

山东大学机器学习课程

实验报告

——实验六：集成学习bagging方法

姓名：陈竞帆

学院：软件学院

班级：软件八班

学号：201500150180

**一、实验目的：**

（1）熟悉python实验软件及相关函数

（2）学习集成学习的概念

（3）利用bagging算法实现mnist数据集的集成学习

**二、实验环境：**

（1）硬件环境：

英特尔® 酷睿™ i5-7500U 处理器

256 GB PCIe® NVMe™ M.2 SSD

8 GB LPDDR3-1866 SDRAM

（2）软件环境：

Windows10家庭版64位操作系统

Python3.6

**三、实验内容**

**3.1 bagging算法概述**

**Bagging算法属于集成学习的一种算法，过程如下：**

**从大小为n的原始数据集D中，分别独立随机地抽取n’个数据（n’<n）形成自助数据集，并且将这个过程独立进行许多次，直到产生很多个独立的自助数据集。然后，每一个自助数据集都被独立地用于训练一个“分量训练器”。**

**在我的实验中，我选择的分量实验器为svm，bpnn和parzen窗方法。其中svm使用的是libsvm自带的方法。Bpnn和parzen都是前面的实验做过的。**

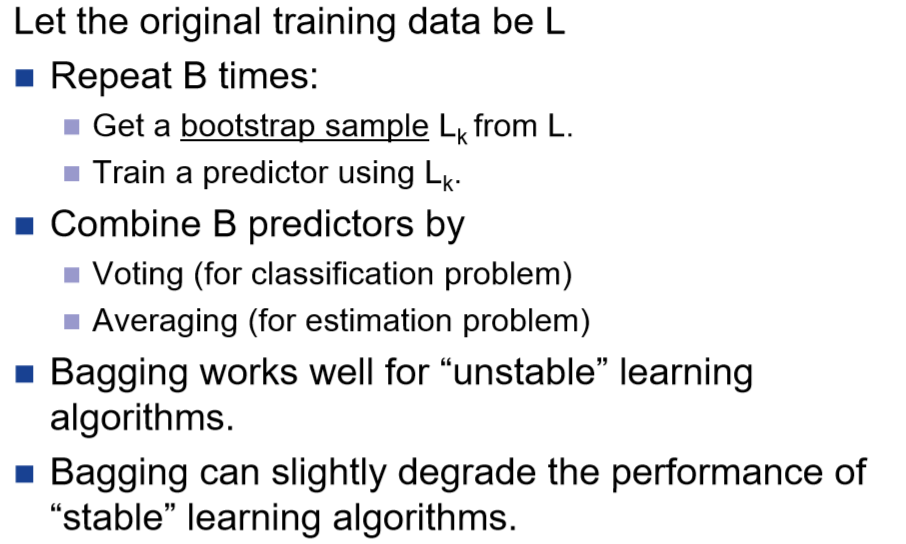
**最终的分类判决将根据这些“分量分类器”各自的判决结果的投票来决定。通常，这些分量分类器的模型形式都是一样的。但是在我的实验中，我选择了不同的分类器。**

**Bagging算法是我们遇到的第一个“多分类器系统”，其中，最后的分类结果取决于许多分量分类器的输出。而bagging法中的最基本的判决规则，就是对各个分量分类器的判决结果使用投票表决原则。**

**在我的代码中，我的投票原则为少数服从多数。**

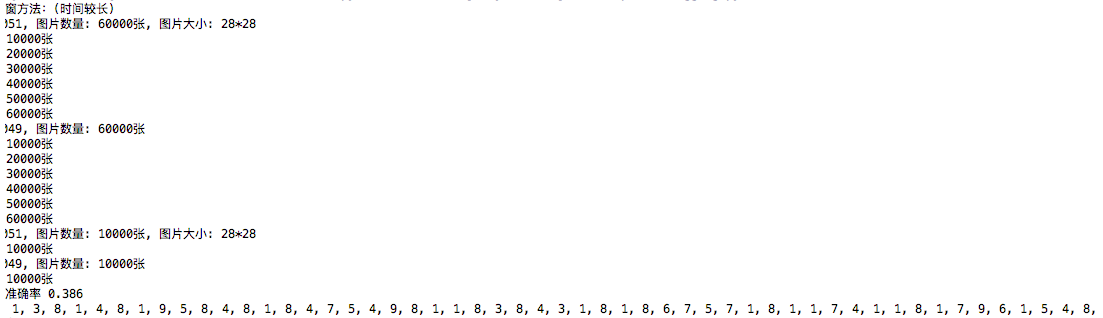
**3.2bagging算法**

设计程序的伪代码，实现图1所示：

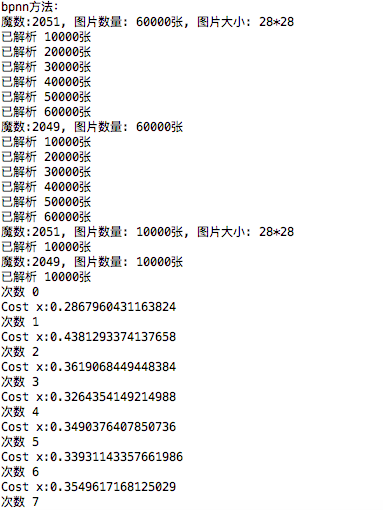


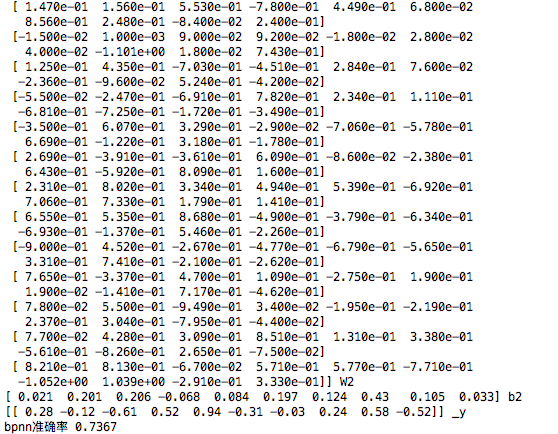
**图1.利用bagging算法的集成学习**

**四、实验结果**

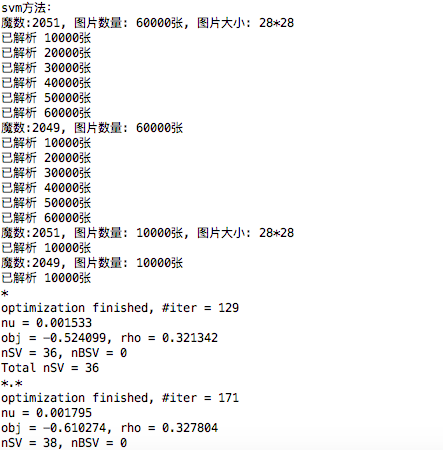


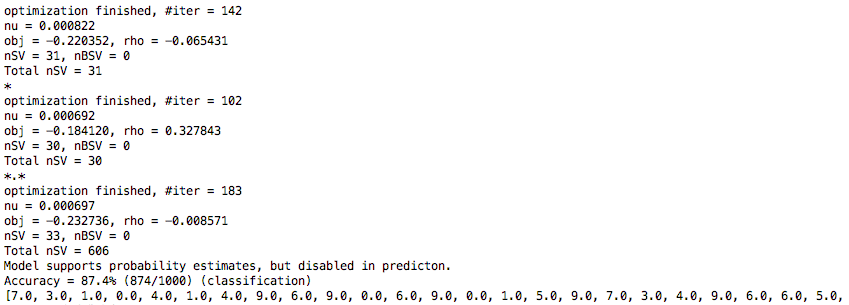
**图2. parzen窗方法得到的预测结果、准确率38.6%**





**图3. 利用两层hidden层结构的bpnn网络预测结果及准确度73.67%**





**图4. 利用svm训练得到的预测结果及准确率 87.4%**

C:\Users\95171\Documents\Tencent Files\951718744\Image\C2C\{7295CE34-CBAC-BAE7-6545-233705E4A689}.png

**图5. 将三个分类器集成，最后通过投票机制得到的预测结果以及准确率90%**

**五、总结与归纳**

**首先，bagging方法更适合于分类器的类别相同，因此如果全部使用bpnn或者svm，那么得到的效果将会更加好。从实验结果我们可以看到，集成学习对准确率的改善有显著的提高。作为本次课程的最后一次实验，是对前面所学知识的一次总结，因此采用了三个不同的分类器。**