KNN 分类算法

主讲人：王石平

Outline

- Motivation
- KNN algorithm
- Variations of KNN
- Project

classification

Algorithm 1 K-nearest neighbors (KNN) algorithm

Input: Training samples $X = \{x_1, \cdots, x_n\}$, their class label $f(X) = (f(x_1), \cdots, f(x_n))$, a new sample (test sample) x, and the number of nearest neighbors k.

Output: the prediction class of x, i.e., $\widehat{f}(x)$.

1: Compute the k-nearest neighbors of x and denote them as $s_1, \cdots, s_k$; // Compute all neighbors.

2: **return** $\widehat{f}(x) \longleftarrow \arg\max_{y \in f(X)} \sum_{i=1}^{k} \delta(y, f(s_i))$ where $\delta(a, b) = 1$ iff $a = b$; // Return the class with highest frequency in the neighbors.

距离加权最近邻算法

$$\widehat{f}(x) \longleftarrow \arg \max_{y \in f(X)} \sum_{i=1}^{k} w_i \delta(y, f(x_i))$$

“回归(regression)” 最近邻算法

$$\widehat{f}(x) \longleftarrow \frac{\sum_{i=1}^{k} f(x_i)}{k}$$

加权“回归(regression)” 最近邻算法

$$\widehat{f}(x) \longleftarrow \frac{\sum_{i=1}^{k} w_i \delta(y, f(x_i)}{\sum_{i=1}^{k} w_i}$$

给定特征空间$X \in \mathbb{R}^{n\times d}$, 和邻域个数$k$, 编程计算所有样本的$k$ 邻域，其输出结果如：

$$\begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

提交日期: 2015-05-21 前.
提交方式：发送到邮箱529836750@qq.com.

Thanks!