

## 习题二

1. 试分析在什么情况下线性模型

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b$$

不需要考虑偏置项  $b$ .

2. 令码长为 9，类别数为 4，给出海明距离下理论最优的 ECOC(纠错输出码, Error Correcting Output Codes) 二元码并证明.
3. 编程实现二分类情况下的线性判别分析 (Linear Discriminant Analysis)<sup>1</sup>, 给出作用于西瓜数据  $3.0\alpha$  上投影向量  $\mathbf{w}$  的结果; 同时在提供的两个 UCI 数据集 (fourclass.csv, heart.csv) 上进行测试<sup>2</sup>, 比较 10 折交叉验证和留一法, 在作业中分别汇报 10 折交叉验证和留一法得到的错误率的均值和标准差<sup>3</sup>.

---

<sup>1</sup>编程题需要提交完整的可执行代码, 代码相应的说明文档可以和题目答案目写在同一份文件中. 最终提交作业时, 请提交 pdf 格式的文档, 并和可执行代码一起压缩到一份 zip 文件中.

<sup>2</sup>题目中需要的数据见附加的压缩文件, 其中, UCI 数据文件的每一行代表一个数据, 数据的特征使用逗号分隔, 每一行最后的数字表示该样本的类别. 西瓜数据来自于《机器学习》教材 89 页表 4.5.

<sup>3</sup>[https://en.wikipedia.org/wiki/Cross-validation\\_\(statistics\)](https://en.wikipedia.org/wiki/Cross-validation_(statistics))