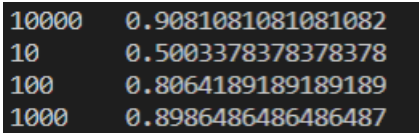


# 山东大学 计算机科学与技术 学院

## 机器学习 课程实验报告

学号：201618130133	姓名：代荣森	班级：2016.4
实验题目：朴素贝叶斯		
实验学时：4	实验日期：2018.11.8	
<b>实验目的：</b> ① 掌握 Naïve Bayes 的基本原理和应用方式 ② 明白 Naïve Bayes 的训练过程和其影响因素		
<b>硬件环境：</b> i5-6200U 8G RAM HD Graphics520		
<b>软件环境：</b> Octave		
<b>实验步骤与内容：</b> 通过学习朴素贝叶斯，使用其公式进行训练。 训练过程：遍历数据，统计每种特征的情况与当前类别同时出现的次数，除以当前类别的总次数得到的概率，组成一个全为概率的三维数组 测试过程：利用 Bayes 定理推导出的公式算出每个类别的概率，选出所有类别中概率最大的作为预测结果，与实际结果进行比较 拉普拉斯平滑：为防止 0 概率对于结果的影响，将每种特征的情况与当前类别同时出现的次数+1，除以（当前类别的总次数+类别数） 训练数据量与正确率如下：  本实验中拉普拉斯平滑对结果影响不大 所有数据均为随机方式从给定训练数据中提取		
<b>结论分析与体会：</b> 最近在学习 python，所以使用 python 来写了这个实验，发现在没有画图要求的情况下别的语言同样能进行一些机器学习项目。（好像 python 也有画图的库不过还不会用） 使用 Naïve Bayes 进行分类的时候，用较少的数据就能得到较高的正确率，随着数据量的增加，准确率会越来越高，但增高的速率会逐渐减慢。		

附录：程序源代码