

K 近邻学习

2020年4月25日 星期六 上午8:53

急切学习 (eager learning) 训练时对样本进行学习

K 近邻: K-Nearest Neighbor (KNN) —— 懒惰学习 (lazy learning)

① 给定测试样本 x , 最近邻样本 z , 最近邻为类别 c 出错概率

$$P(\text{err}) = 1 - \sum_{c \in Y} P(c|x) P(c|z)$$

缓解维数灾难: 降维

① 要求原始空间样本间距离在低空间仍保持 一维缩放 MDS

$$\text{dist}_{ij} = \|z_i - z_j\| \quad (m \text{ 个样本 } z \in \mathbb{R}^{m \times m}, \text{ 行 } i \text{ 到 } j \text{ 为 } \text{dist}_{ij} \text{ 为样本 } x_i \text{ 到 } x_j \text{ 距离})$$

$$B = z^T \cdot z \in \mathbb{R}^{m \times m} \quad (\text{降维后内核 } b_{ij} = z_i^T \cdot z_j)$$

$$\text{dist}_{ij}^2 = \|z_i\|^2 + \|z_j\|^2 - 2\|z_i\| \|z_j\| \cos \theta = b_{ii} + b_{jj} - 2b_{ij}$$

② PCA 主成分分析

③ 核化线性降维 (KPCA)