

机器学习课程 第9次作业

黄昊 20204205

- 9.3 不能，因为 k-means 本身是 NP 问题，且 9.24 是非凸问题，容易陷入局部最优，所以在使用 k 均值时常常多次随机初始化中心点，然后挑选结果最好的一个。
- 9.5 连接性：由于任意 $x' \in D$ 都由 x 密度可达，于是任意 $x_i, x_j \in D$ 都可通过 x 密度相连；
最大性： $x_i \in D \Rightarrow x_i$ 由 x 密度可达， x_j 由 x_i 密度可达 $\Rightarrow x_j$ 由 x 密度可达 $\Rightarrow x_j \in D$
- 9.7 凸聚类：K均值，LVQ
非凸聚类：高斯混合聚类，DBSCAN
- 9.8 同类协方差尽量小，异类中心之间距离尽量大。
- 9.9 对于连续性属性，使用min-max归一化；对于分类属性，如果为有序变量，则等距映射到[0,1]的值域类，否则采用类似独热编码的办法，采取0-1取值。