

# 作业



## ■ 理论题：25%

1) 写出最小二乘求解如下广义线性模型的  $w, b$  详细推到过程

$$y = e^{wx+b}$$

2) 假设有三家工厂A, B, C共同生产一种台灯，他们产品占比和次品率分别如下：

工厂名 (类别名)	产品占比(先验概率)	次品率(条件概率)
A	$0.35 = P(A)$	$0.015 = P(\text{次品}   A)$
B	$0.35 = P(B)$	$0.010 = P(\text{次品}   B)$
C	$0.30 = P(C)$	$0.020 = P(\text{次品}   C)$

某次随机抽检一个样品，该样品是次品概率有多大？如果该样品是次品，则它来自工厂A, B, C 的概率分别有多大？

# 作业



- 实践题：75%=25% x 3
  - 1) 实现线性分类器并在西瓜3.0数据集上用前80%训练、后20%测试时的精度
  - 2) 实现Naïve Bayes分类器并在西瓜3.0数据集上测试k=5重交叉验证精度
  - 3) 比较SVM使用不同（至少4种）核函数时，西瓜3.0数据集上用前80%训练、后20%测试的精度（可使用任意svm算法实现软件包）
- 附加题：(可不做，做对20%额外分)

实现对数几率回归并在西瓜3.0上与线性分类器、NB和SVM做性能比较
- 注意：只允许实践题3)中SVM算法使用已有库中的实现，其他均需自己实现（**发现作弊或抄袭，本次作业0分处理**）。

# 作业提交



- 文档：包括理论题的回答和实践题的结果截图
- 代码：包括所有运行需要的代码和数据
- 把以上两部分压缩在一个压缩包，发送到邮箱wangzihao33@sjtu.edu.cn
- 邮件标题格式：课程ML作业1\_学号\_姓名
- DDL：第5周周日（10月16日）23:59之前（以邮件收到时间为准）