统计机器学习 Homework 06

姓名: 雍崔扬
学号: 21307140051

Problem 1

给定一个二分类训练集 $D_{\text{train}} = \{x^{(i)}, y^{(i)}\}_{i=1}^n$，其中 $y^{(i)} \in \{-1, 1\}$ 而 $x^{(i)} \in \mathbb{R}^d$
考虑软阈值支持向量机:

$$\begin{aligned}
\min_{w,b,\xi} \quad & \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^n \xi_i^2 \\
\text{s.t.} \quad & \xi \succeq 1_n - y \odot (w^{\mathrm{T}} x^{(i)} + b) \\
& \xi \succeq 0_n
\end{aligned} \tag{P}$$

试求其对偶问题.

Solution:

记 $\begin{cases} y = [y^{(1)}, \ldots, y^{(n)}]^{\mathrm{T}} \in \mathbb{R}^n \\ X = [x^{(1)}, \ldots, x^{(n)}] \in \mathbb{R}^{d\times n} \\ \xi = [\xi_1, \ldots, \xi_n]^{\mathrm{T}} \in \mathbb{R}^n \end{cases}$

定义 **Lagrange 函数**为:

$$\begin{aligned}
L(w, b, \xi, \lambda, \nu) &:= \frac{1}{2}\|w\|_2^2 + C\|\xi\|^2 + \lambda^{\mathrm{T}}(1_n - y \odot (X^{\mathrm{T}} w + b1_n) - \xi) - \nu^{\mathrm{T}}\xi \\
&= \frac{1}{2}\|w\|_2^2 + (1_n - by)^{\mathrm{T}}\lambda - \lambda^{\mathrm{T}}(\mathrm{diag}(y)X^{\mathrm{T}})w + (C\xi - \lambda - \nu)^{\mathrm{T}}\xi
\end{aligned}$$

其中 $\lambda, \nu \in \mathbb{R}_+^n$ 为 Langrange 乘子.

KKT 条件为:

$$\begin{cases}
\nabla_w L(w, b, \xi, \lambda, \nu) = w - X\mathrm{diag}(y)\lambda = 0_d \\
\nabla_b L(w, b, \xi, \lambda, \nu) = -y^{\mathrm{T}}\lambda = 0 \\
\nabla_\xi L(w, b, \xi, \lambda, \nu) = 2C\xi - \lambda - \nu = 0_n \\
\xi \succeq 1_n - y \odot (w^{\mathrm{T}} x^{(i)} + b) \\
\xi \succeq 0_n \\
\lambda \succeq 0_n \\
\nu \succeq 0_n \\
\lambda_i(1 - y^{(i)}(w^{\mathrm{T}} x^{(i)} + b) - \xi_i) = 0 \ (i = 1, \ldots, n) \\
\nu_i \xi_i = 0 \ (i = 1, \ldots, n)
\end{cases}$$

定义 **Lagrange 对偶函数**为:

$$\begin{aligned}
\Gamma(\lambda, \nu) &:= \inf_{w,b,\xi} L(w, b, \xi, \lambda, \nu) \\
&= \inf_{w,b,\xi} \left\{ \frac{1}{2}\|w\|_2^2 + (1_n - by)^{\mathrm{T}}\lambda - \lambda^{\mathrm{T}}(\mathrm{diag}(y)X^{\mathrm{T}})w + (C\xi - \lambda - \nu)^{\mathrm{T}}\xi \right\} \quad (\text{substitute } \begin{cases} w = X\mathrm{diag}(y)\lambda \\ y^{\mathrm{T}}\lambda = 0_n \\ 2C\xi - \lambda - \nu = 0_n \end{cases}) \\
&= 1_n^{\mathrm{T}}\lambda + \frac{1}{2}\|X\mathrm{diag}(y)\lambda\|_2^2 - \lambda^{\mathrm{T}}(\mathrm{diag}(y)X^{\mathrm{T}})X\mathrm{diag}(y)\lambda - \frac{1}{4C}\|\lambda + \nu\|^2 \\
&= 1_n^{\mathrm{T}}\lambda - \frac{1}{2}\|X\mathrm{diag}(y)\lambda\|_2^2 - \frac{1}{4C}\|\lambda + \nu\|^2
\end{aligned}$$

于是我们得到软间隔的**对偶问题** (D-punished):

$$\begin{aligned}
\max_{\lambda,\nu} \quad & \Gamma(\lambda, \nu) := 1_n^{\mathrm{T}}\lambda - \frac{1}{2}\|X\mathrm{diag}(y)\lambda\|_2^2 - \frac{1}{4C}\|\lambda + \nu\|^2 \\
\text{s.t.} \quad & y^{\mathrm{T}}\lambda = 0 \\
& \lambda \succeq 0_n \\
& \nu \succeq 0_n
\end{aligned} \tag{D}$$

Problem 2

通常所说的核函数 $k(\cdot,\cdot)$ 默认是**正定核函数** (positive definite kernel function)
一个对称函数 $k:\mathcal{X}\times\mathcal{X}\mapsto\mathbb{R}$ 构成正定核函数的充要条件如下:
(Mercer 定理, 统计学习方法, 定理 7.5)
设 $k:\mathcal{X}\times\mathcal{X}\mapsto\mathbb{R}$ 是对称函数 (即满足 $k(z,x)=k(x,z)$ $(\forall\ x,z\in\mathcal{X})$)
则 $k(\cdot,\cdot)$ 为正定核函数当且仅当对于任意正整数 $m\in\mathbb{Z}_+$ 和 $x_1,\dots,x_m\in\mathcal{X}$ 都有 Gram 矩阵半正定.
其中 $x_1,\dots,x_m\in\mathcal{X}$ 的 Gram 矩阵的定义如下:

$$K:=[k(x_i,x_j)]_{i,j=1}^m$$

假设 $k_1(\cdot,\cdot)$ 和 $k_2(\cdot,\cdot)$ 都是正定核函数，试根据 Mercer 定理证明:

- ① 对于任意 $\alpha_1,\alpha_2\geq 0$，由 $k(x,z):=\alpha_1k_1(x,z)+\alpha_2k_2(x+z)$ 定义的对称函数都是正定核函数
- ② 对于 $\mathcal{X}$ 上的任意实值函数 $f(\cdot)$，由 $k(x,z):=f(x)f(z)$ 定义的对称函数都是正定核函数
- ③ 由 $k(x,z)=k_1(x,z)k_2(x,z)$ 定义的对称函数是正定核函数
- ④ 对于 $\mathcal{X}$ 上的任意实值函数 $f(\cdot)$，由 $k(x,z)=f(x)k_1(x,z)f(z)$ 定义的对称函数都是正定核函数

Part (1)

假设 $k_1(\cdot,\cdot)$ 和 $k_2(\cdot,\cdot)$ 都是正定核函数，试根据 Mercer 定理证明:
对于任意 $\alpha_1,\alpha_2\geq 0$，由 $k(x,z):=\alpha_1k_1(x,z)+\alpha_2k_2(x+z)$ 定义的对称函数都是正定核函数

Proof:
任意给定正整数 $m\in\mathbb{Z}_+$ 和 $x_1,\dots,x_m\in\mathcal{X}$，定义:

$$\begin{aligned}
K_1&:=[k_1(x_i,x_j)]_{i,j=1}^m\\
K_2&:=[k_2(x_i,x_j)]_{i,j=1}^m\\
K&:=[k(x_i,x_j)]_{i,j=1}^m\\
&=[\alpha_1k_1(x_i,x_j)+\alpha_2k_2(x_i,x_j)]_{i,j=1}^m\\
&=\alpha_1K_1+\alpha_2K_2
\end{aligned}$$

对于任意 $z\in\mathcal{X}$，我们都有:

$$\begin{aligned}
z^{\mathrm{T}}Kz&=z^{\mathrm{T}}(\alpha_1K_1+\alpha_2K_2)z\\
&=\alpha_1z^{\mathrm{T}}K_1z+\alpha_2z^{\mathrm{T}}K_2z\quad(\text{note that }\alpha_1,\alpha_2\geq 0)\\
&\geq 0
\end{aligned}$$

这意味着 K 半正定.
根据 Mercer 定理可知由 $k(x,z):=\alpha_1k_1(x,z)+\alpha_2k_2(x+z)$ 定义的对称函数是正定核函数.

Part (2)

假设 $k_1(\cdot,\cdot)$ 和 $k_2(\cdot,\cdot)$ 都是正定核函数，试根据 Mercer 定理证明:
对于 $\mathcal{X}$ 上的任意实值函数 $f(\cdot)$，由 $k(x,z):=f(x)f(z)$ 定义的对称函数都是正定核函数.

Proof:
任意给定正整数 $m\in\mathbb{Z}_+$ 和 $x_1,\dots,x_m\in\mathcal{X}$，定义:

$$\begin{aligned}
y&:=[f(x_1),\dots,f(x_m)]^{\mathrm{T}}\\
K&:=[k(x_i,x_j)]_{i,j=1}^n\\
&=[f(x_i)f(x_j)]_{i,j=1}^n\\
&=yy^{\mathrm{T}}
\end{aligned}$$

对于任意 $z\in\mathcal{X}$，我们都有:

$$\begin{aligned}
z^{\mathrm{T}}Kz&=z^{\mathrm{T}}[k(x_i,x_j)]_{i,j=1}^mz\\
&=z^{\mathrm{T}}yy^{\mathrm{T}}z\\
&=(y^{\mathrm{T}}z)^2\\
&\geq 0
\end{aligned}$$

这意味着 K 半正定.
根据 Mercer 定理可知由 $k(x,z) := f(x)f(z)$ 定义的对称函数是正定核函数.

Part (3)

假设 $k_1(\cdot,\cdot)$ 和 $k_2(\cdot,\cdot)$ 都是正定核函数，试根据 Mercer 定理证明:
由 $k(x,z) = k_1(x,z)k_2(x,z)$ 定义的对称函数是正定核函数.

- **Definition:**
  元素对应相乘的乘积 (entrywise product) 称为 **Hadamard 乘积**，也称为 **Schur 乘积**
  给定 $A,B \in \mathbb{C}^{m\times n}$，记 A,B 的 Hadamard 乘积为 $A \odot B := [a_{ij}b_{ij}]_{i,j=1}^{m,n} \in \mathbb{C}^{m\times n}$
- **Lemma 1: (Matrix Analysis 引理 $7.5.2$)**
  对于任意 $A,B \in \mathbb{C}^{n\times n}$ 和 $x,y \in \mathbb{C}^n$，我们都有:

  $$x^{\mathrm{H}}(A\odot B)y = \mathrm{tr}\left(\mathrm{diag}(\bar{x})A\cdot \mathrm{diag}(y)B^{\mathrm{T}}\right)$$

  Proof:
  注意到:

  $$\begin{aligned} \mathrm{diag}(\bar{x})A &= [\bar{x}_i a_{ij}]_{i,j=1}^n \\ \mathrm{diag}(y)B^{\mathrm{T}} &= [y_i b_{ji}]_{i,j=1}^n \end{aligned}$$

  于是我们有:

  $$\begin{aligned} \mathrm{tr}\left(\mathrm{diag}(\bar{x})A\cdot \mathrm{diag}(y)B^{\mathrm{T}}\right) &= \sum_{i=1}^n\sum_{j=1}^n [\mathrm{diag}(\bar{x})A]_{i,j}[\mathrm{diag}(y)B^{\mathrm{T}}]_{j,i} \\ &= \sum_{i=1}^n\sum_{j=1}^n (\bar{x}_i a_{ij})(y_j b_{ij}) \\ &= \sum_{i=1}^n\sum_{j=1}^n \bar{x}_i (a_{ij}b_{ij}) y_j \\ &= x^{\mathrm{H}}(A\odot B)y \end{aligned}$$

- **Lemma 2: (Schur 乘积定理, Matrix Analysis 定理 $7.5.3$)**
  设 $A,B \in \mathbb{C}^{n\times n}$ 为 Hermite 阵:
  若 A,B 半正定，则 $A\odot B$ 半正定.

  Proof:
  假设 A,B Hermite 半正定.
  对于任意 $x\in\mathbb{C}^n$，记 $C := \mathrm{diag}(x)(B^{\mathrm{T}})^{\frac{1}{2}} = \mathrm{diag}(x)\bar{B}^{\frac{1}{2}}$
  则根据 **Lemma 1** 我们有:

  $$\begin{aligned} x^{\mathrm{H}}(A\odot B)x &= \mathrm{tr}\left(\mathrm{diag}(\bar{x})A\cdot \mathrm{diag}(x)B^{\mathrm{T}}\right) \\ &= \mathrm{tr}\left(\mathrm{diag}(\bar{x})A\cdot \mathrm{diag}(x)\bar{B}\right) \\ &= \mathrm{tr}\left(\bar{B}^{\frac{1}{2}}\mathrm{diag}(\bar{x})A\cdot \mathrm{diag}(x)\bar{B}^{\frac{1}{2}}\right) \\ &= \mathrm{tr}\left(C^{\mathrm{H}}AC\right) \end{aligned}$$

  注意到合同变换保持 Hermite 半正定性，可知 $C^{\mathrm{H}}AC$ 也为 Hermite 半正定阵.
  因此其对角元均为非负实数，于是有 $x^{\mathrm{H}}(A\odot B)x = \mathrm{tr}\left(C^{\mathrm{H}}AC\right) \geq 0$ 成立.
  根据 $x\in\mathbb{C}^n$ 的任意性可知 $A\odot B$ Hermite 半正定.

Proof:
任意给定正整数 $m\in\mathbb{Z}_+$ 和 $x_1,\dots,x_m\in\mathcal{X}$，定义:

$$\begin{aligned} K_1 &:= [k_1(x_i,x_j)]_{i,j=1}^m \\ K_2 &:= [k_2(x_i,x_j)]_{i,j=1}^m \\ K &:= [k(x_i,x_j)]_{i,j=1}^m \\ &= [k_1(x_i,x_j)k_2(x_i,x_j)]_{i,j=1}^m \\ &= K_1\odot K_2 \end{aligned}$$

由于 K_1, K_2 半正定，故根据 **Lemma 2 (Schur 乘积定理)** 可知 $K = K_1 \odot K_2$ 也半正定.
根据 Mercer 定理可知由 $k(x,z) = k_1(x,z)k_2(x,z)$ 定义的对称函数是正定核函数.

Part (4)

假设 $k_1(\cdot,\cdot)$ 和 $k_2(\cdot,\cdot)$ 都是正定核函数，试根据 Mercer 定理证明:
对于 $\mathcal{X}$ 上的任意实值函数 $f(\cdot)$，由 $k(x,z) = f(x)k_1(x,z)f(z)$ 定义的对称函数都是正定核函数.

Proof:
任意给定正整数 $m \in \mathbb{Z}_+$ 和 $x_1,\dots,x_m \in \mathcal{X}$，定义:

$$\begin{aligned} y &:= [f(x_1),\dots,f(x_m)]^{\mathrm{T}} \\ K_1 &:= [k_1(x_i,x_j)]_{i,j=1}^m \\ K &:= [k(x_i,x_j)]_{i,j=1}^m \\ &= [f(x_i)k_1(x_i,x_j)f(x_j)]_{i,j=1}^m \\ &= \mathrm{diag}(y)K_1\mathrm{diag}(y) \end{aligned}$$

对于任意 $z \in \mathcal{X}$，我们都有:

$$\begin{aligned} z^{\mathrm{T}}Kz &= z^{\mathrm{T}}\mathrm{diag}(y)K_1\mathrm{diag}(y)z \\ &= (\mathrm{diag}(y)z)^{\mathrm{T}}K_1(\mathrm{diag}(y)z) \\ &\geq 0 \end{aligned}$$

这意味着 K 半正定.
根据 Mercer 定理可知由 $k(x,z) = f(x)k_1(x,z)f(z)$ 定义的对称函数是正定核函数.

The End