

代码说明以及实验报告中的数据来源

2013012245 基科 31 白可

2015 年 10 月 8 日星期四

语言: Matlab

主要步骤是:

1. 通过训练集得到正态分布的参数并做图
2. 利用正态分布参数对训练集进行预测并得到正确率
3. 利用正态分布参数对测试集进行预测并得到正确率
4. 调整阈值, 计算不同阈值下 (2, 3) 中的正确率。

一共有两个文件 train.m 和 ROC.m, 各自可以独立运行

train.m 主要进行了实验一和二

ROC.m 主要进行了实验三

两个文件大致内容相同, ROC.m 增添了改变阈值的循环, 从而实现画出 ROC 的目的。

重要参数都已经写在了代码的注释中, 该文档将对可调参数进行说明。

程序中只有唯一一个需要调节的参数: li, 它代表了“阈值”, 以下详细说明:

实验一:

最小错误率贝叶斯决策:

$$l(x) = \frac{p(x|\omega_m)}{p(x|\omega_w)} > (<) \frac{P(\omega_w)}{P(\omega_m)}$$
$$li = \frac{P(\omega_w)}{P(\omega_m)}$$

在实验一中, 我分别令它取

1.48 (样本中女生数目/男生数目)

1 (假设男女概率相同)

0.5 (清华男女比 2:1)

1/6 (理工科院系男女比)

从而得到了不同的训练集正确率和测试集正确率

实验二:

最小风险贝叶斯决策:

$$l(x) = \frac{p(x|\omega_m)}{p(x|\omega_w)} > (<) \frac{P(\omega_w) \lambda_{12} - \lambda_{22}}{P(\omega_m) \lambda_{21} - \lambda_{11}}$$
$$li = \frac{P(\omega_w) \lambda_{12} - \lambda_{22}}{P(\omega_m) \lambda_{21} - \lambda_{11}}$$

在试验中我令 $\lambda_{12} = 4, \lambda_{21} = 1, \frac{P(\omega_w)}{P(\omega_m)} = 0.5$, 则 $li = 2$

结果中输出了, TP (预测男, 实际男) FP (预测男, 实际女) FN (预测女, 实际男), TN (预测女, 实际女) 四个数字, 以及真阳性率和假阳性率。

实验三:

其中 time 参数可调。在我的试验中, 我令 time 的取值为 0.01, 以 0.01 为步长到 2.5, 阈值是 time 的两倍。

如果需要画出比较好的图像。不同的实验有所不同, 需要对其进行调整。