

第4章 支持向量机2

Exercise 4.10

在线性SVM中，我们假设分类边界的形式为 $\mathbf{w}^T \mathbf{x} + b$ ，其中包含一个偏置项 $b \in \mathbb{R}$ 。但是，学习一个不含偏置项的SVM分类器，即使用 $\mathbf{w}^T \mathbf{x}$ 作为分类边界也是可行的。

- (a) 如果不含偏置项，原始空间中的优化问题是什么样的？
(b) 如果不含偏置项，给出其对偶形式，如果关于拉格朗日乘子 λ 的解是 λ^* ，则最优的决策边界 \mathbf{w}^* 是？
(c) 当希望存在偏置项时，如果希望上述方法仍然是有用的。给定一个训练数据集 $(\mathbf{x}_i, y_i) (1 \leq i \leq n)$ ，我们可将任意 $\mathbf{x} \in \mathbb{R}^d$ 转换为 \mathbb{R}^{d+1} 空间中的 $\hat{\mathbf{x}}$ ，这只需要通过对 \mathbf{x} 增加一个额外的维度即可，所增加的维度总有一个常数值1。假设 λ^* 是关于对偶形式的最优解，并假设分类边界为 $\mathbf{w}^T \mathbf{x} + b$ ，则最优的 b^* 值是多少

Exercise 4.11

Consider the general 2-nd polynomial kernel $K_2(\mathbf{x}; \mathbf{x}') = (\xi + \gamma \mathbf{x}^T \mathbf{x}')^2$. Which of the following transform can be used to derive this kernel?

- A. $\phi(\mathbf{x}) = (1, \sqrt{2\gamma}x_1, \dots, \sqrt{2\gamma}x_d, \gamma x_1^2, \dots, \gamma x_d^2)$
B. $\phi(\mathbf{x}) = (\xi, \sqrt{2\gamma}x_1, \dots, \sqrt{2\gamma}x_d, x_1^2, \dots, x_d^2)$
C. $\phi(\mathbf{x}) = (\xi, \sqrt{2\gamma\xi}x_1, \dots, \sqrt{2\gamma\xi}x_d, x_1^2, \dots, x_d^2)$
D. $\phi(\mathbf{x}) = (\xi, \sqrt{2\gamma\xi}x_1, \dots, \sqrt{2\gamma\xi}x_d, \gamma x_1^2, \dots, \gamma x_d^2)$

Exercise 4.12

At the optimal solution of

$$\begin{aligned} \min \quad & \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^N \xi_i \\ \text{s. t.} \quad & y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \xi_i \geq 0 \quad (i = 1, 2, \dots, N) \end{aligned}$$

assume that $y_1(\mathbf{w}^T \mathbf{x}_1 + b) = -10$. What is the corresponding ξ_1 ?

- A. 1
B. 11
C. 21
D. 31

实践题

Exercise 4.1

汽车状态预测

车辆的状态分为四类: * unacc (Unacceptable 状况很差) * acc (Acceptable 状况一般) * good (Good 状况好) * vgood (Very good 状况非常好)

那我们又是通过什么来判断这辆车的状态好坏呢? * buying (购买价: vhigh, high, med, low) * maint (维护价: vhigh, high, med, low) * doors (几个门: 2, 3, 4, 5more) * persons (载人数: 2, 4, more) * lug_boot (贮存空间: small, med, big) * safety (安全性: low, med, high)

请使用支持向量机模型去学习给出的数据，然后通过学到的模型去预测测试集数据？

数据集说明：

data_train： 训练集

data_test： 测试集

description: 数据说明

地址： https://github.com/jjw12345/machine_learning/tree/main/svm

提交说明：

最后提交代码和测试集的预测数据。