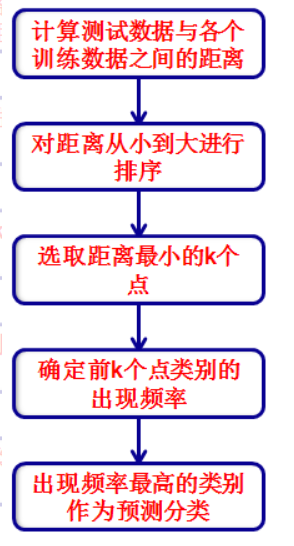
《机器学习》读书笔记

1. 算法概述（算法名称及原理）

KNN：如果一个样本在特征空间中的k个最相邻的样本中的大多数属于某一个类别，则该样本也属于这个类别，并具有这个类别上样本的特性。该方法在确定分类决策上只依据最邻近的一个或者几个样本的类别来决定待分样本所属的类别。

1. 算法设计（流程图及主要分段代码，附详细代码注释）





1. 选用数据（数据集描述，包括来源，行数，列数，格式等）

数据集是sklearn.datasets库中的鸢尾花数据

描述信息如下：

* 共有150行5列，
* 其中属性信息如下：

1. 萼片长度cm，
2. 萼片宽度cm，
3. 花瓣长度cm，
4. 花瓣宽度cm
5. 类：   
   - Iris Setosa   
   - Iris Versicolour   
   - Iris Virginica
6. 评价方法（说明训练集和测试集分配方法及评价指标）

十折[交叉验证](https://baike.baidu.com/item/%E4%BA%A4%E5%8F%89%E9%AA%8C%E8%AF%81/8543100" \t "https://baike.baidu.com/item/%E5%8D%81%E6%8A%98%E4%BA%A4%E5%8F%89%E9%AA%8C%E8%AF%81/_blank)。将数据集分成十份，轮流将其中9份作为训练数据，1份作为测试数据，进行试验。每次试验都会得出相应的正确率（或差错率）。10次的结果的正确率（或差错率）的平均值作为对算法精度的估计，一般还需要进行多次10折交叉验证（例如10次10折交叉验证），再求其均值，作为对算法准确性的估计。

1. 实验结果截图



1. 实验结果分析及比较

对于鸢尾花数据集，与c4.5，cart算法相比KNN的验证结果要更优





1. 遇到的问题及解决方法，实践心得

总体上来说KNN算法的实现要比之前的决策树算法要更容易一些，在对欧式距离的实现上，一开始没有想到直接用行列向量的方法，而是采用了更加复杂的循环，后来在查阅资料的时候偶然发现有更容易的方法，可能之后在学习过程中需要考虑同样的问题如何更简单的解决。