

## DEAP 数据库

EEG\_feature.txt 包含了1216个脑电信号样本的160维特征，每行为一个样本，每列为一种特征。特征从左至右分别是每个脑电电极的theta（1-32列）、slow alpha（33-64列）、alpha（65-96列）、beta（1-97128列）、gamma（129-160列）波段的脑电特征。

subject\_video.txt 包含了1216个脑电信号对应的32名被试和38段视频信息，其中包含两列。第一列是对应的被试编号，第二列是对应的视频编号。

EEG\_feature.txt 与 subject\_video.txt和valence\_arousal\_label.txt中每行都是一一对应的，例如subject\_video.txt的第二行就是EEG\_feature.txt中第二个样本（第二行）的被试和视频信息；valence\_arousal\_label.txt的第二行也是EEG\_feature.txt中第二个样本（第二行）的愉悦度和唤醒度标签。valence\_arousal\_label.txt中第一列为愉悦度标签，**1代表positive，2代表negative**；第二列为唤醒度标签，**1代表high，2代表low**。

## 数据预处理

根据valence\_arousal\_label.txt标签将样本分为四类

分类1: positive high [1,1]

分类2: positive low [1,2]

分类3: negative high [2,1]

分类4: negative low [2,2]

EEG\_pca\_feature.txt 是经过PCA降维处理后的二维坐标，用于作图显示分类结果使用

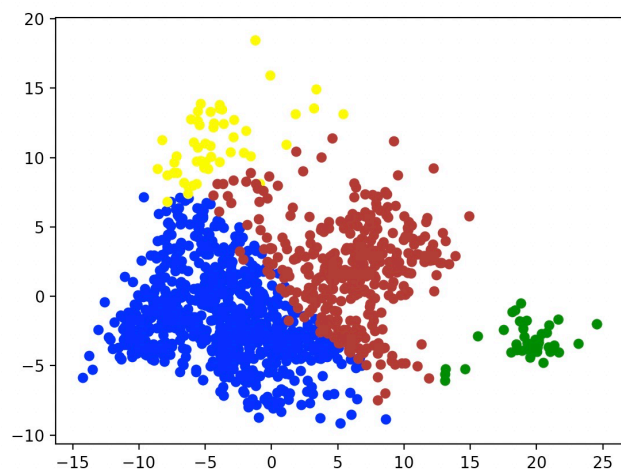
## 最大最小距离法聚类

入参：

thita：距离阈值，取值范围(0,1)

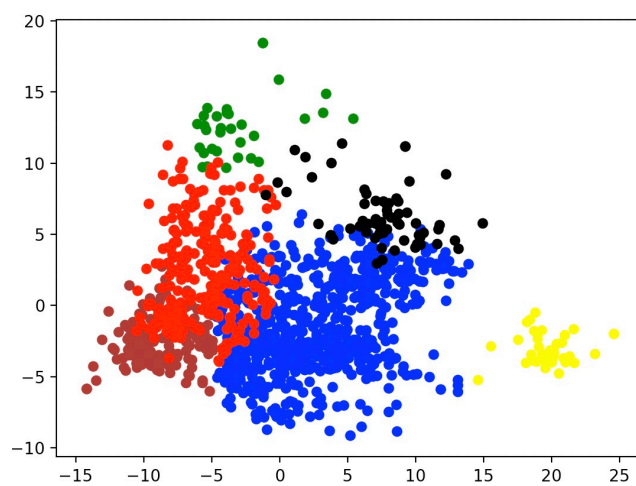
Z1:初始聚类中心

thita 取0.6； Z1=0（第1个点）



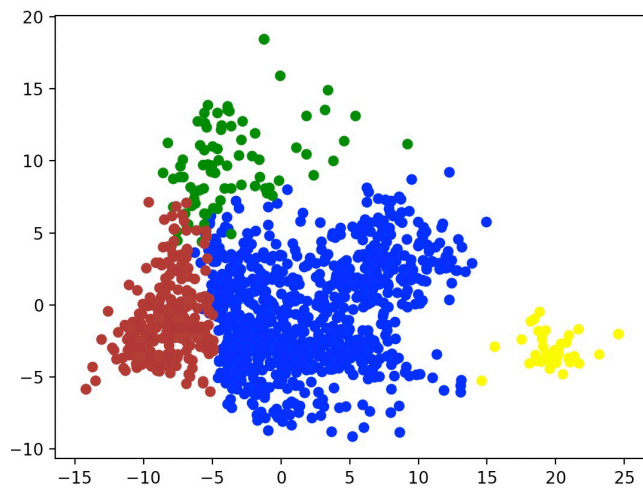
图一

thita取0.6; Z1=500（第501个点）



图二

thita取0.77; Z1=500（第1个点）



图三

#### 情况一，调整初始聚类中心

由上图一、图二可知，不同的初始聚类中心对聚类结果甚至聚类个数有较大影响，初始中心影响比较随机

#### 情况二，调整距离阈值

由上图二、图三可以看到，thita的取值对分类结果以及聚类个数有较大影响，thita越小分类越多

## kMeans聚类

参数：

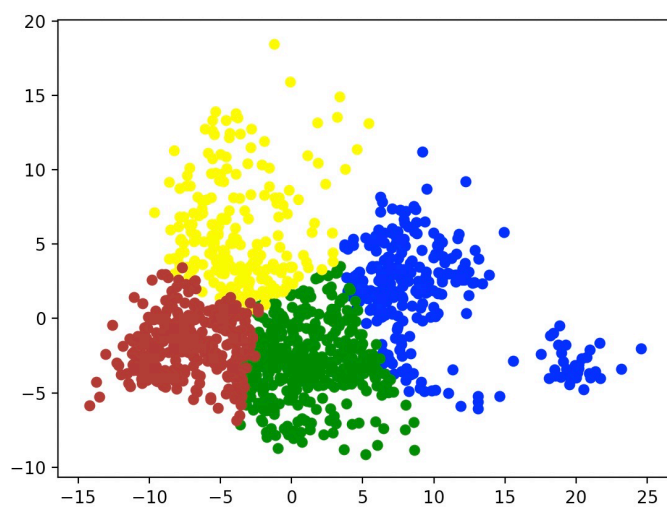
Z：初始聚类中心

K：聚类个数

有数据背景可知分类个数有四个，K值固定为4，可调初始整聚类中心Z

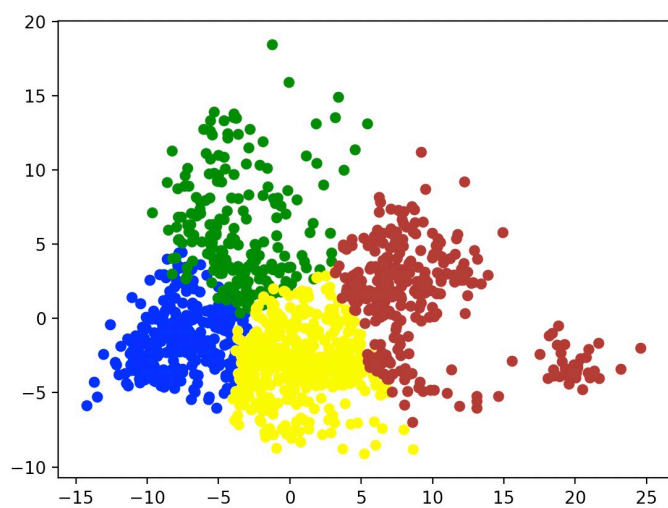
根据标签从每类样本选取一个样本作为初始聚类中心

情况一：Z = [dataSet[530],dataSet[531],dataSet[532],dataSet[533]]



图四

情况二：  $Z = [\text{dataSet}[0], \text{dataSet}[1], \text{dataSet}[2], \text{dataSet}[7]]$

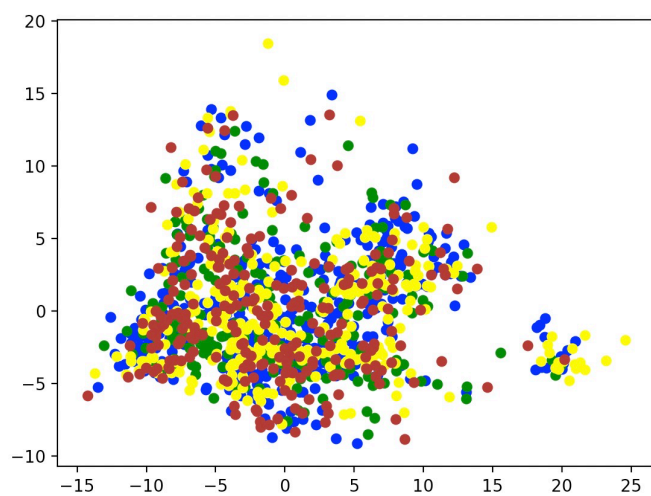


图五

由图四、图五可知初始聚类中心选择对聚类结果有轻微影响，基本聚类一致

## 标签实际分类

python3 util.py



图六

'blue', 聚类1 (1, 1)

'green', 聚类2 (1, 2)

'yellow', 聚类3 (2, 1)

'brown', 聚类4 (2, 2)

由上图实际样本分类结果可以发现，样本分布并不线性可分，跟MinMax与kMeans分类结果比较发现，都不能正确的将样本分类，综上该样本集并不适合简单的kMeans或者MinMax分类