

---

---

# 人工智能导论:作业三实验报告

薛正海(181220063、xuezh@smail.nju.edu.cn)

(南京大学 人工智能学院, 南京 210093)

## 1 机器学习方法介绍

### 1.1 朴素贝叶斯分类器

为解决分类问题,朴素贝叶斯分类器借助概率学中的贝叶斯公式,利用先验信息和条件概率来计算样本属于某一类别的后验概率 $p(C|F_1, \dots, F_n)$ <sup>[1]</sup>,其中 $C$ 为类别, $F_i$ 为各特征.朴素贝叶斯分类器的一个基本假设是样本的各特征之间独立,这一假设简化了贝叶斯公式中的条件概率计算.

具体而言,由贝叶斯公式,
$$p(C|F_1, \dots, F_n) = \frac{p(C) p(F_1, \dots, F_n|C)}{p(F_1, \dots, F_n)}$$
其中先验 $p(C)$ 可由样本中各类别出现的次数统计而得, $p(F_1, \dots, F_n|C)$ 由独立性的假设可化为各特征在类别的条件下的条件概率的连乘积,同样可由各样本统计而得.分母部分对各类别都相同,故可不加考虑.最后判断出的类别是各类别中后验概率最高的一类,这便是最大后验概率的决策准则.

### 1.2 Logistic Regression

Logistic Regression 试图拟合一个非线性函数,使得该函数的输出恰好是分类的结果.理想中最优的函数应是 0-1 跃迁函数,但此函数不可导.Sigmoid 函数是该函数的一个可导的替代品.直觉上,我们应令正样本的 Sigmoid 函数输出接近于 1,负样本的输出接近于 0.从极大似然估计的角度<sup>[2]</sup>,设 $t_i$ 为二分类的类别, $y_i$ 为分类函

数的输出结果,则我们应最大化 $\prod_{n=1}^N \{y_n^{t_n} * (1-y_n)^{(1-t_n)}\}$ .取对数,得 $\sum_{n=1}^N \{t_n \ln y_n + (1-t_n) \ln(1-y_n)\}$ ,这便是 Logistic Regression 的目标函数.注意 $t_i$ 为 1 时,若 $y_i$ 较小则受到较大的惩罚, $t_i$ 为 0 时则相反,与直觉相符.

从 KL 散度的角度<sup>[3]</sup>,极大似然估计实际上是最小化了拟合出的数据分布与数据的真实分布的 KL 散度.观察 Logistic Regression 的目标函数不难看出其实际计算了两个分布之间的交叉熵 $H(p, q) = -\sum p(x) \log q(x) = -y \log \hat{y} - (1-y) \log(1-\hat{y})$ ,而交叉熵 $H(p, q) = D_{KL}(p|q) + H(p)$ 与 KL 散度之间只误差真实分布的熵值,后者是一个常数.

### 1.3 随机森林

随机森林为了解决决策树的高方差问题,构造多棵决策树.每棵随机树的训练样本由原数据集进行 Bagging,即有放回采样生成.这样可保证训练样本的无偏性,同时为训练样本引入随机性,降低每棵树之间的相关性和整个随机森林的方差.在每棵树的训练中,特征也被随机分配.最终的分类结果由所有树进行投票决定.

## 2 方法性能对比

### 2.1 单个方法的超参数调节

1. 由朴素贝叶斯分类器的分类公式,其不存在可供调节的超参数.

2. Logistic Regression 可调节的超参数为 Ridge value,即正则化系数.在 80%样本作为训练集,其余 20%样本作为测试集时,不同 Ridge value 对应分类器的性能如下:

Ridge Value	准确率	Ridge Value	准确率
1e-8(默认)	65.97%	3e-2	74.31%
1e-5	72.22%	0.1	74.31%
<b>1e-3</b>	<b>75.00%</b>	<b>0.3</b>	<b>75.69%</b>
1e-2	75.00%	1.0	74.31%

可见,默认设置的 Ridge Value 易导致过拟合,在测试集上的准确率不高.总的来说,正则化系数在 1e-3 数量级左右时测试集准确率相对较高,与通常情况相符.值得注意的是正则化系数设置得较大时模型的准确率并未显著降低,这可能是因为训练数据中“发射”类样本较多,模型总是倾向于将未知样本分为“发射”类,分类较为容易.

3. 随机森林可调节的参数有决策树的棵数,每棵决策树的最大深度及可用特征的数量.

棵数	深度	可用特征	准确率	棵数	深度	可用特征	准确率
10	无限制	100	80.56%	30	无限制	50	77.78%
<b>30</b>	<b>无限制</b>	<b>100</b>	<b>81.94%</b>	30	无限制	100	81.94%
100	无限制	100	80.56%	<b>30</b>	<b>无限制</b>	<b>150</b>	<b>82.64%</b>
300	无限制	100	79.86%	30	无限制	200	80.56%
30	12	150	77.08%	30	12	100	80.55%
30	25	150	79.86%	30	25	100	81.25%
<b>30</b>	<b>50</b>	<b>150</b>	<b>82.64%</b>	<b>30</b>	<b>50</b>	<b>100</b>	<b>81.94%</b>
<b>30</b>	<b>100</b>	<b>150</b>	<b>82.64%</b>	<b>30</b>	<b>100</b>	<b>100</b>	<b>81.94%</b>

由实验数据可见,随机森林中树的棵数过少或过多时分类的准确率都会降低.棵数过少可能导致决策树的 Diversity 不足,投票效果不佳;棵数过多可能导致程序运行效率降低,在给定时限内无法对每棵树进行足够训练.类似地,可用特征过多或过少时分类的准确率也会降低.过多时可能导致每棵树的相关性增加,过少时会影响每棵树的训练.调节每棵决策树的深度可以防止随机森林发生过拟合,但实验中似乎未出现严重的过拟合,不同深度对实验结果影响不大.

2.2 多个方法的比较

朴素贝叶斯由于其“朴素”的性质,结果依赖于特征之间的独立性,但本环境中的特征不是互相独立的(空白像素的周围更倾向于有空白像素等),故朴素贝叶斯分类器的效果相对较差.

就 Logistic Regression 而言,其分类边界是线性的,且性能依赖于手动提取的高层特征的好坏.常见的 practice 是“特征工程+Logistic Regression”.但本环境中各特征都是低层次像素级的,相对而言维数较高,没有经过进一步特征提取,故普通的 Logistic Regression 效果相对较差.但随机森林拥有非线性建模的能力,对特征的要求较低,故在本环境中分类准确率最高.

3 修改特征提取方法

观察用原特征提取方法训练出的智能体的行为,我发现其只学出了不停发射炮弹的动作,对我尝试躲避 aliens 发出的子弹的左右移动没有完全模仿.这可能是因为特征只反映了当前帧中游戏的信息,并不能体现出子弹的下落过程.我想出的改进方法是借用 LSTM 的思想,将过去几帧的游戏画面加入特征之中.这样特征能反映出不同帧之间的联系,使模型可以对 aliens 发出的子弹进行建模,识别出有危害性的子弹和无危害性的 ImmovablePositions.在 featureExtract 中修改的代码如下:



---

以上就是我本次作业报告的全部内容,感谢您的阅读!

**References:**

- [1] [https://en.wikipedia.org/wiki/Naive\\_Bayes\\_classifier](https://en.wikipedia.org/wiki/Naive_Bayes_classifier)
- [2] <https://blog.csdn.net/daunxx/article/details/51816588>
- [3] Arjovsky, Martin, Soumith Chintala, and Léon Bottou. "Wasserstein gan." arXiv preprint arXiv:1701.07875 (2017).