

统计学习：第三章

1 参照图 3.1, 在二维空间中给出实例点, 画出 k 为 1 和 2 时的 k 近邻法构成的空间划分, 并对其进行比较, 体会 k 值选择与模型复杂度及预测准确率的关系。

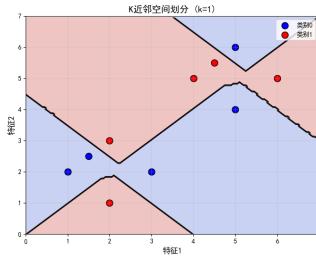


图 1: $K=1$ 时 K 近邻算法的二维空间划分

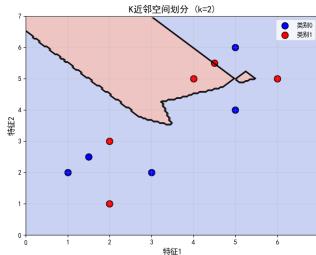


图 2: $K=2$ 时 K 近邻算法的二维空间划分

$k=1$ 模型边界更加清晰, 但其实存在过拟合, $k=2$ 模型变得简单, 但是近似误差也在增大。

2 利用例题 3.2 构造的 kd 树求点 $z = (3,4.5)$ 的最近邻点。

答: $(3, 4.5)$ 在 $(4, 7)$ 下, 距离约为 2.69。回到父节点 $(5, 4)$, 计算距离为 1.80, 为当下最近点; 其另一子节点 $(2, 3)$ 可能与其相交, 计算距离为 1.80, 为当下最近点。再考虑父节点 $(7, 2)$ 由于不相交, 搜索停止。故最近点为 $(2, 3)$ 。

3 参照算法 3.3, 写出输出为 2 的 k 近邻的。