机器学习作业

郭英明

2020/10/26

1

写出下列线性规划问题的对偶问题

$$\min_{x\in\mathbb{R}^2} c^T x$$

$$s.t. \bar{A}x - \bar{b} \leq 0, \tag{1}$$

$$Ax - b = 0 \tag{2}$$

其中，$c \in \mathbb{R}^2, \bar{A} \in \mathbb{R}^{m*n}, \bar{b} \in \mathbb{R}^m, A \in \mathbb{R}^{p*n}, b \in \mathbb{R}^p$

解答: 对应的拉格朗日函数为:

$$L(x,\lambda,\upsilon) = c^T x + \lambda^T(\bar{A}x - \bar{b}) + \upsilon^T(Ax - b)$$

拉格朗日函数的最小值函数为:

$$g(\lambda,\upsilon) = \inf_x L(x,\lambda,\upsilon) \tag{3}$$

$$= \inf_x (c + \bar{A}^T\lambda + A^T\upsilon)^T x - \bar{b}^T\lambda - b^T\upsilon \tag{4}$$

所以可以得到:

$$g(\lambda,\upsilon) = \begin{cases} -\bar{b}^T\lambda - b^T\upsilon, c + \bar{A}^T\lambda + A^T\upsilon = 0 \\ -\infty, others \end{cases} \tag{5}$$

所以对偶问题可以写成:

$$\max_{\lambda,\upsilon} -\bar{b}^T\lambda - b^T\upsilon$$

$$\boldsymbol{s.t.} c + \bar{A}^T\lambda + A^T\upsilon = 0, \tag{6}$$

$$\lambda \geq 0 \tag{7}$$

2

软间隔支持向量机还可以定义为一下形式：

$$\min_{\omega,b,\xi} \frac{1}{2}\left\|\boldsymbol{\omega}\right\|^2 + C\sum_{i=1}^m \xi_i^2$$
$$\boldsymbol{s.t.} y_i(\boldsymbol{\omega}^T\boldsymbol{x}_i + b) \geq 1-\xi_i, i = 1, \cdots, m.$$

试求其对偶形式，写出其KKT条件，并根据KKT条件讨论训练样本的分布情况

解答:

对应的拉格朗日函数为：

$$L(\omega, b, \xi, \lambda) = \frac{1}{2}\left\|\boldsymbol{\omega}\right\|^2 + C\sum_{i=1}^m \xi_i^2 + \sum_{i=1}^m \lambda_i(1-\xi_i - y_i(\boldsymbol{\omega}^T\boldsymbol{x}_i + b)$$

对于上述拉格朗日函数，若要求得其最小值函数，对于$\boldsymbol{\omega}, b, \xi_i$分别求偏导数并使之等于0得到如下式子：

$$\frac{\partial L}{\partial \boldsymbol{\omega}} = \boldsymbol{\omega} - \sum_{i=1}^m \lambda_i y_i \boldsymbol{x}_i = 0$$
$$\frac{\partial L}{\partial b} = \sum_{i=1}^m \lambda_i y_i = 0$$
$$\frac{\partial L}{\partial \xi_i} = 2C\xi_i - \lambda_i = 0$$

将上述式子代入拉格朗日函数即可得对偶问题为：

$$\max_{\lambda} \sum_{i=1}^m \lambda_i - \frac{1}{2}\sum_{i=1}^m\sum_{j=1}^m \lambda_i\lambda_j yiyj\boldsymbol{x}^T\boldsymbol{x} - \frac{1}{4C}\sum_{i=1}^m \lambda_i^2$$

$$\boldsymbol{s.t.} \sum_{i=1}^m \lambda_i yi = 0, \tag{8}$$

$$\lambda \geq 0 \tag{9}$$

其KKT条件如下：

$$\begin{cases} \lambda_i \geq 0 \\ 1-\xi_i - y_i(\boldsymbol{\omega}^T\boldsymbol{x}_i + b \leq 0 \\ \lambda_i(1-\xi_i - y_i(\boldsymbol{\omega}^T\boldsymbol{x}_i + b) = 0 \end{cases} \tag{10}$$

训练样本的分布情况如下：

$\lambda_i = 0$时 $y_i(\boldsymbol{\omega}^T\boldsymbol{x}_i + b > 1-\xi_i$，则该样本不会对$f(x)$有任何影响

$\lambda_i > 0$时 $y_i(\boldsymbol{\omega}^T\boldsymbol{x}_i + b = 1-\xi_i$，则该样本是支持向量。

当$0 < \xi < 1$即$0 < \frac{\lambda_i}{2C} < 1$时，则该样本落在最大间隔内部

当$\xi > 1$即$\frac{\lambda_i}{2C} > 1$时，则该样本被错误分类