中国科学院大学
University of Chinese Academy of Sciences

# 模式识别与机器学习

## 081203M04004H

# Chap 7 课程作业解答

2022 年 11 月 10 号

*Professor:* 黄庆明


**学生: 周胤昌**
**学号: 202228018670052**
**学院: 网络空间安全学院**
**所属专业: 网络空间安全**
**方向: 安全协议理论与技术**

# Problem 1

给定如下训练数据集

$$X = \begin{pmatrix} 3 & 4 & 1 \\ 3 & 3 & 1 \end{pmatrix}, y = \begin{pmatrix} 1 & 1 & -1 \end{pmatrix}^{\mathrm{T}}$$

通过求解 SVM 的**原始问题**来求解最大间隔的分离超平面.

**Solution:** 通过求解如下最优化问题来得到最优分类器的参数 $(\boldsymbol{w}^*, b^*)$:

$$\begin{cases} \min\limits_{\boldsymbol{w},b} \dfrac{1}{2}\|\boldsymbol{w}\|_2^2 \\ \text{s.t. } y^{(i)}\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}^{(i)} + b\right) \geq 1, i = 1, \cdots, N \end{cases} \Rightarrow \begin{cases} \min\limits_{(w_1,w_2),b} \dfrac{1}{2}\left(w_1^2 + w_2^2\right) \\ 3w_1 + 3w_2 + b \geq 1 \\ 4w_1 + 3w_2 + b \geq 1 \\ -\left(w_1 + w_2 + b\right) \geq 1 \end{cases}$$

要求求解这个严格的凸二次规划问题, 我们可以直接利用 Matlab 中的二次规划问题求解包直接求解. 而二次规划的标准形式为

$$\begin{cases} \min\limits_{\boldsymbol{x}} \dfrac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{H}\boldsymbol{x} + \boldsymbol{f}^{\mathrm{T}}\boldsymbol{x} \\ \boldsymbol{A} \cdot \boldsymbol{x} \leq \boldsymbol{b} \end{cases}$$

对照过来, 可以知道

$$\boldsymbol{H} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \boldsymbol{f} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{A} = \begin{pmatrix} -3 & -3 & -1 \\ -4 & -3 & -1 \\ 1 & 1 & 1 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}$$

于是我们可以在 Matlab 中写出如下代码:

```matlab
H = [1, 0, 0; 0, 1, 0; 0, 0, 0];
f = [0; 0; 0];
A = [-3, -3, -1; -4, -3, -1; 1, 1, 1];
b = [-1; -1; -1];
[x, fval, exitflag, output, lambda] = quadprog(H, f, A, b)
```

执行后便可得到结果 (exitflag=1, 即问题存在最优解):

$$\boldsymbol{x}^* = \begin{pmatrix} w_1^* \\ w_2^* \\ b^* \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.5 \\ -2 \end{pmatrix}, \min_{(w_1,w_2,b)} \frac{1}{2}\left(w_1^2 + w_2^2\right) = 0.25$$

于是最大间隔的分离超平面方程和判别函数分别为

$$\frac{1}{2}x_1 + \frac{1}{2}x_2 - 2 = 0, f_{\boldsymbol{w},b}(\boldsymbol{x}) = \operatorname{sgn}\left(\frac{1}{2}x_1 + \frac{1}{2}x_2 - 2\right)$$

本题的训练样本较少, 所以方法 2 是说: 可以直接观察出支持向量 (而不用求解上述二次规划问题), 从而立即得到具有最大间隔的分离超平面[1]. 具体如下图1所示:

[1]这是因为最终模型仅与**支持向量**有关.

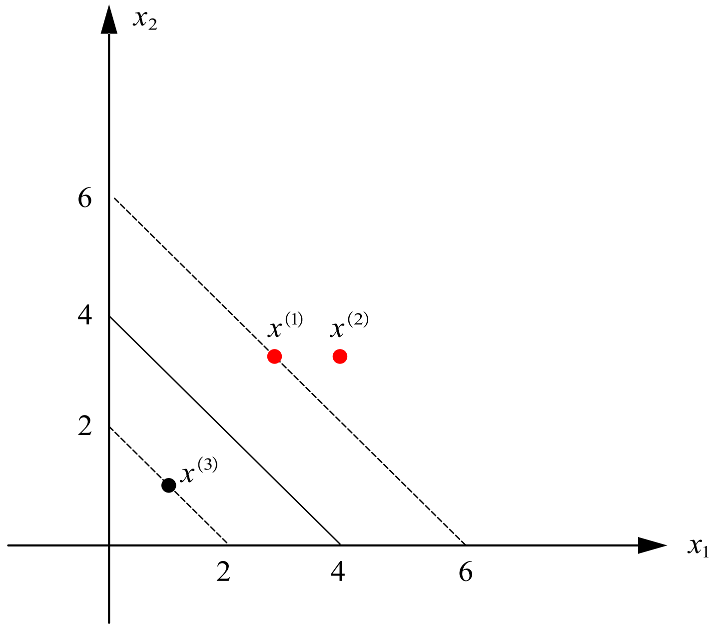


图 1: 最大间隔分离超平面示意图

明显可以看出支持向量为 $x^{(3)}, x^{(1)}$, 所以**这两个点的垂直平分线就是所要求的最大间隔分离超平面**! 即其方程为 $0.5x_1 + 0.5x_2 - 2 = 0$. 而对于上述二次优化问题, 我们也可以利用拉格朗日数乘法求解如下: 构造辅助函数 (只代入支持向量, 因为最终模型只与支持向量有关)

$$L(\boldsymbol{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\left(w_1^2 + w_2^2\right) + \alpha_1(1 - 3w_1 - 3w_2 - b) + \alpha_2(1 + w_1 + w_2 + b)$$

对 $w_1, w_2, b$ 分别求偏导:

$$\begin{cases} w_1 - 3\alpha_1 + \alpha_2 = 0 \\ w_2 - 3\alpha_1 + \alpha_2 = 0 \\ -\alpha_1 + \alpha_2 = 0 \end{cases} \Rightarrow w_1 = w_2 = 2\alpha_1$$

代入辅助函数可得:

$$L(\alpha_1) = -4\alpha_1^2 + 2\alpha_1, L'(\alpha_1) = -8\alpha_1 + 2$$

令 $L'(\alpha_1) = 0$ 得 $\alpha_1 = \frac{1}{4}$, 因此 $w_1 = w_2 = \frac{1}{2}$, 从而 $b \leq -2$, 于是最大间隔的分离超平面方程为

$$\frac{1}{2}x_1 + \frac{1}{2}x_2 - 2 = 0$$

# Problem 2

给定如下训练数据集

$$\boldsymbol{X} = \begin{pmatrix} 3 & 4 & 1 \\ 3 & 3 & 1 \end{pmatrix}, \boldsymbol{y} = \begin{pmatrix} 1 & 1 & -1 \end{pmatrix}^{\mathrm{T}}$$

通过求解 SVM 的**对偶问题**来求解最大间隔的分离超平面.

**Solution:** SVM 的对偶问题为

$$\begin{cases} \max\limits_{\boldsymbol{\alpha}} \left\{ \sum\limits_{i=1}^{N} \alpha_i - \dfrac{1}{2} \sum\limits_{i,j=1}^{N} \alpha_i \alpha_j y^{(i)} y^{(j)} \left\langle \boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)} \right\rangle \right\} \\ \text{s.t. } \alpha_i \geq 0, i = 1, \cdots, N \\ \sum\limits_{i=1}^{N} \alpha_i y^{(i)} = 0 \end{cases}$$

代入上述训练样本具体化对偶问题可得

$$\begin{cases} \max\limits_{\boldsymbol{\alpha}} \left\{ \sum\limits_{i=1}^{3} \alpha_i - \dfrac{1}{2} \left( 18\alpha_1^2 + 25\alpha_2^2 + 2\alpha_3^2 + 21\alpha_1\alpha_2 - 6\alpha_1\alpha_3 - 7\alpha_2\alpha_3 \right) \right\} \\ \alpha_i \geq 0, i = 1, 2, 3 \\ \alpha_1 + \alpha_2 - \alpha_3 = 0 \end{cases}$$

而求解这个二次规划问题, 我们可以利用 Matlab 中的求解包来求解. 而二次规划的标准形式为

$$\begin{cases} \min\limits_{\boldsymbol{x}} \dfrac{1}{2} \boldsymbol{x}^{\mathrm{T}} \boldsymbol{H} \boldsymbol{x} + \boldsymbol{f}^{\mathrm{T}} \boldsymbol{x} \\ \boldsymbol{A} \cdot \boldsymbol{x} \leq \boldsymbol{b} \\ \boldsymbol{A}_{eq} \cdot \boldsymbol{x} = \boldsymbol{b}_{eq} \end{cases}$$

对照过来可以知道:

$$\boldsymbol{H} = \begin{pmatrix} 18 & 10.5 & -3 \\ 10.5 & 25 & -3.5 \\ -3 & -3.5 & 2 \end{pmatrix}, \boldsymbol{f} = \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}$$

$$\boldsymbol{A} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \boldsymbol{b} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{A}_{eq} = \begin{pmatrix} 1 & 1 & -1 \end{pmatrix}, \boldsymbol{b}_{eq} = (0)$$

于是我们可以在 Matlab 中写出如下代码:

```
H = [18, 10.5, -3; 10.5, 25, -3.5; -3, -3.5, 2];
f = [-1; -1; -1];
A = [-1, 0, 0; 0, -1, 0; 0, 0, -1];
b = [0; 0; 0];
A_eq = [1, 1, -1];
b_eq = [0];
[x, fval, exitflag, output, lambda] = quadprog(H, f, A, b, A_eq, b_eq)
```

根据代码输出就可以得到对偶问题的最优解 $\boldsymbol{\alpha}^*$ 和相应的最优值分别为

$$\boldsymbol{\alpha}^* = \begin{pmatrix} 0.25 \\ 0 \\ 0.25 \end{pmatrix}, \max_{\boldsymbol{\alpha}} \left\{ \sum_{i=1}^{3} \alpha_i - \frac{1}{2}\left(18\alpha_1^2 + 25\alpha_2^2 + 2\alpha_3^2 + 21\alpha_1\alpha_2 - 6\alpha_1\alpha_3 - 7\alpha_2\alpha_3\right) \right\} = 0.0625$$

于是原问题的最优解 $(\boldsymbol{w}^*, b^*)$ 为

$$\boldsymbol{w}^* = \sum_{i=1}^{N} \alpha_i^* y^{(i)} \boldsymbol{x}^{(i)} = 0.25 \cdot \begin{pmatrix} 3 \\ 3 \end{pmatrix} + 0 \cdot \begin{pmatrix} 4 \\ 3 \end{pmatrix} - 0.25 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$$

$$b^* = y^{(j)} - \sum_{i=1}^{N} \alpha_i^* y^{(i)} \left\langle \boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)} \right\rangle, \alpha_j^* > 0 \Rightarrow b^* = y^{(1)} - \sum_{i=1}^{3} \alpha_i^* y^{(i)} \left\langle \boldsymbol{x}^{(i)}, \boldsymbol{x}^{(1)} \right\rangle = -2$$

于是分离超平面方程和判别函数的表达式分别为

$$\boldsymbol{w}^{*\mathrm{T}} \boldsymbol{x} + b^* = 0.5\,(x_1 + x_3) - 2 = 0,\ f_{\boldsymbol{w},b}(\boldsymbol{x}) = \operatorname{sgn}\left(\boldsymbol{w}^{*\mathrm{T}} \boldsymbol{x} + b^*\right)$$

对于上面的对偶问题, 我们也可以解析求解: 根据约束条件可知 $\alpha_3 = \alpha_1 + \alpha_2$, 代入目标函数可得

$$\theta_D(\alpha_1, \alpha_2) = -4\alpha_1^2 - 10\alpha_1\alpha_2 - 3\alpha_2^2 + 2\alpha_1 + 2\alpha_2$$

于是

$$\frac{\partial \theta_D(\alpha_1, \alpha_2)}{\partial \alpha_1} = -8\alpha_1 - 10\alpha_2 + 2 = 0 \Rightarrow \alpha_1 = \frac{1}{4} - \frac{5\alpha_2}{4}$$

代入 $\theta_D(\alpha_1, \alpha_2)$ 可得

$$\theta_D(\alpha_2) = -\frac{1}{4}\alpha_2^2 - \frac{1}{2}\alpha_2 + \frac{1}{4} \xrightarrow{\alpha_2 \geq 0} \max_{\alpha_2 \geq 0} \theta_D = \theta_D(0) = \frac{1}{4} \Rightarrow a_2 = 0, \alpha_1 = \alpha_3 = \frac{1}{4}$$

对于原问题最优解的推导与上面的过程一致, 就不赘述了. 我们也可以利用下述的算法1(即 SMO 算法) 来求出 SVM 的对偶问题的解为 $\boldsymbol{\alpha}^* = (0.25, 0, 0.25)^{\mathrm{T}}$:

**Algorithm 1 SMO 算法**

**Input:** 训练数据集 $S = \left\{\left(\boldsymbol{x}^{(i)}, y^{(i)}\right), i = 1, \cdots, N\right\}$, 误差 $\epsilon$

**Output:** $\hat{\alpha} = (\hat{\alpha}_1, \cdots, \hat{\alpha}_N)$

1: 初始化: $\alpha^{(0)} \leftarrow 0, k \leftarrow 0, C \leftarrow +\infty$, 并计算偏移量 $b^{(0)}$;

2: 初始化误差项: $E_i \leftarrow g\left(\boldsymbol{x}^{(i)}\right) - y^{(i)}$;

3: **while** 不满足 KKT 条件且 $\exists i \in \{1, 2, \cdots, N\}$, s.t. $E_i \geq \epsilon$ **do**

4: $k \leftarrow k + 1$ 并选择待优化的变量: $\alpha_1^{(k)}, \alpha_2^{(k)}$, 然后求解优化问题的解 $\alpha_1^{(k+1)}, \alpha_2^{(k+1)}$:

$$\alpha_2^{\text{new,unclipped}} \leftarrow \alpha_2^{(k)} + \frac{y^2(E_1 - E_2)}{\eta},,\ \alpha_2^{(k+1)} \leftarrow \begin{cases} H, & \text{若}\alpha_2^{\text{new,unclipped}} > H \\ \alpha_2^{\text{new,unclipped}}, & \text{若}L \leq \alpha_2^{\text{new,unclipped}} \leq H \\ L, & \text{若}\alpha_2^{\text{new,unclipped}} < L \end{cases}$$

$$\eta \leftarrow K_{11} + K_{22} - 2K_{12},\ \alpha_1^{(k+1)} \leftarrow \alpha_1^{(k)} + y^{(1)} y^{(2)} \left(\alpha_2^{(k)} - \alpha_2^{(k+1)}\right)$$

5: 更新 $\alpha \leftarrow \alpha^{(k+1)}$, 更新 $E_i \leftarrow g\left(\boldsymbol{x}^{(i)}\right) - y^{(i)}$, 计算 $b^{(k+1)}$;

6: **end while**

7: **end {SMO}**

# Problem 3

推导软间隔 SVM 的对偶形式.

**Solution:** 软间隔分类器 SVM 的原问题如下:

$$\begin{cases} \min\limits_{\boldsymbol{w},b} \dfrac{1}{2}\|\boldsymbol{w}\|_2^2 + C\sum\limits_{i=1}^{N}\xi_i \\ y^{(i)}\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}^{(i)} + b\right) \geq 1 - \xi_i, i = 1, \cdots, N \\ \xi_i \geq 0, i = 1, \cdots, N \end{cases}$$

于是我们可以构造如下的广义拉格朗日函数:

$$L(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\eta}) = \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\alpha_i\left[y^{(i)}\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}^{(i)} + b\right) - 1 + \xi_i\right] - \sum_{i=1}^{N}\eta_i\xi_i$$

先固定拉格朗日乘子 $\boldsymbol{\alpha}, \boldsymbol{\eta}$, 关于 $\boldsymbol{w}, b, \xi_i$ 对 $L(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\eta})$ 做优化 (最小化) 得到 $\theta_D(\boldsymbol{\alpha}, \boldsymbol{\eta})$:

$$\begin{cases} \nabla_{\boldsymbol{w}} L(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\eta}) = \boldsymbol{w} - \sum\limits_{i=1}^{N}\alpha_i y^{(i)}\boldsymbol{x}^{(i)} = 0 \\ \dfrac{\partial}{\partial b} L(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\eta}) = -\sum\limits_{i=1}^{N}\alpha_i y^{(i)} = 0 \\ \dfrac{\partial}{\partial \xi_i} L(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\eta}) = C - \alpha_i - \eta_i = 0 \end{cases}$$

将上述 3 个条件代入广义拉格朗日函数可得到 $\theta_D(\boldsymbol{\alpha}, \boldsymbol{\eta})$:

$$\begin{aligned} \theta_D(\boldsymbol{\alpha}, \boldsymbol{\eta}) &= \frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\alpha_i\left[y^{(i)}\left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x}^{(i)} + b\right) - 1 + \xi_i\right] - \sum_{i=1}^{N}\eta_i\xi_i \\ &= \frac{1}{2}\sum_{i=1}^{N}\alpha_i y^{(i)}\left(\boldsymbol{x}^{(i)}\right)^{\mathrm{T}}\sum_{j=1}^{N}\alpha_j y^{(j)}\boldsymbol{x}^{(j)} - \sum_{i=1}^{N}\alpha_i\left[y^{(i)}\left(\sum_{j=1}^{N}\alpha_j y^{(j)}\left(\boldsymbol{x}^{(j)}\right)^{\mathrm{T}}\boldsymbol{x}^{(i)} + b\right) - 1 + \xi_i\right] \\ &+ C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\eta_i\xi_i \\ &= \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j y^{(i)}y^{(j)}\left\langle\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}\right\rangle - \sum_{i,j=1}^{N}\alpha_i\alpha_j y^{(i)}y^{(j)}\left\langle\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}\right\rangle - b\cdot\sum_{i=1}^{N}\alpha_i y^{(i)} \\ &+ \sum_{i=1}^{N}\alpha_i + \sum_{i=1}^{N}\left(C - \alpha_i - \eta_i\right)\xi_i \\ &= -\frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j y^{(i)}y^{(j)}\left\langle\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}\right\rangle + \sum_{i=1}^{N}\alpha_i \end{aligned}$$

即得到

$$\theta_D(\boldsymbol{\alpha}, \boldsymbol{\eta}) = \min_{\boldsymbol{w},b,\xi} L(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\eta}) = \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j y^{(i)}y^{(j)}\left\langle\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}\right\rangle = \theta_D(\boldsymbol{\alpha})$$

于是最大化 $\theta_D(\boldsymbol{\alpha})$, 即可得出下述的对偶问题:

$$\begin{cases} \max\limits_{\boldsymbol{\alpha}} \left\{ \sum\limits_{i=1}^{N} \alpha_i - \frac{1}{2} \sum\limits_{i,j=1}^{N} \alpha_i \alpha_j y^{(i)} y^{(j)} \left\langle \boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)} \right\rangle \right\} \\ \sum\limits_{i=1}^{N} \alpha_i y^{(i)} = 0 \\ 0 \le \alpha_i = \underbrace{C - \eta_i \le C}_{\because \eta_i \ge 0 (\text{乘子非负})}, i = 1, \cdots, N \end{cases}$$

假设 $\boldsymbol{\alpha}^* = \left(\alpha_1^*, \cdots, \alpha_N^*\right)$ 是上述对偶问题的最优解, 那么原问题的解可以如下求解: 显然 $\boldsymbol{w}^*$ 可以直接写出:

$$\boldsymbol{w}^* = \sum_{i=1}^{N} \alpha_i^* y^{(i)} \boldsymbol{x}^{(i)}$$

而对于 $b^*$, 需要找到一个相关的等式, 而根据库恩塔克条件 (KKT) 可得

$$\alpha_i \left[ y^{(i)} \left( \boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}^{(i)} + b \right) - 1 + \xi_i \right] = 0, \eta_i \xi_i = 0$$

因此若想要得到 $b$ 的等式, 则根据互补松弛性可知, 只需要 $\alpha_i > 0 (\neq 0)$, 但是 $\xi_i$ 又是不能知道的, 所以需要想办法把它去掉 (即需要使得 $\xi_i = 0$), 于是又根据互补松弛性可知, 再需要令 $\eta_i > 0 (\neq 0)$ 即可使得 $\xi_i = 0$, 从而可得出 $b$ 的等式

$$y^{(i)} \left( \boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}^{(i)} + b \right) = 1 \Rightarrow b = y^{(i)} - \boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}^{(i)}$$

因此若想求得 $b^*$, 则需要挑选样本 $\left(\boldsymbol{x}^{(j)}, y^{(j)}\right)$, 使得样本满足 $\alpha_j^* > 0, \eta_j > 0$ (即 $0 < \alpha_j^* < C$), 于是可求得

$$b^* = y^{(j)} - \boldsymbol{w}^{\mathrm{T}} \boldsymbol{x}^{(j)} = y^{(j)} - \sum_{i=1}^{N} \alpha_i^* y^{(i)} \left( \boldsymbol{x}^{(i)} \right)^{\mathrm{T}} \boldsymbol{x}^{(j)} = y^{(j)} - \sum_{i=1}^{N} \alpha_i^* y^{(i)} \left\langle \boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)} \right\rangle$$

于是可求得分类超平面方程和判别函数分别为

$$\left(\boldsymbol{w}^*\right)^{\mathrm{T}} \boldsymbol{x} + b^* = 0, f_{\boldsymbol{w},b}(\boldsymbol{x}) = \operatorname{sgn}\left\{ \left(\boldsymbol{w}^*\right)^{\mathrm{T}} \boldsymbol{x} + b^* \right\}$$

# Problem 4

高斯核的形式如下

$$\mathcal{K}(\boldsymbol{x},\boldsymbol{z})=\exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{z}\|^2}{2\sigma^2}\right)$$

请证明高斯核可以表示为无限维特征向量的内积.

**Solution:** 可以将高斯核作如下展开:

$$\begin{aligned}\mathcal{K}(\boldsymbol{x},\boldsymbol{z})&=\exp\left\{-\frac{(\boldsymbol{x}-\boldsymbol{z})^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{z})}{2\sigma^2}\right\}\\&=\exp\left\{-\frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}-2\boldsymbol{x}^{\mathrm{T}}\boldsymbol{z}+\boldsymbol{z}^{\mathrm{T}}\boldsymbol{z}}{2\sigma^2}\right\}\\&=\exp\left(-\frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}{2\sigma^2}\right)\cdot\exp\left(\frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{z}}{\sigma^2}\right)\cdot\exp\left(-\frac{\boldsymbol{z}^{\mathrm{T}}\boldsymbol{z}}{2\sigma^2}\right)\end{aligned}$$

由于指数函数 $y=e^x$ 的泰勒级数在任一点 (实数) 都收敛, 所以我们可以将上式的中间项做泰勒展开:

$$\exp\left(\frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{z}}{\sigma^2}\right)=\sum_{n=0}^{\infty}\frac{\left(\frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{z}}{\sigma^2}\right)^n}{n!}=\sum_{n=0}^{\infty}\frac{\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{z}\right)^n}{n!\sigma^{2n}}$$

于是核函数可以表示为

$$\mathcal{K}(\boldsymbol{x},\boldsymbol{z})=\exp\left(-\frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}{2\sigma^2}\right)\cdot\exp\left(-\frac{\boldsymbol{z}^{\mathrm{T}}\boldsymbol{z}}{2\sigma^2}\right)\cdot\sum_{n=0}^{\infty}\frac{\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{z}\right)^n}{n!\sigma^{2n}}$$

不妨设 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{z}=\sum_{i=1}^{k}x_iz_i$. 为了进行后续推导, 我们需要声明一下**推广的二项式定理**, 即

$$\left(\sum_{i=1}^{k}x_i\right)^n=\sum_{l=1}^{L}\frac{n!}{n_{l_1}!\cdot n_{l_2}!\cdots n_{l_k}!}x_1^{n_{l_1}}\cdot x_2^{n_{l_2}}\cdots x_k^{n_{l_k}}$$

其中 $\sum_{i=1}^{k}n_{l_i}=n, L=\frac{(n+k-1)!}{n!(k-1)!}$($L$ 也被称作多项式系数). 于是有:

$$\begin{aligned}\mathcal{K}(\boldsymbol{x},\boldsymbol{z})&=\exp\left(-\frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}{2\sigma^2}\right)\cdot\exp\left(-\frac{\boldsymbol{z}^{\mathrm{T}}\boldsymbol{z}}{2\sigma^2}\right)\cdot\sum_{n=0}^{\infty}\frac{\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{z}\right)^n}{n!\sigma^{2n}}\\&=\exp\left(-\frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}{2\sigma^2}\right)\cdot\exp\left(-\frac{\boldsymbol{z}^{\mathrm{T}}\boldsymbol{z}}{2\sigma^2}\right)\cdot\sum_{n=0}^{\infty}\frac{1}{n!\sigma^{2n}}\sum_{l=1}^{L}\frac{n!}{n_{l_1}!\cdot n_{l_2}!\cdots n_{l_k}!}(x_1z_1)^{n_{l_1}}\cdot(x_2z_2)^{n_{l_2}}\cdots(x_kz_k)^{n_{l_k}}\\&=\sum_{n=0}^{\infty}\sum_{l=1}^{L}\frac{\exp\left(-\frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}{2\sigma^2}\right)\cdot\exp\left(-\frac{\boldsymbol{z}^{\mathrm{T}}\boldsymbol{z}}{2\sigma^2}\right)}{\sigma^{2n}\cdot n_{l_1}!\cdot n_{l_2}!\cdots n_{l_k}!}\left(x_1^{n_{l_1}}\cdot x_2^{n_{l_2}}\cdots x_k^{n_{l_k}}\right)\cdot\left(z_1^{n_{l_1}}\cdot z_2^{n_{l_2}}\cdots z_k^{n_{l_k}}\right)\\&=\sum_{n=0}^{\infty}\sum_{l=1}^{L}\underbrace{\frac{\exp\left(-\frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}{2\sigma^2}\right)}{\sqrt{\sigma^{2n}\cdot n_{l_1}!\cdot n_{l_2}!\cdots n_{l_k}!}}\left(x_1^{n_{l_1}}\cdot x_2^{n_{l_2}}\cdots x_k^{n_{l_k}}\right)}_{:=\varphi_{n_l}(\boldsymbol{x})}\cdot\underbrace{\frac{\exp\left(-\frac{\boldsymbol{z}^{\mathrm{T}}\boldsymbol{z}}{2\sigma^2}\right)}{\sqrt{\sigma^{2n}\cdot n_{l_1}!\cdot n_{l_2}!\cdots n_{l_k}!}}\left(z_1^{n_{l_1}}\cdot z_2^{n_{l_2}}\cdots z_k^{n_{l_k}}\right)}_{:=\varphi_{n_l}(\boldsymbol{z})}\\&=\sum_{n=0}^{\infty}\sum_{l=1}^{L}\varphi_{n_l}(\boldsymbol{x})\cdot\varphi_{n_l}(\boldsymbol{z})\end{aligned}$$

我们令

$$\boldsymbol{\Phi}_n(\boldsymbol{x}) = [\varphi_{n_1}(\boldsymbol{x}), \varphi_{n_2}(\boldsymbol{x}), \cdots, \varphi_{n_L}(\boldsymbol{x})], \boldsymbol{\Phi}_n(\boldsymbol{z}) = [\varphi_{n_1}(\boldsymbol{z}), \varphi_{n_2}(\boldsymbol{z}), \cdots, \varphi_{n_L}(\boldsymbol{z})]$$

再令

$$\mathcal{K}_n(\boldsymbol{x}, \boldsymbol{z}) = \langle \boldsymbol{\Phi}_n(\boldsymbol{x}), \boldsymbol{\Phi}_n(\boldsymbol{z}) \rangle$$

于是有

$$\mathcal{K}(\boldsymbol{x}, \boldsymbol{z}) = \sum_{n=0}^{\infty} \sum_{l=1}^{L} \varphi_{n_l}(\boldsymbol{x}) \cdot \varphi_{n_l}(\boldsymbol{z}) = \sum_{n=0}^{\infty} \langle \boldsymbol{\Phi}_n(\boldsymbol{x}), \boldsymbol{\Phi}_n(\boldsymbol{z}) \rangle = \sum_{n=0}^{\infty} \mathcal{K}_n(\boldsymbol{x}, \boldsymbol{z})$$

显然高斯核可以表示为无限维特征向量的内积, 而且可知高斯核也可以表示成无限个核函数的线性组合.

## Problem 5

请证明, 无论数据空间的维数如何, 仅由两个数据点 $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}$(每个类一个样本) 所组成的数据集足以确定最大间隔超平面. 充分解释你的答案, 包括给出硬间隔 SVM(即 $\boldsymbol{w}$) 作为 $\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}$ 函数的解的明确公式.

**Solution:** 显然两个样本都是**支持向量**, 所以这两个点的垂直平分超平面即为硬间隔 SVM 的表达式. 不妨设 $\boldsymbol{x}^{(1)} = \left(x_1^{(1)}, x_2^{(1)}, \cdots, x_n^{(1)}\right)^{\mathrm{T}}, \boldsymbol{x}^{(2)} = \left(x_1^{(2)}, x_2^{(2)}, \cdots, x_n^{(2)}\right)^{\mathrm{T}}$, 先计算这两点连线的方向向量, 也就是垂直平分超平面的法向量:

$$\boldsymbol{n} = \boldsymbol{w} = \boldsymbol{x}^{(1)} - \boldsymbol{x}^{(2)} = \left(x_1^{(1)} - x_1^{(2)}, x_2^{(1)} - x_2^{(2)}, \cdots, x_n^{(1)} - x_n^{(2)}\right)^{\mathrm{T}}$$

而垂直平分超平面一定经过这两点连线的中点, 即

$$P_0 = \frac{\boldsymbol{x}^{(1)} + \boldsymbol{x}^{(2)}}{2} = \frac{1}{2}\left(x_1^{(1)} + x_1^{(2)}, x_2^{(1)} + x_2^{(2)}, \cdots, x_n^{(1)} + x_n^{(2)}\right)^{\mathrm{T}}$$

于是根据点法式即可写出如下垂直平分超平面的方程:

$$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} + b = 0 = \sum_{i=1}^{n}\left(x_i^{(1)} - x_i^{(2)}\right)x_i + b = 0 \xrightarrow{\text{代入点} P_0} b = -\sum_{i=1}^{n}\frac{1}{2}\left(x_i^{(1)} + x_i^{(2)}\right)\left(x_i^{(1)} - x_i^{(2)}\right)$$

于是最大间隔超平面方程为:

$$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} + b = \sum_{i=1}^{n}\left(x_i^{(1)} - x_i^{(2)}\right)\left[x_i - \frac{1}{2}\left(x_i^{(1)} + x_i^{(2)}\right)\right] = 0$$