

## 练习题

1. 设训练样本数目为  $N$ ，原始观测空间的维数为  $m$ 。支持向量机所对应的二次规划问题可以使用 SMO(Sequential Minimal Optimization)算法求解。SMO 算法不断地求解仅涉及两个优化变量的二次规划问题。设给定训练样本为  $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$ ，对偶问题为：

$$\min_{\{\alpha_i\}_{i=1}^N} \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j - \sum_{i=1}^N \alpha_i \quad \text{s.t.} \quad \sum_{i=1}^N \alpha_i y_i = 0, \quad \alpha_i \geq 0, \text{ for } i = 1, \dots, N.$$

具体地，在每次迭代时，选中两个优化变量  $\alpha_{i^*}$  和  $\alpha_{j^*}$ ，同时保持其它变量固定，求解关于  $\alpha_{i^*}$  和  $\alpha_{j^*}$  的二次规划问题。请结合线性支持向量机（分为线性可分和线性不可分情况），给出在每次迭代过程中关于两个变量  $\alpha_{i^*}$  和  $\alpha_{j^*}$  的解。

2. 某公司提出某设计方案，拟通过征集网络调查确定喜欢该方案的用户真实比例：调查选项为“喜欢”和“不喜欢”，每个用户只能选择一个选项且只能提交一次结果。假设 1 个用户被邀请参与调查，公司需要支付 RMB 1.0 元。若公司希望能以至少 99.99% 的把握确保估计误差小于 1%。请尝试对调查成本给出预算。

3. 什么是经验风险最小化准则一致性的充要条件？经验风险最小化准则一致性的充要条件的意义是什么？

4. 什么是经验风险最小化准则一致收敛速度？试借助 Sauer 引理和 VC 不等式，推导经验风险最小化准则一致收敛速度的界。

5. 什么是学习速率？什么是学习速率的界？如何实现对学习问题中泛化误差的有效控制？