统计(机器)学习 Homework-7

(1) 线性支持向量机还可以定义为以下形式:

$$\min_{w,b,\xi} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i^2$$

$$\text{s.t.} \quad y_i(w\cdot x_i + b) \geqslant 1-\xi_i, \quad i=1,2,\cdots,N$$

$$\xi_i \geqslant 0, \quad i=1,2,\cdots,N$$

试求其对偶形式.

解: 写出 Lagrange 函数:

$L(w,b,\xi,\alpha_i,\mu_i)$

$$= \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i^2 - \sum_{i=1}^{N}\alpha_i\{y_i(w\cdot x_i+b)-1+\xi_i\} - \sum_{i=1}^{N}\mu_i\xi_i$$

其中 $\alpha_i \geqslant 0$, $\mu_i \geqslant 0$

首先关于 w, b, ξ 进行极小化:

$$\nabla_w L = w - \sum_{i=1}^{N}\alpha_i y_i x_i = 0$$

$$\nabla_b L = -\sum_{i=1}^{N}\alpha_i y_i = 0$$

$$\nabla_{\xi_i} L = 2C\xi_i - \alpha_i - \mu_i = 0$$

代入 $L(w,b,\xi_i,\alpha_i,\mu_i)$ 可得:

$$\min_{w,b,\xi_i} L = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j x_i\cdot x_j - \sum_{i=1}^{N}\frac{(\alpha_i+\mu_i)^2}{4C} + \sum_{i=1}^{N}\alpha_i$$

于是对偶问题为:

$$\max_{\alpha_i, \mu_i} -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j x_i\cdot x_j - \sum_{i=1}^{N}\frac{(\alpha_i+\mu_i)^2}{4C} + \sum_{i=1}^{N}\alpha_i$$

$$\text{s.t.} \quad \sum_{i=1}^{N}\alpha_i y_i = 0$$

$$2C\xi_i - \alpha_i - \mu_i = 0$$

$$\alpha_i \geqslant 0$$

$$\mu_i \geqslant 0$$

$\square$

(2) 证明内积的正整数幂函数:

$$K(x,z) = (x\cdot z)^p$$

是正定核函数，这里 p 是正整数, $x, z \in \mathbf{R}^n$.

证明: 首先讨论 $p=1$ 的情况:

若 $p=1$, 则 $K(x,z) = x\cdot z$, 记此时Gram矩阵为 $K^{(1)}$

对于 $\forall\ x_1, \cdots, x_m \in \mathbb{R}^n$

Gram矩阵满足 $K^{(1)}_{ij} = x_i\cdot x_j$

对于 $\forall\ \mu_1, \cdots, \mu_m \in \mathbb{R}$, 记 $\vec{\mu} = (\mu_1, \cdots, \mu_m)^T$

$\vec{\mu}^T\cdot K^{(1)}\vec{\mu}$

$= \sum_{i,j}\mu_i\mu_j\cdot(x_i\cdot x_j)$

$= (\mu_1 x_1 + \cdots + \mu_m x_m)^T \cdot (\mu_1 x_1 + \cdots + \mu_m x_m)$

$= \| \mu_1 x_1 + \cdots + \mu_m x_m \|_2^2 \geq 0$

$\Rightarrow K^{(1)}$ 半正定

则 $p=1$ 时结论成立

当 $p>1$ 时，先证下述引理：

Lemma：若 $A \succeq 0$，$B \succeq 0$，则 $C = A \circ B \succeq 0$

其中 C 为 A 和 B 按元素进行点积（Hadamard积）

证明：$x^T (A \circ B) x$

$$= \sum_i \sum_j x_i x_j a_{ij} b_{ij}$$

$$= \sum_i \sum_j x_i a_{ij} x_j b_{ij}$$

$$= tr(diag(x) \cdot A \, diag(x) \cdot B)$$

其中 $diag(x) = \begin{bmatrix} x_1 & & 0 \\ & \ddots & \\ 0 & & x_m \end{bmatrix}$

由 trace 的性质：

$tr(diag(x) \cdot A \cdot diag(x) \cdot B)$

$= tr(B^{1/2} \, diag(x) \cdot A \, diag(x) \cdot B^{1/2})$

其中 $B = U\Lambda U^T$，$B^{1/2} = U\Lambda^{1/2}U^T$

令 $C = B^{1/2}\,\mathrm{diag}(x)$

则 $C^T = \mathrm{diag}(x)\cdot B^{1/2}$

根据惯性定理：

$$A \succeq 0 \Rightarrow CAC^T \succeq 0$$

由于 $CAC^T \succeq 0$，则 CAC^T 的特征值非负

$\Rightarrow \mathrm{tr}(B^{1/2}\mathrm{diag}(x)A\,\mathrm{diag}(x)B^{1/2}) = \mathrm{tr}(CAC^T) \geq 0$

$\Rightarrow x^T(A \circ B)x \geq 0$

$\Rightarrow C = A \circ B \succeq 0$

引理证毕

现假设 $p < J$ 时结论均成立

当 $p = J$ 时：

$$K(x,z) = (x\cdot z)^J = (x\cdot z)^{J-1}\cdot(x\cdot z)$$

记 $p=j$ 时 Gram 矩阵为 $K^{(j)}$，则有：

$$K^{(J)} = K^{(J-1)} \circ K^{(1)}$$

根据归纳假设，$K^{(J-1)} \succeq 0$，而 $p=1$ 时有 $K^{(1)} \succeq 0$

再由 Lemma 可知，$K^{(J-1)} \circ K^{(1)} \succeq 0$，即 $K^{(J)} \succeq 0$

因此对于 $\forall p \geq 1, p \in \mathbb{Z}$，都有 $K^{(p)} \succeq 0$，即 $K(x,z)$ 是正定核函数。 $\square$