N_neighbors:指K

Algorithm:brute暴力法、kd-tree用kd树、ball-tree球树,数据越大越需要用后者

Leaf_size:停止建子树的叶子节点数,默认30

P: 距离度量方法,euclidean欧式距离、manhattan曼哈顿距离、chebyshev切比雪夫距离、minkowski闵可夫斯基距离

n-jobs: 指定cpu计算个数,默认-1,即全部

sklearn.neighbors.KNeighborsClassifier 参数 3个要素: 距离度量、K值选择、分类决策规则

距离有欧式距离、曼哈顿距离、切比雪夫距离等

K值影响很大,所以一般先取一个较小的数,然后通过交 叉验证得到最优。K值越小,模型越复杂

由输入实例的K个近邻的训练实例中的多数类决定输入实 例的类,多数表决规则等价于经验分险最小化

构造根节点,一般选中位数

以该点的坐标轴对空间就行划分(如x轴)

生成两个子节点(因为上面依照x轴划分,所以是X值最 靠近的点)

以两个子节点的坐标轴对空间就行划分(注意上面用了x 轴,这里就用y轴,x,y交替使用)

不断重复

这样KD树比较平衡,但未必效率最优

当训练集、距离度量、K值以及分类决策规则确定后,其 结果唯一确定(不存在优化)

注意

KNN

KD树

总概