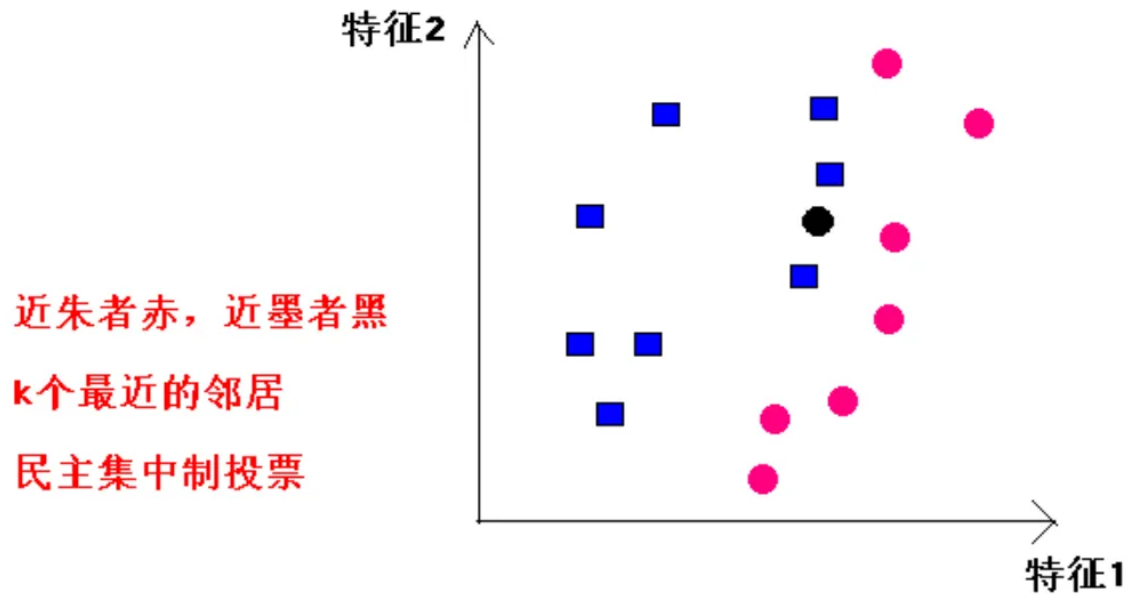


KNN算法

一、什么是KNN

k-近邻算法 (k-Nearest Neighbour algorithm)，又称为KNN算法，是机器学习中原理最简单的算法。主要用在分类任务中，当然也可以应用在回归任务中。

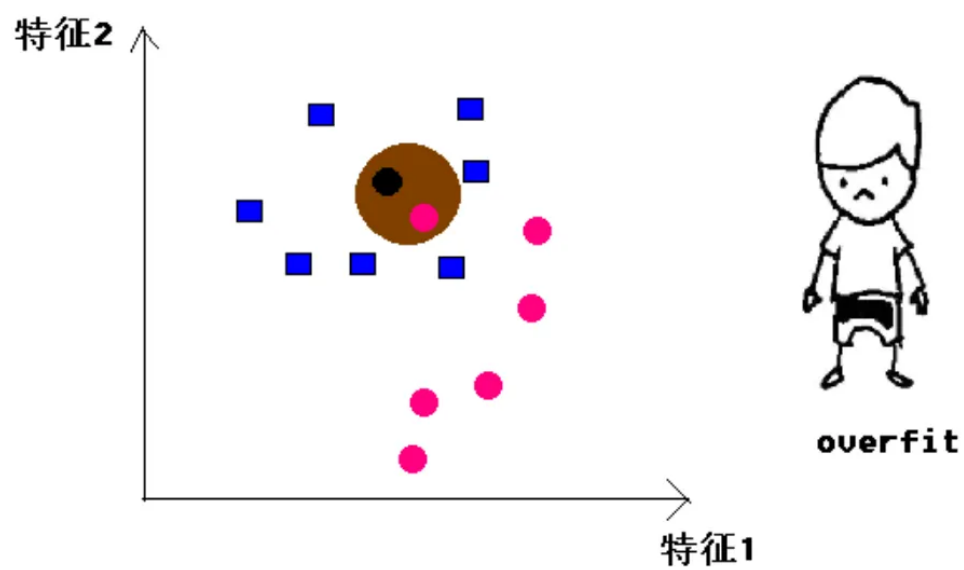
KNN的工作原理：给定一个已知标签类别的训练数据集，输入没有标签的新数据后，在训练数据集中找到与新数据最邻近的k个实例，如果这k个实例的多数属于某个类别，那么新数据就属于这个类别。可以简单理解为：由那些离X最近的k个点来投票决定X归为哪一类。



二、k值的选择

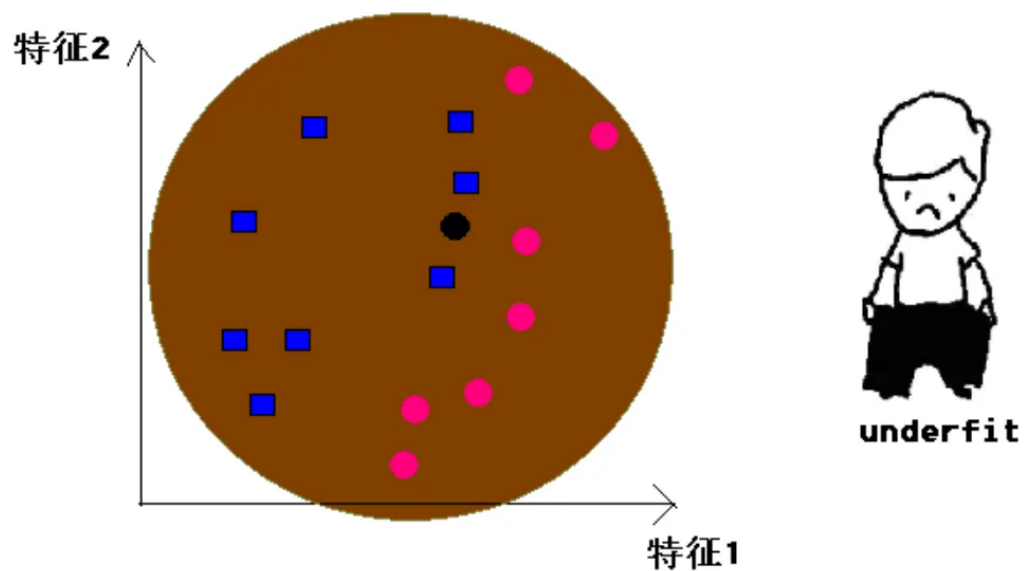
2.1 K太小导致“过拟合”

- K太小导致“过拟合”
- 容易把噪声学进来, 忽略了数据的真实分布情况



2.2 K太大导致“欠拟合”

- K太大导致“欠拟合”
- 决策效率低



所以：

1. k不能太小，也不能太大
2. Fit= 最优拟合
3. 需要超参数调参

三、距离的度量

1. 曼哈顿距离
2. 欧式距离

四、演示KNN的算法原理

五、KNN实现分类