

Chapter 5: Unsupervised Learning

输入数据没有标签信息，需要通过算法模型来挖掘数据内在的结构和模式

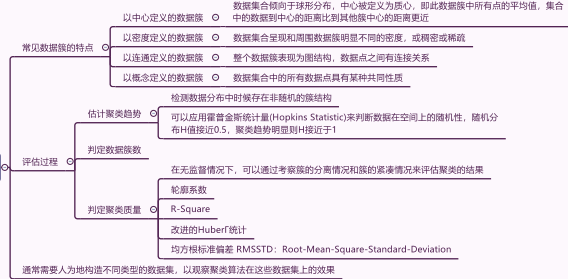
主要类别

- 数据聚类
 - 多次迭代找到数据的最优分割
 - 通过数据之间的内在关系把样本划分为若干类别，使得同类样本之间相似度高，不同类别之间相似度低
- 特征变量关联
 - 利用各种相关性分析方法来找别变量之间的关系

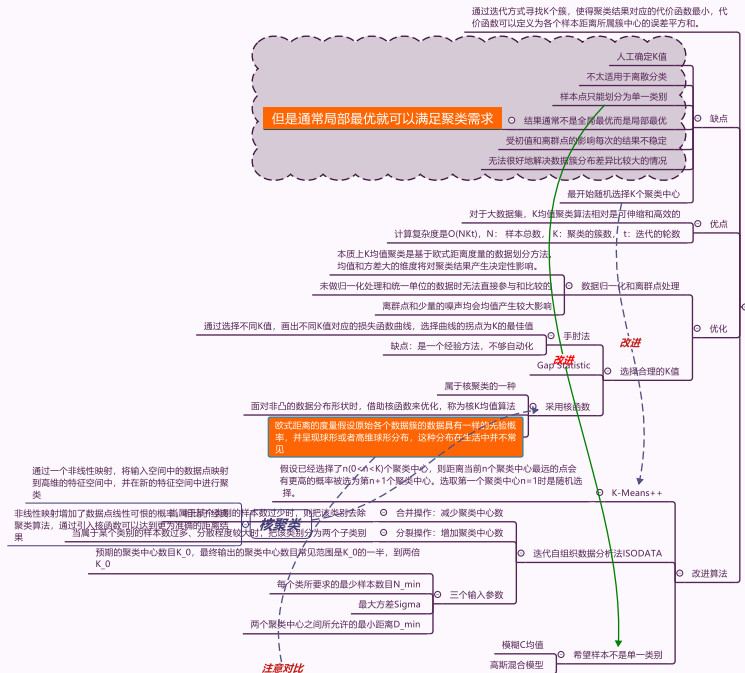
自组织映射神经网络 SOM: Self-Organizing Map



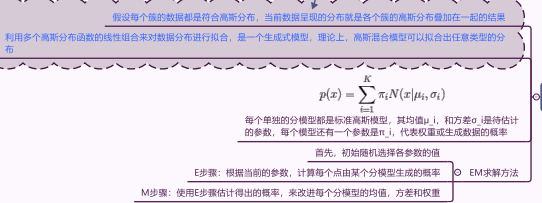
聚类算法的评估



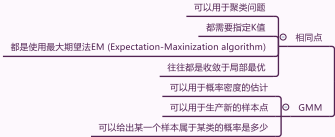
K 均值聚类



高斯混合模型 GMM: Gaussian Mixed Model



K-Means vs GMM



2018.12.16 Sunday. By KuKuXia@gitHub.com

ML/DL/RL 深度交流讨论群: 622545311, 加群时请备注: 姓名-方向-公司 or 高校