线性判别函数习题

姓名： 甘云冲

学号： 2101213081

1.

$$
\begin{aligned}
J(\mathbf{w}) &= \frac{(P_1\mu_1 - P_2\mu_2)^2}{P_1\sigma_1^2 + P_2\sigma_2^2} \\
&= \frac{(P_1\mathbf{w}^T\mathbf{m}_1 - P_2\mathbf{w}^T\mathbf{m}_2)^2}{P_1\mathbf{w}^T\mathbf{S}_1\mathbf{w} + P_1\mathbf{w}^T\mathbf{S}_2\mathbf{w}} \\
&= \frac{\mathbf{w}^T(P_1\mathbf{m}_1 - P_2\mathbf{m}_2)(P_1\mathbf{m}_1 - P_2\mathbf{m}_2)^T\mathbf{w}}{\mathbf{w}^T(P_1\mathbf{S}_1 + P_2\mathbf{S}_2)\mathbf{w}}
\end{aligned}
$$

定义

$$
\mathbf{S}_b^\star = (P_1\mathbf{m}_1 - P_2\mathbf{m}_2)(P_1\mathbf{m}_1 - P_2\mathbf{m}_2)^T, \mathbf{S}_w^\star = P_1\mathbf{S}_1 + P_2\mathbf{S}_2
$$

于是有

$$
J(\mathbf{w}) = \frac{\mathbf{w}^T\mathbf{S}_b^\star\mathbf{w}}{\mathbf{w}^T\mathbf{S}_w^\star\mathbf{w}}
$$

利用Fisher线性判别结论有

$$
\hat{\mathbf{w}} = \mathbf{S}_w^{\star-1}(P_1\mathbf{m}_1 - P_2\mathbf{m}_2) = (P_1\mathbf{S}_1 + P_2\mathbf{S}_2)^{-1}(P_1\mathbf{m}_1 - P_2\mathbf{m}_2)
$$

2.

(1)

采用反证法，假设凸包交集不为空，至少存在一点$\mathbf{z}$，有$\mathbf{z} \in S(A) \bigcap S(B)$，于是有：

$$\mathbf{z}=\sum_{i=1}^{n}a_i\mathbf{x}_i=\sum_{i=1}^{m}b_i\mathbf{y}_i$$

由于线性可分，假设存在超平面$\Pi:\mathbf{w}^T\mathbf{x}+b=0$将A和B分开，且：

$$\mathbf{w}^T\mathbf{x}_i+b<0<\mathbf{w}^T\mathbf{y}_j+b$$

不妨令：

$$\epsilon_1=\max\left\{\mathbf{w}^T\mathbf{x}_i+b\middle|\mathbf{x}_i\in A\right\}<0<\min\left\{\mathbf{w}^T\mathbf{y}_i+b\middle|\mathbf{y}_i\in B\right\}=\epsilon_2$$

将$\mathbf{z}$带入：

$$\begin{gathered}\mathbf{w}^T\mathbf{z}+b=\sum_{i=0}^{n}a_i\mathbf{w}^T\mathbf{x}_i+b<(\epsilon_1-b)\sum_{i=0}^{n}a_i+b=\epsilon_1\\ \mathbf{w}^T\mathbf{z}+b=\sum_{i=0}^{n}b_i\mathbf{w}^T\mathbf{y}_i+b>(\epsilon_2-b)\sum_{i=0}^{n}b_i+b=\epsilon_2\\ \Rightarrow\epsilon_2<\epsilon_1\end{gathered}$$

矛盾，所以它们的交集为空。

(2)

由于A和B线性可分，所以SVM的解存在，不妨设在A、B侧的支持平面分别为：

$$\Pi_A:\mathbf{w}^T\mathbf{x}=p,\qquad\Pi_B=\mathbf{w}^T\mathbf{x}=q$$

以及

$$\mathbf{w}^T\mathbf{x}\le p,\forall\mathbf{x}\in A,\qquad\mathbf{w}^T\mathbf{x}\ge q,\forall\mathbf{x}\in B$$

分离超平面即为：

$$\Pi:\mathbf{w}^T\mathbf{x}=\frac{p+q}{2}$$

最大间隔即等价于：

$$\max_{\mathbf{w}}\frac{q-p}{||\mathbf{w}||}$$

同时有等价于：

$$
\min_{\mathbf{w}} \frac{1}{2}||\mathbf{w}||^2 - (q-p)
$$
$$
s.t. \begin{cases} \mathbf{w}^T\mathbf{x} \le p, & \mathbf{x} \in A, \\ \mathbf{w}^T\mathbf{x} \ge q, & \mathbf{x} \in B \end{cases}
$$

于是拉格朗日函数为：

$$
L = \frac{1}{2}||\mathbf{w}||^2 - (q-p) - \sum_{i=1}^{n} \alpha_i(p - \mathbf{w}^T\mathbf{x}_i) - \sum_{i=1}^{m} \beta_i(\mathbf{w}^T\mathbf{y}_i - q)
$$

于是：

$$
\begin{aligned}
\frac{\partial L}{\partial \mathbf{w}} &= \mathbf{w} + \sum_{i=1}^{n} \alpha_i\mathbf{x}_i - \sum_{i=1}^{m} \beta_i\mathbf{y}_i \\
\frac{\partial L}{\partial p} &= 1 - \sum_{i=1}^{n} \alpha_i \\
\frac{\partial L}{\partial q} &= \sum_{i=1}^{m} \beta_i - 1
\end{aligned}
$$

令上式都为0，则对偶优化问题变为：

$$
\min_{\mathbf{w}} \frac{1}{2}\Big|\Big|\sum_{i=1}^{m} \beta_i\mathbf{y}_i - \sum_{i=1}^{n} \alpha_i\mathbf{x}_i\Big|\Big|^2
$$
$$
s.t. \begin{cases} \sum_{i=1}^{n} \alpha_i = 1, & \alpha_i \ge 0 \\ \sum_{i=1}^{m} \beta_i = 1, & \beta_i \ge 0 \end{cases}
$$

这即为凸包$S(A)$和$S(B)$的距离的最近点。

3.

(1)

拉格朗日函数为：

$$
L = \frac{1}{2}||\mathbf{w}||^2 - v\rho + \frac{1}{n}\sum_{i=1}^{n} \xi_i - \alpha\rho - \sum_{i=1}^{n} \alpha_i[y_i(\mathbf{w}^T\mathbf{x}_i + b) - \rho + \xi_i] - \sum_{i=1}^{n} \beta_i\xi_i
$$

于是：

$$
\begin{aligned}
\frac{\partial L}{\partial \mathbf{w}} &= \mathbf{w} - \sum_{i=1}^{n} \alpha_i y_i \mathbf{x_i} \\
\frac{\partial L}{\partial b} &= -\sum_{i=1}^{n} \alpha_i y_i \\
\frac{\partial L}{\partial \rho} &= -v - \alpha + \sum_{i=1}^{n} \alpha_i \\
\frac{\partial L}{\partial \xi_i} &= \frac{1}{n} - \alpha_i - \beta_i
\end{aligned}
$$

令上式都为0，即为：

$$
\begin{aligned}
L &= \frac{1}{2}||\mathbf{w}||^2 - v\rho + \frac{1}{n}\sum_{i=1}^{n}\xi_i - \alpha\rho - \sum_{i=1}^{n}\alpha_i[y_i(\mathbf{w}^T\mathbf{x}_i + b) - \rho + \xi_i] - \sum_{i=1}^{n}\beta_i\xi_i \\
&= \frac{1}{2}||\mathbf{w}||^2 - \sum_{i=1}^{n}\alpha_i y_i(\mathbf{w}^T\mathbf{x}_i + b) \\
&= \frac{1}{2}||\mathbf{w}||^2 - \mathbf{w}^T\sum_{i=1}^{n}\alpha_i y_i \mathbf{x}_i - b\sum_{i=1}^{n}\alpha_i y_i \\
&= -\frac{1}{2}||\mathbf{w}||^2 \\
&= -\frac{1}{2}\sum_{i,j=1}^{n} y_i y_j \alpha_i \alpha_j \mathbf{x}_i^T \mathbf{x}_j
\end{aligned}
$$

对偶形式为：

$$
\begin{aligned}
&\min_{\alpha}\left[\frac{1}{2}\sum_{i,j=1}^{n} y_i y_j \alpha_i \alpha_j \mathbf{x}_i^T \mathbf{x}_j\right] \\
s.t. \quad & 0 \le \alpha_i \le \frac{1}{n} \qquad i = 1, \ldots, n \\
& \sum_{i=0}^{n} \alpha_i y_i = 0
\end{aligned}
$$

(2)

对偶形式：

$$
\begin{aligned}
&\min_{\alpha}\left[\frac{1}{2}\sum_{i,j=1}^{n} y_i y_j \alpha_i \alpha_j \mathbf{x}_i^T \mathbf{x}_j - \sum_{i=0}^{n}\alpha_i\right] \\
&s.t. \quad 0 \le \alpha_i \le \frac{1}{n\hat{\rho}} \qquad i = 1, \ldots, n \\
&\sum_{i=0}^{n}\alpha_i y_i = 0
\end{aligned}
$$

由于$\hat{\rho} > 0$，由KKT条件$\alpha\rho = 0 \Rightarrow \alpha = 0$，所以有$v = \sum_{i=1}^{n}\alpha_i$。

所以上式的对偶形式可改为：

$$
\begin{aligned}
&\min_{\alpha}\left[\frac{1}{2}\sum_{i,j=1}^{n} y_i y_j \alpha_i \alpha_j \mathbf{x}_i^T \mathbf{x}_j - v\right] \\
&s.t. \quad 0 \le \hat{\alpha}_i \le \frac{1}{n} \qquad i = 1, \ldots, n \\
&\sum_{i=0}^{n}\hat{\alpha}_i y_i = 0
\end{aligned}
$$

其中$\hat{\alpha}_i = \alpha_i\rho$，$v$为常数。

$$
\begin{aligned}
&\min_{\alpha}\left[\frac{1}{2}\sum_{i,j=1}^{n} y_i y_j \hat{\alpha}_i \hat{\alpha}_j \mathbf{x}_i^T \mathbf{x}_j\right] \\
&s.t. \quad 0 \le \hat{\alpha}_i \le \frac{1}{n} \qquad i = 1, \ldots, n \\
&\sum_{i=0}^{n}\hat{\alpha}_i y_i = 0
\end{aligned}
$$

和(1)中的对偶形式相同，所以该线性支持向量机与v-SVM的分类器等价。

(3)

由KKT条件可以知道：

$$
\alpha\rho = 0
$$

同时有：

$$
\begin{aligned}
-v - \alpha + \sum_{i=1}^{n}\alpha_i = 0 \\
1 - \sum_{i=1}^{n}\alpha_i - \sum_{i=1}^{n}\beta_i = 0
\end{aligned}
$$

由于$\beta_i \geq 0$，有$\sum_{i=1}^{n} \alpha_i < 1$。

所以$v > 1 > \sum_{i=1}^{n} \alpha_i$：

$$\alpha = \sum_{i=1}^{n} \alpha_i - v \neq 0$$

一定有$\rho = 0$。