Machine Learning and Pattern Recognition

东北大学“智能+X”新工科课程系列

机器学习与模式识别

东北大学 信息科学与工程学院
人工智能系 、智能感知与机器人研究所
陈东岳

Solving Linear SVM

求解线性SVM

CHAPTER ONE

SVM对偶问题推导

Derivation of SVM Dual Problem

第一节 SVM对偶问题推导

SVM问题数学描述

$$\min_{\boldsymbol{\omega},\gamma} \frac{1}{2}\|\boldsymbol{\omega}\|^2$$

$$\text{s.t.} \quad y_i\left(\boldsymbol{\omega}^T \boldsymbol{x}_i + \gamma\right) \geq 1, \qquad i = 1,2,\dots,m$$

SVM拉格朗日乘子式

$$L(\omega,\gamma,\boldsymbol{\alpha}) = \frac{\|\boldsymbol{\omega}\|^2}{2} + \sum_{i=1}^{m} \alpha_i(1 - y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + \gamma))$$

$$\min_{\boldsymbol{\omega},\gamma}\left[\max_{\boldsymbol{\alpha}:\alpha_i\geq 0} L(\boldsymbol{\omega},\gamma,\boldsymbol{\alpha})\right] \qquad \max_{\boldsymbol{\alpha}:\alpha_i\geq 0}\left[\min_{\boldsymbol{\omega},\gamma} L(\boldsymbol{\omega},\gamma,\boldsymbol{\alpha})\right]$$

原始问题　　　　对偶问题

第一节 SVM对偶问题推导

SVM拉格朗日对偶问题求解

- 首先求解：

$$\min_{\boldsymbol{\omega},\gamma} L(\boldsymbol{\omega},\gamma,\boldsymbol{\alpha}) = \min_{\boldsymbol{\omega},\gamma}\left[\frac{1}{2}\|\boldsymbol{\omega}\|^2 + \sum_{i=1}^{m}\alpha_i\left(1-y_i(\boldsymbol{\omega}^T\boldsymbol{x}_i+\gamma)\right)\right]$$

- 分别令函数$L(\omega,\gamma,\boldsymbol{\alpha})$ 对$\boldsymbol{\omega},\gamma$ 求偏导，并使其等于0。

$$\frac{\partial L}{\partial \omega} = \mathbf{0} \Rightarrow \boldsymbol{\omega} = \sum_{i=1}^{m}\alpha_i y_i \boldsymbol{x}_i \qquad \frac{\partial L}{\partial \gamma} = 0 \Rightarrow 0 = \sum_{i=1}^{m}\alpha_i y_i$$

- 整理上式可得：

$$\min_{\boldsymbol{\omega},\gamma} L(\boldsymbol{\omega},\gamma,\boldsymbol{\alpha}) = \boxed{\frac{1}{2}\left(\sum_{i=1}^{m}\alpha_i y_i \boldsymbol{x}_i\right)^T\left(\sum_{i=1}^{m}\alpha_i y_i \boldsymbol{x}_i\right)} + \sum_{i=1}^{m}\alpha_i - \boxed{\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j}$$

$$= \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j$$

第一节 SVM对偶问题推导

▶ 线性SVM的拉格朗日对偶问题

变形后的优化问题：

$$\max_{\boldsymbol{\alpha}}\left[\sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j\right]$$

约束条件为：

$$\sum_{i=1}^{m}\alpha_i y_i=0$$

$$\alpha_i\geq 0, i=1,2,\dots,m$$

CHAPTER THREE

求解对偶问题

Solving Dual Problem

第三节 求解对偶问题

▶ 重写对偶问题

$$\max_{\alpha}\left[\sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j\right]$$

$$s.t.\ \ \sum_{i=1}^{m}\alpha_i y_i=0,\ \alpha_i\geq 0,i=1,2,\dots,m$$

▶ 最优解必须满足原始问题的KKT条件

$$\begin{cases}\alpha_i\geq 0\\ y_i\left(\boldsymbol{\omega}^T\boldsymbol{x}_i+\gamma\right)-1\geq 0\\ \alpha_i\left(y_i\left(\boldsymbol{\omega}^T\boldsymbol{x}_i+\gamma\right)-1\right)=0\end{cases},i=1,2,\dots,m$$

第三节 求解对偶问题

▶ 如何求解最优解α_i——SMO算法

- 基本思路：

  - 该算法以满足KKT条件为目标，将多个变量的综合优化问题，转化为一系列单个变量优化的迭代问题。

- 算法过程：

- 1) 随机或等值初始化α_i，计算$\boldsymbol{\omega}$和γ；
- 2) 挑选两个破坏KKT条件最严重的乘子α_k和α_l进行优化，其他乘子不变；
- 3) 根据条件$\sum_{i=1}^{m} \alpha_i y_i = 0$，将$\alpha_l$转换为$\alpha_k$的函数；
- 4) 通过求取目标函数关于α_k的导数来计算α_k和α_l的封闭解；
- 5) 更新$\alpha_k, \alpha_l, \boldsymbol{\omega}$和$\gamma$，重复步骤2)和5)直到满足KKT条件，停止。

第三节 求解对偶问题

▶ 如何求解最优解α_i ——SMO算法

-1) 随机或等值初始化α_i，计算$\boldsymbol{\omega}$和 γ

$$\frac{\partial L}{\partial \boldsymbol{\omega}} = \mathbf{0} \Longrightarrow \boldsymbol{\omega} = \sum_{i=1}^{m} \alpha_i y_i \boldsymbol{x}_i \qquad \alpha_i\left(y_i\left(\boldsymbol{\omega}^T \boldsymbol{x}_i + \gamma\right) - 1\right) = 0$$

$$\boldsymbol{\omega} = \sum_{i=1}^{m} \alpha_i y_i \boldsymbol{x}_i \qquad y_s\left(\boldsymbol{\omega}^T \boldsymbol{x}_s + \gamma\right) - 1 = 0$$

$$S = \{i | \alpha_i > 0, i = 1,2, \dots, m\} \qquad \gamma = \frac{1}{|S|} \sum_{s \epsilon S} [y_s - \boldsymbol{\omega}^T \boldsymbol{x}_s]$$

第三节 求解对偶问题

▶ 如何求解最优解α_i ——SMO算法

- 2) 挑选两个破坏KKT条件最严重的乘子α_k和α_l进行优化

$$\begin{cases} \alpha_i \geq 0 \\ y_i\left(\boldsymbol{\omega}^T \boldsymbol{x}_i + \gamma\right) - 1 \geq 0 \\ \alpha_i\left(y_i\left(\boldsymbol{\omega}^T \boldsymbol{x}_i + \gamma\right) - 1\right) = 0 \end{cases} \quad , i = 1,2, \dots, m$$

$$\boldsymbol{\omega} = \sum_{i=1}^{m} \alpha_i y_i \boldsymbol{x}_i$$

$$\gamma = \frac{1}{|S|} \sum_{s \in S} \left[y_s - \boldsymbol{\omega}^T \boldsymbol{x}_s\right]$$

第三节 求解对偶问题

▶ 如何求解最优解α_i ——SMO算法

- 3)根据条件 $\sum_{i=1}^{m}\alpha_i y_i = 0$ ，将α_l转换为α_k的函数；

$$\frac{\partial L}{\partial \gamma} = 0 \Rightarrow \sum_{i=1}^{m}\alpha_i y_i = 0 \Rightarrow \alpha_k y_k + \alpha_l y_l = c \Rightarrow \alpha_l = y_l(c - \alpha_k y_k)$$

$$\theta_D(\alpha) = \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j$$

$$D_{k,l} = \sum_{i\neq k,l}^{m}\sum_{j\neq k,l}^{m}\alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j$$

$$\theta_D(\alpha) = C_{k,l} + \alpha_k + \alpha_l - \frac{1}{2}\left[D_k\alpha_k + D_l\alpha_l + 2\alpha_k\alpha_l y_k y_l \boldsymbol{x}_k^T\boldsymbol{x}_l + D_{k,l}\right]$$

$$C_{k,l} = \sum_{i\neq k,l}\alpha_i$$

$$D_k = \sum_{j\neq l}\alpha_j y_k y_j \boldsymbol{x}_k^T\boldsymbol{x}_j$$

$$D_l = \sum_{i\neq k}\alpha_i y_i y_l \boldsymbol{x}_i^T\boldsymbol{x}_l$$

▶ 如何求解最优解α_i——SMO算法

- 4)通过求取目标函数关于α_k的导数来计算α_k和α_l的封闭解；

  - ➢ 求取目标函数关于α_k的导数

  - ➢ 令导数为0，求α_k和α_l的最优解

  - ➢ $\theta_D(\boldsymbol{\alpha})=\theta_D(\alpha_k)$是$\alpha_k$的二次函数，导数$\frac{\partial \theta_D(\alpha_k)}{\partial \alpha_k}=0$具有封闭解。

  - ➢ 将封闭解作为$\alpha_k(t+1)$和$\alpha_l(t+1)$数值

第三节 求解对偶问题

▶ 如何求解最优解α_i ——SMO算法

- 5) 更新$\alpha_k, \alpha_l, \boldsymbol{\omega}$和$\gamma$，重复步骤2)和5)直到满足KKT条件，停止。

  - 在第$t+1$次迭代时，使用新的数值$\alpha_k(t+1)$和$\alpha_l(t+1)$来代替旧的数值$\alpha_k(t)$和$\alpha_l(t)$，更新$\boldsymbol{\omega}$和 γ：

$$\boldsymbol{\omega} = \sum_{i=1}^{m} \alpha_i y_i \boldsymbol{x}_i$$

$$\gamma = \frac{1}{|S|}\sum_{s \epsilon S}[y_s - \boldsymbol{\omega}^T \boldsymbol{x}_s]$$

$$S = \{i|\alpha_i > 0, i = 1,2,\dots,m\}$$

  - 终止条件即要求KKT条件在精度ϵ内得到满足。

$$\begin{cases} \alpha_i \geq 0 \\ y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + \gamma) - 1 \geq 0 \\ \alpha_i(y_i(\boldsymbol{\omega}^T \boldsymbol{x}_i + \gamma) - 1) = 0 \end{cases} ,i = 1,2,\dots,m$$

总 结

- 线性SVM求解
  - 对偶问题推导
    - SVM原始问题描述
    - SVM拉格朗日乘子式
    - 求取对偶问题
  - 对偶问题求解
    - 重写对偶问题
    - 讨论KKT条件
    - SMO算法
      - 基本思路
      - 算法步骤

M L & P R

THANK YOU

感谢聆听

MADE BY DONGYUE CHEN