第六章，K近邻算法

1. 优点：精度高，对异常值不敏感（受异常值的影响小），无数据输入的假定
2. 缺点：计算复杂度高，空间复杂度高，适用数据范围小（数值型和标称型）
3. 距离的计算方法：LP距离，欧式距离（LP中P=2），曼哈顿距离
4. 近似误差：可以理解为对现有训练集的训练误差
5. 估计误差：对测试集的测试误差
6. 近似误差关注训练集，如果近似误差小了会出现过拟合的现象，对现有的训练集能有很好的预测，但是对未知的测试样本将会出现较大偏差的预测。模型本身不是最接近最佳模型。

   估计误差关注测试集，估计误差小了说明对未知数据的预测能力好。模型本身最接近最佳模型。
7. K值的选择
    1. 如果K值较小，则学习的近似误差会减小，但学习的估计误差会增大，对噪声敏感，且K值的减小容易发生过拟合，模型变得复杂
    2. 若K值较大，则学习的近似误差会增大，但是估计误差会减少，模型简单
8. 分类决策规则：多数表决规则（经验风险最小化）