Problem I:

自博16 陈燕杰 2016310721

① 现有N个点 $\{x_1,\cdots,x_N\}\subseteq\mathbb{R}^p$，每个点都对应有一个label $y_i\in\{-1,1\}$

记超平面方程为 $\beta^T x^*=0$，其中：

$\beta^T=[w_1,w_2,\cdots,w_p \mid \beta_0]$ 为增广权向量

$x^*=\begin{bmatrix}x_1\\x_2\\\vdots\\x_p\\1\end{bmatrix}$ 为增广样本向量

$\{x_1,\cdots,x_N\}$ 是可分的，意味着：

$\forall(x_i,y_i)\in\{(x_1,y_1),\cdots,(x_N,y_N)\}$，有：$\begin{cases}y_i=-1 \Rightarrow \beta^T x_i^*<0\\ y_i=1 \Rightarrow \beta^T x_i^*>0\end{cases}$

这等价于：$\forall i\in\{1\ldots N\}$，$y_i\beta^T x_i^*>0$ 恒成立 (1)

引入余量 d，上面(1)式可写作：

$\exists d>0$，s.t. $y_i\beta^T x^*\geq d$

这等价于：

$$y_i\beta^T\frac{x^*}{\|x^*\|}\geq\frac{d}{\|x^*\|}$$

取 β_{sep}^T 为 $\frac{\|x^*\|}{d}\beta^T$ 可得 $y_i\beta_{sep}^T\cdot\frac{x^*}{\|x^*\|}\geq 1$，即 $y_i\beta_{sep}^T z_i\geq 1$

② 根据更新规则：$\beta_{new}=\beta_{old}+y_i z_i$；

$$\beta_{new}-\beta_{sep}=\beta_{old}-\beta_{sep}+y_i z_i$$

取2-范数：

$$\|\beta_{new}-\beta_{sep}\|^2=\|\beta_{old}-\beta_{sep}+y_i z_i\|^2$$

$$=\|\beta_{old}-\beta_{sep}\|^2+\|y_i z_i\|^2+2y_i(\beta_{old}-\beta_{sep})^T z_i$$

$$=\|\beta_{old}-\beta_{sep}\|^2+y_i^2\|z_i\|^2+2y_i(\beta_{old}-\beta_{sep})^T z_i$$

其中，$y_i\in\{-1,1\}$，$y_i^2=1$，$z_i=\frac{x_i}{\|x_i\|}$ 故其二范数必为1，故 $y_i^2\|z_i\|^2=1$

$$2y_i(\beta_{old}-\beta_{sep})^T z_i=2(y_i\beta_{old}z_i-y_i\beta_{sep}z_i)$$

β_{old} 把 z_i 分错了，$\therefore y_i\beta_{old}z_i<0$

β_{sep} 把 z_i 分对了，$\therefore y_i\beta_{sep}z_i\geq1$ 即 $-y_i\beta_{sep}z_i\leq-1$

$\therefore 2y_i(\beta_{old}-\beta_{sep})^T z_i\leq-2$

综上，$\|\beta_{new}-\beta_{sep}\|^2\leq\|\beta_{old}-\beta_{sep}\|^2-1$

Problem 2 自博16 陈斯杰 2016310721

证明 Hahn-Banach 定理的几何形式在 R^n 空间中的特例.

证明：考察两列向量 $S_1=\{x_1,\cdots,x_n\}$，$S_2=\{y_1,\cdots,y_m\}$

设 S_1 和 S_2 线性可分，分类函数为 $g(x)=w^Tx+w_0$

$$\begin{cases} x\in S_1 & g(x)>0 \\ x\in S_2 & g(x)<0 \end{cases}$$

考察 S_1 的凸包中的一个点 x，有：

$$x=\sum_{i=1}^n \alpha_i x_i$$

$$\begin{aligned} g(x) &= w^Tx+w_0 \\ &= w^T\left(\sum_{i=1}^n \alpha_i x_i\right)+w_0 \\ &= \sum_{i=1}^n(\alpha_i w^T x_i)+\left(w_0\sum_{i=1}^n \alpha_i\right) \qquad \left(\sum_{i=1}^n \alpha_i=1,\ \alpha_i\geq 0\right) \\ &= \sum_{i=1}^n\left[\alpha_i\,(w^Tx_i+w_0)\right]>0 \end{aligned}$$

在此用反证法，假设 S_1 的凸包中的点 x，同时也在 S_2 的凸包中，即 x 也可表示为：

$$x=\sum_{i=1}^m \beta_i y_i$$

$$\begin{aligned} g(x) &= w^Tx+w_0 \\ &= \sum_{i=1}^m \beta_i\,(w^Ty_i+w_0)>0 \qquad (*) \end{aligned}$$

但是 $y_i\in S_2$，$g(y_i)=w^Ty_i+w_0<0$

由 $\sum_{i=1}^m \beta_i=1$，$\beta_i\geq 0$，

$$\sum_{i=1}^m \beta_i(w^Ty_i+w_0)\leq 0, \qquad (**)$$

(*) 和 (**) 矛盾，故反证假设不成立，S_1 的凸包中的点是不可能同时在 S_2 的凸包中的。

Problem 3

自博16 陈斯杰 2016310721

① 原问题为：$\begin{cases} \min\limits_{w} \frac{1}{2}w^Tw \\ s.t.\ w^T x^{(i)} \geq 1 \quad \forall i \in \{1,\dots,m\} \end{cases}$

约束为不等式约束，为 $w^T x^{(i)} - 1 \geq 0$ 引入松弛变量 α_i，$\alpha_i \geq 0$，$\sum \alpha_i = 0$

那么，$\alpha_i(1-w^Tx^{(i)}) = 0 \quad \forall i \in \{1,\dots,m)$

其拉格朗日函数为：

$$L(w,\vec{\alpha}) = \frac{1}{2}w^Tw + \sum_{i=1}^{m} \alpha_i(1-w^Tx^{(i)})$$

其中 $\vec{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_m)$

$$\frac{\partial L}{\partial w} = \frac{1}{2}\cdot 2w + \sum_{i=1}^{m}\alpha_i(1-\frac{\partial}{\partial w}w^Tx^{(i)})$$

$$= w + \sum_{i=1}^{m}\alpha_i(1-x^{(i)})$$

$$= w + (\sum_{i=1}^{m}\alpha_i) - \sum_{i=1}^{m}\alpha_i x^{(i)}$$

$$= 0$$

$$\Rightarrow w = \sum_{i=1}^{m}\alpha_i x^{(i)}$$

w 满足：

$$\max_{\alpha \geq 0}\min_{w}\left(\frac{1}{2}w^Tw + \sum_{i=1}^{m}\alpha_i(1-w^Tx^{(i)})\right)$$

$$= \max_{\alpha \geq 0}\left\{\frac{1}{2}(\sum_{i=1}^{m}\alpha_i x^{(i)})^T(\sum_{i=1}^{m}\alpha_i x^{(i)}) + \sum_{i=1}^{m}\alpha_i[1-(\sum_{i=1}^{m}\alpha_i x^{(i)})^T x^{(i)}]\right\}$$

$$= \max_{\alpha \geq 0}\left\{\sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j x^{(i)T}x^{(j)}\right\}$$

② 可以通过核函数完成

因为：$w = \sum_{i=1}^{m}\alpha_i x^{(i)}$

在对一个点 p 进行判断时

$$w^Tp = (\sum_{i=1}^{m}\alpha_i x^{(i)})\cdot p$$

$$= \sum_{i=1}^{m}([\alpha_i x^{(i)}]p)$$

可以写作内积形式

Problem 4:

自博16 陈斯杰 2016310721

$$\begin{cases} \min \ \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \\ s.t. \quad y_i w^T x_i \geq 1-\xi_i \quad (i=1\ldots n) \\ \qquad \xi_i \geq 0 \quad (i=1\ldots n) \end{cases}$$

此问题可以转化为拉格朗日函数的鞍点问题

$$\min_{w,b,\xi_i} \max_{\alpha} L(\vec{w}, b, \vec{\alpha}) = \frac{1}{2}w^T w + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\{y_i[(w^T x_i)+b]-1+\xi_i\} - \sum_{i=1}^{n}\gamma_i\xi_i$$

其中 $\alpha_i \geq 0$, $\gamma_i \geq 0$

对 $L(\vec{w}, b, \vec{\alpha})$ 求导并令其为0，代入，

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{n}\alpha_i\{y_i x_i\} = 0 \quad \Rightarrow w = \sum_{i=1}^{n}\alpha_i y_i x_i$$

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^{n}\alpha_i y_i = 0$$

$$\frac{\partial L}{\partial \xi_i} = C - \gamma_i$$

得对偶问题

$$\begin{cases} \max \ Q(\alpha) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j x_i^T x_j \\ s.t. \ \sum_{i=1}^{n} y_i\alpha_i = 0 \\ 0 \leq \alpha_i \leq C \\ i \in \{1..n\} \end{cases}$$

可以写成无约束形式

$$\min_{w,b} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\max\{1-y_i(w^T x_i+b), 0\}$$

记 $L(w, b, x_i, y_i) = \max\{1-y_i(w^T x_i+b), 0\}$

L 表示样本距离分类超平面的距离小于间隔的阈值时，带来的惩罚。

Problem 5 自博16 陈斯杰 2016310721

设 K 是一个 mercer 核

ϕ 将 x 映射为 $\phi(x)$ 后，$\phi(x)$ 的可分性变好了。

那么理论上，可以继续如下更新 θ

$$\theta^{(i+1)} = \theta^{(i)} + \alpha\left(y^{(i)} - h_{\theta^{(i)}}[\phi(x^{(i+1)})]\right)\cdot\phi(x^{(i+1)})$$

题中提及了，由于更新的方式如此，这就决定了

$\theta^{(i+1)}$ 的值永远都只是 $\phi(x^{(i)})$ 值的线性组合，

即：

$$\theta^{(i)} = \sum_{j=1}^{i}\beta_j\phi(x^{(j)})$$

那么 $\theta^{(i)}$ 的值就可以用 i 个 β_j 的值来表出（$j=1,\ldots,i$）

② 记 $h_{\theta^{(i)}}(x^{(i+1)}) = g(\theta^{(i)T}\cdot\phi(x^{(i+1)}))$

$$g(\theta^{(i)T}\cdot\phi(x^{(i+1)})) = g\left[\left(\sum_{j=1}^{i}\beta_j\phi(x^{(j)})\right)\cdot\phi(x^{(i+1)})\right]$$

$$= g\left[\sum_{j=1}^{i}\beta_j\phi(x^{(j)})\cdot\phi(x^{(i+1)})\right]$$

$$= g\left[\sum_{j=1}^{i}\beta_j K(x^{(j)}, x^{(i+1)})\right]$$

（$x^{(j)}$ 和 $x^{(i+1)}$ 是低维的，其核函数也好算）

③ $$\theta^{(i)} = \theta^{(i-1)} + \alpha\left[y^{(i)} - h_{\theta^{(i-1)}}(\phi(x^{(i)})\right]\phi(x^{(i)})$$

$$= \theta^{(i-1)} + \alpha\left[y^{(i)} - g(\theta^{(i-1)T}\phi(x^{(i)}))\right]\phi(x^{(i)})$$

$$\theta^{(i)} = \sum_{j=1}^{i}\beta_j\phi(x^{(j)}) \Rightarrow \theta^{(i)} - \theta^{(i-1)} = \beta_i\phi(x^{(i)})$$

$$\therefore \beta_i = \frac{\theta^{(i)} - \theta^{(i-1)}}{\phi(x^{(i)})} = \alpha\left[y^{(i)} - g(\theta^{(i-1)T}\phi(x^{(i)}))\right]$$

$$= \alpha\left\{y^{(i)} - g\left[\sum_{j=1}^{i-1}K(x^{(j)}, x^{(i)})\right]\right\}$$ （代入）

$\therefore$ 综上 $$\theta^{(i+1)} = \alpha\left\{y^{(i+1)} - g\left[\sum_{j=1}^{i}K(x^{(j)}, x^{(i+1)})\right]\right\}$$