Naïve Bayes and AdaBoost

陈晓宇 MF1633007 traumcxy@163.com

1. 实现细节

在本实验中使用了Naïve Bayes和 AdaBoost算法，下面分别给出它们的伪代码：

Naïve Bayes:

假设所有属性相互独立， 基于属性独立条件性假设, ,其中d为属性数目，为在第个属性上的取值。

因此，只要事先将朴素贝叶斯分类器涉及的所有概率估值存储好就行了，这样在进行预测的时候只需要“查表”即可以进行判别。

AdaBoost：

输入：训练集；训练算法A，训练轮数T

过程：

1. 
2. for t = 1,2,…,T do
   1. 
   2. 
   3. If  then break;
   4. 
   5. (其中为归一化函数，使得)
3. end for

输出：

1. 实验
2. 参数设置
   1. 因为要采取10-cross validation，而dataSet1的数据只有277个，因此最后一组有34个数据，dataset2每组100个；
   2. 对于离散型的特征，直接根据特征取值建立分支即可，而对于数值型的，则预先进行处理，按最小到最大分为4个等级，然后用等级去代替原来的数值型的数据，这样就将数值型的转为离散型的了；
   3. Adaboost算法中，迭代次数设为6，即每个分类器最多由6个弱的朴素贝叶斯分类器组成。

|  |  |  |
| --- | --- | --- |
| 平均值（标准偏差） | 数据集1 | 数据集2 |
| Bayes-adaboost | 65.93%(0.0796) | 70%（0.0490） |