Statistical Learning HW7

丁泽澍 18307110088

(1) 解：引入拉格朗日乘子 α 和 μ，则原问题的拉格朗日函数为：

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

其中 $\alpha_i,\mu_i\geq 0$

则：dual function. $\theta(\alpha,\mu)=\inf\limits_{w,b,\xi} L(w,b,\xi;\alpha,\mu)$

∴求解极小问题：

$$\begin{cases}\dfrac{\partial L}{\partial w}=w-\sum_{i=1}^{N}\alpha_i y_i x_i=0\\[2ex]\dfrac{\partial L}{\partial b}=-\sum_{i=1}^{N}\alpha_i y_i=0\\[2ex]\dfrac{\partial L}{\partial \xi_i}=2C\xi_i-\alpha_i-\mu_i=0\end{cases},\ \text{for } i=1,2,\cdots,N$$

∴代入有：$\theta(\alpha,\mu)=-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j x_i^T x_j+\sum_{i=1}^{N}\alpha_i-\frac{1}{4C}\sum_{i=1}^{N}(\alpha_i+\mu_i)^2$

[$x_i^T x_j$ 即 $(x_i\cdot x_j)$，内积]

∴dual problem 为：

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

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

$$\alpha_i\geq 0,\ \mu_i\geq 0\quad \text{for } i=1,2,\cdots,N$$

(2) 证明：由正定核函数的充要条件，只需证 $K=[K(x_i,x_j)]_{m\times m}$（即 kernel function $K(x,z)=(x\cdot z)^p$ 对应的 Gram 矩阵）半正定：

设任取一个 m 维向量 $c=(c_1,c_2,\cdots,c_m)^T\in R^m$，则有：

$$c^T K c=\sum_{i=1}^{N}\sum_{j=1}^{N}c_i c_j K_{ij}=\sum_{i=1}^{N}\sum_{j=1}^{N}c_i c_j (x_i\cdot x_j)^p \qquad (\#)$$

易证明：对任意$\boxed{\text{整}}$正数 p，均存在一个映射（记为 φ），使得 $(x_i\cdot x_j)^p=\varphi(x_i)\cdot\varphi(x_j)$

"·"为内积

⇒ 用数学归纳法，假设 x_i 与 x_j 为 n 维向量：

1> $p=1$ 时，令 $\varphi(x)=x$ 即可，成立 $x_i\cdot x_j=\varphi(x_i)\cdot\varphi(x_j)$

2> 假设 ~~p=1时~~ $p=n-1$ 时命题成立，即 ∃ 映射 φ，s.t. $(x_i\cdot x_j)^{n-1}=\varphi(x_i)\cdot\varphi(x_j)$

则 $p=n$ 时：$(x_i\cdot x_j)^n=(x_i\cdot x_j)(x_i\cdot x_j)^{n-1}=(x_i\cdot x_j)[\varphi(x_i)\cdot\varphi(x_j)]$

可令 $G(x)=\begin{bmatrix}x_1\varphi(x)\\x_2\varphi(x)\\\vdots\\x_n\varphi(x)\end{bmatrix}$ 若 $\varphi(x)\in R^m$ 则 $G(x)\in R^{mn}$

∴ 由内积定义, $G(x)\cdot G(y) = (x_1y_1)[\varphi(x)\cdot\varphi(y)] + (x_2y_2)[\varphi(x)\cdot\varphi(y)] + \cdots$

$= (x\cdot y)[\varphi(x)\cdot\varphi(y)]$

∴ 有 $(x_i\cdot x_j)^n = (x_i\cdot x_j)[\varphi(x_i)\cdot\varphi(x_j)] = G(x_i)\cdot G(x_j)$

即 $p=n$ 时 命题亦成立

综上，得 对 $\forall p\in N^+$, ∃一个映射(记为φ), s.t. $(x_i\cdot x_j)^p = \varphi(x_i)\cdot\varphi(x_j)$

(该映射对不同的p不同)

代入上面结果于(#)式, 即有: $\exists\varphi$, s.t.

$$c^TKc = \sum_{i=1}^{N}\sum_{j=1}^{N} c_ic_j(x_i\cdot x_j)^p = \sum_{i=1}^{N}\sum_{j=1}^{N}[\varphi(x_i)\cdot\varphi(x_j)]c_ic_j$$

$$= \left(\sum_{i=1}^{N} c_i\varphi(x_i)\right)\left(\sum_{j=1}^{N} c_j\varphi(x_j)\right) = \left\|\sum_{i=1}^{N} c_i\varphi(x_i)\right\|_2^2 \geq 0$$

∴ K半正定 ⟺ $K(x,z)=(x\cdot z)^p$ 为正定核　　证毕　　#

~~(3) 解: 支持向量为: ⑤~~

~~支持向量定义为 在超平面 $w^Tx+b=1$ 或 -1 上的点, 图中仅有⑤满足该定义 (且④有可能是错分类的点)~~

(3) 解: 支持向量为: ② ③ ④ ⑥

原因: 在软间隔最大化求解中, 根据KKT条件要求, 若对偶变量 $\alpha_i>0$, (参照讲义中的表示)

则有 $y_i(w^Tx_i+b) = 1-\xi_i$ 即该样本为支持向量

具体而言, 分为以下几种:

1> ~~$0<\alpha_i$~~ $0<\alpha_i<C$ 则 $\xi_i=0$ $(\mu_i>0)$ ⟹ 恰在间隔边界, 对应图中⑥

2> $\alpha_i=C$ 则 $\mu_i=0$ ⟹
- $0<\xi_i<1$, 正确分类但在间隔平面与分离平面间, 对应图中②
- $\xi_i=1$, 落在分离平面上, 图中无此点.
- $\xi_i>1$, 错分类, 对应图中③和④

∴ 支持向量为 ②③ ~~③④~~ ④⑥