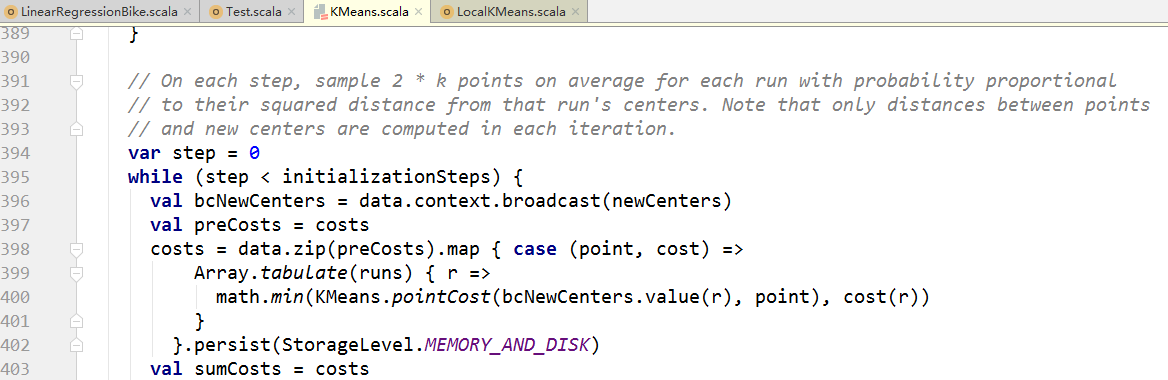
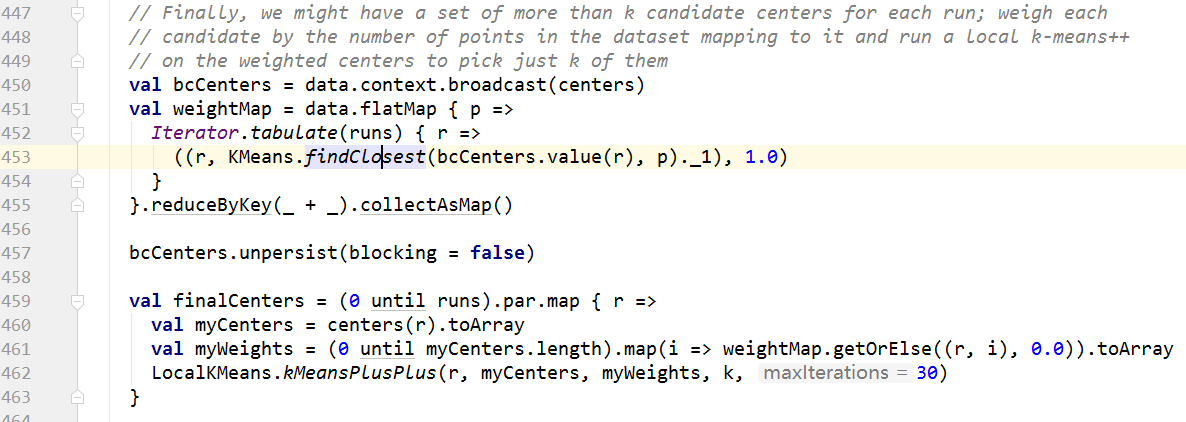
395行是我们需要迭代5次的位置



462行是当我们得到3000个候选聚类中心后，使用本地KMeans++来进行选择1000个聚类中的方法



48行是按照权重，随机抽取一个作为第一个聚类中心

53行是重点，这里的kmeans++计算距离的方式有些改动，它不仅仅算距离，还要乘上样本的权重，把这个结果做为概率。

