第10章 降维与度量学习

1. KNN

基础：给定测试样本，基于某种距离度量找出训练集中与其最靠近的k个训练样本，然后基于这k个邻居进行预测

- 分类：投票法 — k样本中出现最多的类别标记（多数表决，加权多数表决）
- 回归：平均法 — k样本的实值输出标记的平均值作为预测结果

给定测试样本x，若最邻近样本为z，则最近邻分类器出错概率即

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

三因素：k值的选择（k小易过拟合，k大易欠拟合），距离测量（公式）

决策规则（分类，回归）

算法：
- 暴力 brute
- kd-tree，ball-tree，BBF tree，MVP tree

kd-tree：从m个样本的n维特征中，分别计算n个特征取值的方差，用方差最大的第k维特征n_k作为根节点，取值为n_{kv}作为划分点，<该值的为左子树，>该值的右子树

2. 低维嵌入和PCA

降维：通过某种数学变换将原始高维属性空间转变为低维子空间

即高维空间中的低维嵌入　　多维缩放（MDS）

PCA：通过某种线性投影，将高维的数据映射到低维的空间中表示，并且期望在所投影的维度上数据的方差最大，以此使用较少维度

正交属性空间中存在超平面：

- 最近重构性：样本到超平面的距离足够近
- 最大可分性：样本点在这个超平面上的投影尽可能分开

原理：X是已中心化(z-score)过的数据矩阵。样本点 x_i 在超平面上二投影 W^Tx_i。投影后的点在各维度上方差最大化。方差和 $\frac{1}{n}\sum W^Tx_ix_i^TW$。

目标函数：$\max_W tr(W^TXX^TW)$

$L=W^TXX^TW+\lambda(I-W^TW)$　　$\frac{\partial L}{\partial W}=2XX^TW-2\lambda W$

令 $\frac{\partial L}{\partial W}=0 \rightarrow XX^TW=\lambda W$　　$W^TXX^TW=W^T\lambda W=\lambda$

步骤：数据中心化 → 求中心化矩阵X的协方差矩阵 → 特征值/向量

→ 从大到小排列，获取最前面的k列 → 得到矩阵W

3. 核化线性降维和流形学习

基于核技巧对线性降维进行核化

$k(x_i,x_j)=\phi(x_i)^T\phi(x_j)$　　$z_j=W_j^T\phi(x)=\sum_{i=1}^{m}\alpha_i^j\phi(x_i)^T\phi(x)$

$=\sum_{i=1}^{m}\alpha_i^jk(x_i,x_2)$

KPCA计算量大

流形学习 { 等度量映射　最短路径 MDS算法
　　局部线性嵌入　LLE