

1、kNN算法的优缺点

优点

- 简单好用，容易理解，精度高，理论成熟，既可以用来做分类也可以用来做回归；
- 可用于数值型数据和离散型数据；
- 训练时间复杂度为 $O(n)$ ；无数据输入假定；
- 对异常值不敏感

缺点

- 计算复杂性高；空间复杂性高；
- 样本不平衡问题（即有些类别的样本数量很多，而其它样本的数量很少）；
- 一般数值很大的时候不用这个，计算量太大。但是单个样本又不能太少，否则容易发生误分。
- 最大的缺点是无法给出数据的内在含义。

k近邻 法三要素：距离度量、k值得选择和分类决策规则。常用的距离度量是欧氏距离及更一般的 L_p 距离。k值小时，模型更复杂，容易发生过拟合；k值大时，模型更简单。k值的选择反映了对近似误差与估计误差之间的权衡，通常又交叉验证选择最优的k。常用的分类决策规则是多数表决，对应于经验风险最小化。

ps:对近似误差与估计误差的理解，

https://blog.csdn.net/qg_35664774/article/details/79448076做了总结。