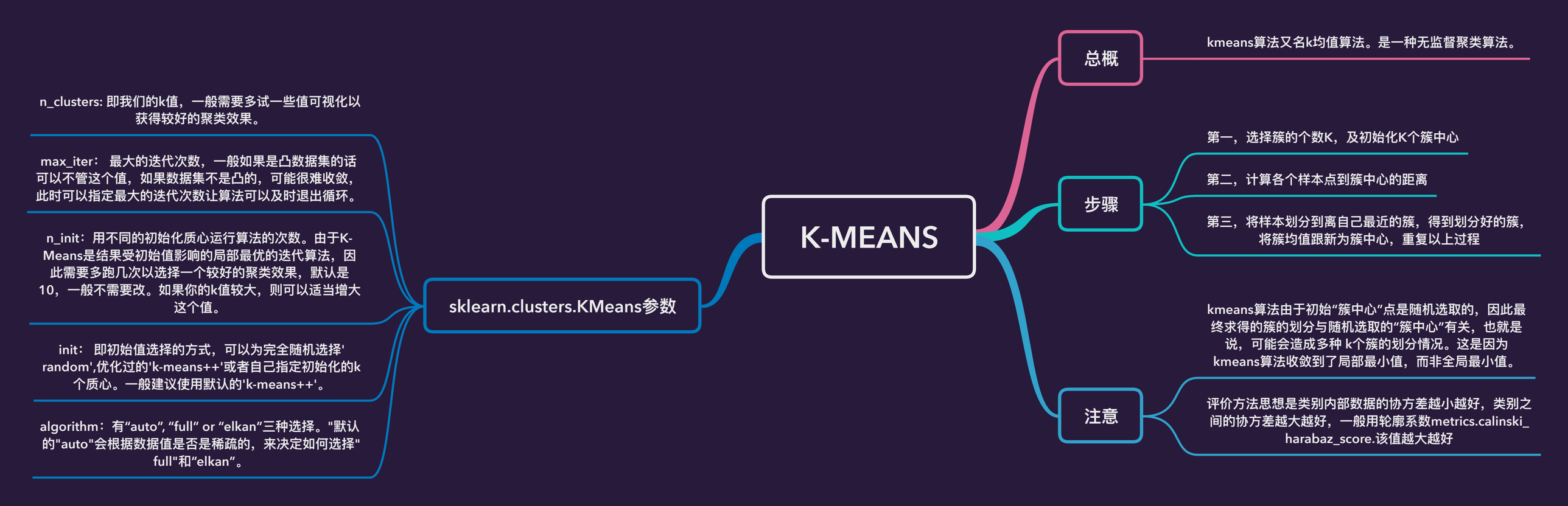
K-MEANS
总概
kmeans算法又名k均值算法。是一种无监督聚类算法。
步骤
第一，选择簇的个数K，及初始化K个簇中心
第二，计算各个样本点到簇中心的距离
第三，将样本划分到离自己最近的簇，得到划分好的簇，将簇均值跟新为簇中心，重复以上过程
注意
kmeans算法由于初始“簇中心”点是随机选取的，因此最终求得的簇的划分与随机选取的“簇中心”有关，也就是说，可能会造成多种 k个簇的划分情况。这是因为kmeans算法收敛到了局部最小值，而非全局最小值。
评价方法思想是类别内部数据的协方差越小越好，类别之间的协方差越大越好，一般用轮廓系数metrics.calinski_harabaz_score.该值越大越好
sklearn.clusters.KMeans参数
n_clusters: 即我们的k值，一般需要多试一些值可视化以获得较好的聚类效果。
max_iter： 最大的迭代次数，一般如果是凸数据集的话可以不管这个值，如果数据集不是凸的，可能很难收敛，此时可以指定最大的迭代次数让算法可以及时退出循环。
n_init：用不同的初始化质心运行算法的次数。由于K-Means是结果受初始值影响的局部最优的迭代算法，因此需要多跑几次以选择一个较好的聚类效果，默认是10，一般不需要改。如果你的k值较大，则可以适当增大这个值。
init： 即初始值选择的方式，可以为完全随机选择'random',优化过的'k-means++'或者自己指定初始化的k个质心。一般建议使用默认的'k-means++'。
algorithm：有“auto”, “full” or “elkan”三种选择。"默认的"auto"会根据数据值是否是稀疏的，来决定如何选择"full"和“elkan”。