

《统计学习方法》(李航)

归纳总结三

K 近邻法

西安交通大学

联系方式: [williamyi96@gmail.com](mailto:williamyi96@gmail.com)

易凯

2017 年 1 月 24 日

目录	2
----	---

## 目录

<b>1 K 近邻法</b>	<b>3</b>
1.1 K 近邻算法 . . . . .	3
1.2 K 近邻模型 . . . . .	3
1.3 K 近邻法的实现: KD 树 . . . . .	3

# 1 K 近邻法

## 1.1 K 近邻算法

K 近邻算法是基本而且简单的分类和回归方法。其基本思路是首先确定输入实例点的  $k$  个最近邻实例点，然后利用这  $k$  个近邻的实例点采用诸如多数表决的方式决定该点所属的类。

## 1.2 K 近邻模型

**K 近邻法三要素：** 距离度量、 $k$  值选择、分类决策规则

**模型** K 近邻模型对应于基于训练数据集对特征空间的一个划分。K 近邻法中，当训练集、距离度量、 $k$  值和分类策略规则确定之后，其结果就具有了唯一性。

**距离测量** 常用的距离测量方法有欧氏距离法和  $L_p$  距离法。

**K 值选择** K 值的选择要适当，过小噪声引起的估计误差会增大；过大的分类本身有不具备代表性。K 值在数据量有限的情况下一般使用交叉验证法进行确定。

**分类决策规则** 分类决策一般采取多数表决的形式，其对应的是经验风险的最小化。

## 1.3 K 近邻法的实现：KD 树

K 近邻法的实现需要考虑如何快速搜索  $K$  个最近邻点。KD 树是一种便于对  $K$  维空间中的数据进行快速检索的数据结构。

同时，KD 树是一棵二叉树，表示的是对维空间的一个划分，其每个结点对应于  $K$  维空间划分中的一个超矩形区域。

利用 KD 树可以省去大部分数据点的搜索，从而减少搜索的计算量。