

# 机器学习第二次作业

2020年10月

## 1 带权重支持向量机

考虑带权重的支持向量机，其中  $C > 0$ ，权重常数  $\gamma_i \in (0, +\infty)$ ：

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|_2^2 + C \sum_{i=1}^n \gamma_i \xi_i$$

$$\text{s.t.} \quad y_i (\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, \xi_i \geq 0, 1 \leq i \leq n$$

- (1) 求该问题的Lagrange方程。
- (2) 求该问题的对偶形式 (Dual Form)。
- (3) 已知非线性映射  $\phi$  和其对应的核函数  $k(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{x}) \cdot \phi(\mathbf{y})$ ，若在上述带权重支持向量机中使用该核函数，请写出对应的原问题和对偶问题。

## 2 SVM实践

本题中我们使用SVM进行糖尿病识别，糖尿病识别数据集已提供在附件中。请完成以下几个任务：

- (1) 编程实现使用SGD优化的线性SVM和基于核函数的非线性SVM（例如基于高斯核），并在数据集上对它们进行训练与测试。（助教提供了线性SVM的代码框架供参考，也可自己从头实现）
- (2) 使用验证集数据对自己实现的SVM进行超参数调优，找到合适的超参数。
- (3) 调用sklearn中的SVM模型在数据集上进行训练与测试，调整核函数与超参数，与自己实现的结果进行比较分析。

### 2.1 代码要求

本题只允许使用python作为编程语言，在自己实现SVM的环节中不允许直接调用sklearn的SVM接口。

### 2.2 提交要求

本题需要提交实验代码并形成完整的实验报告。