687) ^ 2 + (3.424321 - 3.424321) ^ 2) = 0

i = 1, j = 1, distance = sqrt((-3.453687 - 1.658985) ^ 2 + (3.424321 - 4.285136) ^ 2) = 5.184633

i = 2, j = 0, distance = sqrt((4.838138 + 3.453687) ^ 2 + (-1.151539 - 3.424321) ^ 2) = 9.470631

i = 2, j = 1, distance = sqrt((4.838138 - 1.658985) ^ 2 + (-1.151539 - 4.285136) ^ 2) = 6.297972

i = 3, j = 0, distance = sqrt((-5.379713 + 3.453687) ^ 2 + (-3.362104 - 3.424321) ^ 2) = 7.054441

i = 3, j = 1, distance = sqrt((-5.379713 - 1.658985) ^ 2 + (-3.362104 - 4.285136) ^ 2) = 10.393438

从上面的计算结果可以看出，

dataSet[0]离centroids[1]更近

dataSet[1]离centroids[0]更近

dataSet[2]离centroids[1]更近

dataSet[3]离centroids[0]更近

则clusterAssment =

[[1, 0]

[0, 0]

[1, 39.664448]

[0, 0]]

四、求质心centroids