THU-70250043-0, Pattern Recognition (Spring 2021) **Homework: 5**

SVM

Lecturer: Changshui Zhang zcs@mail.tsinghua.edu.cn

Hong Zhao vzhao@tsinghua.edu.cn

Student: Jingxuan Yang yangjx20@mails.tsinghua.edu.cn

Problem 1

Given an unlabeled set of examples $\left\{x^{(1)}, x^{(2)}, \ldots, x^{(n)}\right\}$ the one-class SVM algorithm tries to find a direction ω that maximally separates the data from the origin. (This turns out to be useful for anomaly detection.)

More precisely, it solves the (primal) optimization problem:

$$\begin{aligned}\min_{\omega} \ & \frac{1}{2}\omega^\top\omega \\ \text{s.t.} \ & \omega^\top x^{(i)} \geqslant 1,\ i = 1, 2, \ldots, n\end{aligned} \tag{1}$$

A new test example x is labeled 1 if $\omega^\top x \geqslant 1$, and 0 otherwise.

1.1 The primal optimization problem for the one-class SVM was given above. Write down the corresponding dual optimization problem. Simplify your answer as much as possible. In particular, ω should not appear in your answer.

解: Lagrange 函数为

$$L(\omega, \alpha) = \frac{1}{2}\omega^\top\omega - \sum_{i=1}^{n} \alpha_i(\omega^\top x^{(i)} - 1) \tag{2}$$

其对 ω 的偏导数为

$$\frac{\partial L(\omega, \alpha)}{\partial \omega} = \omega - \sum_{i=1}^{n} \alpha_i x^{(i)} \tag{3}$$

令此偏导数为 0, 可得

$$\omega = \sum_{i=1}^{n} \alpha_i x^{(i)} \tag{4}$$

代入 Lagrange 函数, 可得对偶函数为

$$Q(\alpha) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j[x^{(i)}]^\top x^{(j)} + \sum_{i=1}^{n}\alpha_i \tag{5}$$

所以, 对偶问题为

$$\begin{aligned} \max_{\alpha}\ & Q(\alpha) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j[x^{(i)}]^\top x^{(j)} + \sum_{i=1}^{n}\alpha_i \\ \text{s.t.}\quad & \alpha_i \geqslant 0,\ i = 1, 2, \ldots, n \end{aligned} \tag{6}$$

1.2 Can the one-class SVM be kernelized (both in training and testing)? Justify your answer.

解: 由于样本点只以内积形式存在, 则可以使用核函数进行训练和测试.

记核函数为 $K(\cdot,\cdot)$, 则在训练时需要解如下问题

$$\begin{aligned} \max_{\alpha}\ & Q(\alpha) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j K(x^{(i)}, x^{(j)}) + \sum_{i=1}^{n}\alpha_i \\ \text{s.t.}\quad & \alpha_i \geqslant 0,\ i = 1, 2, \ldots, n \end{aligned} \tag{7}$$

求得上式最优解 α^* 后, 对新样本 x 测试时计算

$$f(x) = \sum_{i=1}^{n}\alpha_i^* K(x^{(i)}, x) \tag{8}$$

若 $f(x) \geqslant 1$ 则将 x 标记为 1, 否则标记为 0.

Problem 2

Consider a dataset with 2 points in 1-D: $x_1 = 0$, $x_2 = \sqrt{2}$ with labels $y_1 = -1$, $y_2 = 1$. Consider mapping each point to 3-D using the feature vector $\Phi = [1, \sqrt{2}x, x^2]^\top$. (This is equivalent to using a second order polynomial kernel.) The max margin classifier has the form:

$$\begin{aligned} \min_{w}\ & \|w\|^2 \\ \text{s.t.}\quad & y_1(w^\top\Phi(x_1) + b) \geqslant 1 \\ & y_2(w^\top\Phi(x_2) + b) \geqslant 1 \end{aligned} \tag{9}$$

2.1. Write down a vector that is parallel to the optimal vector w.

解: 与最优解 w 平行的一个向量可以为

$$\bar{w} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \tag{10}$$

2.2. What is the value of the margin that is achieved by this w? Hint 1: recall that the margin is the distance from each support vector to the decision boundary. Hint 2: think about the geometry of 2 points in space, with a line separating one from the other.

解: 根据 Hint 1 对 margin 的定义可知

$$\text{margin} = \sqrt{2} \tag{11}$$

2.3. Solve for w, using the fact the margin is equal to $1/\|w\|$.

解: 由 margin 值可知

$$\|w\| = \frac{1}{\sqrt{2}} \tag{12}$$

则

$$w = \frac{\|w\|}{\|\bar{w}\|}\bar{w} = \begin{bmatrix} 0 \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} \tag{13}$$

2.4. Write down the form of the discriminant function $f(x) = w^\top \Phi(x) + b$ as an explicit function of x.

解: 由上题可知

$$f(x) = w^\top \Phi(x) + b = \frac{\sqrt{2}}{2}x + \frac{1}{2}x^2 + b \tag{14}$$

又 $y_1 f(x_1) = y_2 f(x_2) = 1$ 可知

$$b = -1 \tag{15}$$

则判别函数为

$$f(x) = \frac{1}{2}x^2 + \frac{\sqrt{2}}{2}x - 1 \tag{16}$$

Problem 3

The exclusive-OR is the simplest problem that cannot be solved using a linear discriminant operating directly on the features. The points $k = 1, 3$ at $\boldsymbol{x} = (1, 1)^\top$ and $\boldsymbol{x} = (-1, -1)^\top$ are in category ω_1 (red in Figure 1), while $k = 2, 4$ at $\boldsymbol{x} = (1, -1)^\top$ and $\boldsymbol{x} = (-1, 1)^\top$ are in ω_2 (black in Figure 1). Following the approach of SVM, we use kernel function to map the features to a higher dimension space where they can be linearly separated.

3.1 Consider the polynomial kernel of degree 2:

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x_i}^\top \boldsymbol{x_j} + 1)^2, \tag{17}$$

where $\boldsymbol{x}_i = (x_{i1}, x_{i2})^\top$ and $\boldsymbol{x}_j = (x_{j1}, x_{j2})^\top$.

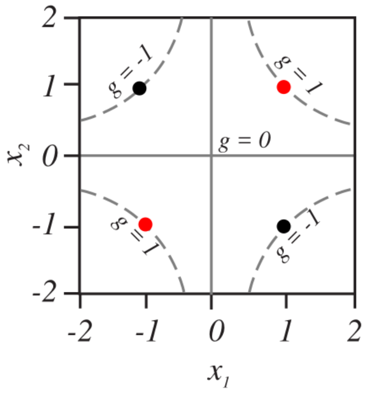

图 1: The XOR problem in the original (x_1, x_2) feature space.

Show that it corresponds to mapping

$$\Phi(x_1,x_2) = \left[1, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1x_2, x_1^2, x_2^2\right]. \tag{18}$$

解: 对核函数有

$$\begin{aligned} K(\boldsymbol{x}_i, \boldsymbol{x}_j) &= (\boldsymbol{x_i}^\top \boldsymbol{x_j} + 1)^2 \\ &= (x_{i1}x_{j1} + x_{i2}x_{j2} + 1)^2 \\ &= 1 + x_{i1}^2 x_{j1}^2 + x_{i2}^2 x_{j2}^2 + 2x_{i1}x_{j1} + 2x_{i2}x_{j2} + 2x_{i1}x_{j1}x_{i2}x_{j2} \end{aligned} \tag{19}$$

又映射的内积为

$$\Phi(x_{i1}, x_{i2}) \cdot \Phi(x_{j1}, x_{j2}) = 1 + 2x_{i1}x_{j1} + 2x_{i2}x_{j2} + 2x_{i1}x_{j1}x_{i2}x_{j2} + x_{i1}^2 x_{j1}^2 + x_{i2}^2 x_{j2}^2 \tag{20}$$

则

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \Phi(x_{i1}, x_{i2}) \cdot \Phi(x_{j1}, x_{j2}) \tag{21}$$

因此核函数对应的映射为 $\Phi(\cdot)$.

3.2 Derive the dual problem in the 6-dimensional space with Lagrange multipliers α_i, $i = 1, 2, 3, 4$ as the only variants.

解: 记 $y_1 = 1$, $y_2 = -1$, $y_3 = 1$, $y_4 = -1$,

$$\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{x}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \boldsymbol{x}_3 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \quad \boldsymbol{x}_4 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \tag{22}$$

则原问题为

$$\begin{aligned} &\min_{\boldsymbol{\omega}} \ \frac{1}{2}\|\boldsymbol{\omega}\|_2^2 \\ &\text{s.t.} \quad y_i[\boldsymbol{\omega}^\top \Phi(\boldsymbol{x}_i) - b] - 1 \geqslant 0, \ i = 1, 2, 3, 4 \end{aligned} \tag{23}$$

引入对偶变量 $\boldsymbol{\alpha}$, 则 Lagrange 函数为

$$L(\boldsymbol{\omega}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\boldsymbol{\omega}\|_2^2 - \sum_{i=1}^{4} \alpha_i [y_i \boldsymbol{\omega}^\top \Phi(\boldsymbol{x}_i) - y_i b - 1] \tag{24}$$

其对 $\boldsymbol{\omega}$ 的偏导数为

$$\frac{\partial L(\boldsymbol{\omega}, b, \boldsymbol{\alpha})}{\partial \boldsymbol{\omega}} = \boldsymbol{\omega} - \sum_{i=1}^{4} \alpha_i y_i \Phi(\boldsymbol{x}_i) \tag{25}$$

令此偏导数为 0, 可得

$$\boldsymbol{\omega} = \sum_{i=1}^{4} \alpha_i y_i \Phi(\boldsymbol{x}_i) \tag{26}$$

Lagrange 函数对 b 的偏导数为

$$\frac{\partial L(\boldsymbol{\omega}, b, \boldsymbol{\alpha})}{\partial b} = \sum_{i=1}^{4} \alpha_i y_i \tag{27}$$

令此偏导数为 0, 可得

$$\sum_{i=1}^{4} \alpha_i y_i = 0 \tag{28}$$

将 $\boldsymbol{\omega}$ 代入 Lagrange 函数, 可得对偶函数为

$$Q(\boldsymbol{\alpha}) = -\frac{1}{2} \sum_{i=1}^{4} \sum_{j=1}^{4} \alpha_i \alpha_j y_i y_j \Phi(\boldsymbol{x}_i) \cdot \Phi(\boldsymbol{x}_j) + \sum_{i=1}^{4} \alpha_i \tag{29}$$

所以, 对偶问题为

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \ & Q(\boldsymbol{\alpha}) = -\frac{1}{2} \sum_{i=1}^{4} \sum_{j=1}^{4} \alpha_i \alpha_j y_i y_j \Phi(\boldsymbol{x}_i) \cdot \Phi(\boldsymbol{x}_j) + \sum_{i=1}^{4} \alpha_i \\ \text{s.t.} \quad & \sum_{i=1}^{4} \alpha_i y_i = 0 \\ & \alpha_i \geqslant 0, \ i = 1, 2, 3, 4 \end{aligned} \tag{30}$$

3.3 Solve the dual problem analytically (without programming). What are the support vectors? Hint: use the symmetry of the problem.

解: 对偶函数为

$$\begin{aligned} Q(\boldsymbol{\alpha}) &= -\frac{1}{2} \sum_{i=1}^{4} \sum_{j=1}^{4} \alpha_i \alpha_j y_i y_j \Phi(\boldsymbol{x}_i) \cdot \Phi(\boldsymbol{x}_j) + \sum_{i=1}^{4} \alpha_i \\ &= \sum_{i=1}^{4} \alpha_i - \frac{1}{2} \sum_{i=1}^{4} \sum_{j=1}^{4} \alpha_i \alpha_j y_i y_j K(\boldsymbol{x}_i, \boldsymbol{x}_j) \end{aligned} \tag{31}$$

其中 $\forall\ i, j = 1, 2, 3, 4$, 核函数

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \begin{cases} 9, & \text{if } i = j \\ 1, & \text{otherwise} \end{cases} \tag{32}$$

则对偶函数对 α_i 的偏导为

$$\begin{aligned} \frac{\partial Q(\boldsymbol{\alpha})}{\partial \alpha_i} &= 1 - \alpha_i y_i y_i K(\boldsymbol{x}_i, \boldsymbol{x}_i) - \sum_{j \neq i} \alpha_j y_i y_j K(\boldsymbol{x}_i, \boldsymbol{x}_j) \\ &= 1 - 9\alpha_i - y_i \sum_{j \neq i} \alpha_j y_j, \quad \forall\ i = 1, 2, 3, 4 \end{aligned} \tag{33}$$

又由对偶问题约束可知

$$\sum_{j \neq i} \alpha_j y_j = \sum_{j=1}^{4} \alpha_j y_j - \alpha_i y_i = -\alpha_i y_i, \quad \forall\ i = 1, 2, 3, 4 \tag{34}$$

代入偏导数表达式可得

$$\frac{\partial Q(\boldsymbol{\alpha})}{\partial \alpha_i} = 1 - 9\alpha_i + \alpha_i y_i y_i = 1 - 8\alpha_i, \quad \forall\ i = 1, 2, 3, 4 \tag{35}$$

令这些偏导数为 0, 有

$$\alpha_i = \frac{1}{8}, \quad \forall\ i = 1, 2, 3, 4 \tag{36}$$

则由互补松弛 (complementary slackness) 性可知, $\boldsymbol{x}_i$, $i = 1, 2, 3, 4$ 都是支持向量.

3.4 Derive the final discriminant function $g(\boldsymbol{x})$ and the decision hyperplane.

解: 由上题可知

$$\begin{aligned} \boldsymbol{\omega} &= \sum_{i=1}^{4} \alpha_i y_i \Phi(\boldsymbol{x}_i) \\ &= \frac{1}{8} \left[\begin{pmatrix} 1 \\ \sqrt{2} \\ \sqrt{2} \\ \sqrt{2} \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ \sqrt{2} \\ -\sqrt{2} \\ -\sqrt{2} \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ -\sqrt{2} \\ -\sqrt{2} \\ \sqrt{2} \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ -\sqrt{2} \\ \sqrt{2} \\ -\sqrt{2} \\ 1 \\ 1 \end{pmatrix} \right] \\ &= \frac{1}{8} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 4\sqrt{2} \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \frac{\sqrt{2}}{2} \\ 0 \\ 0 \end{pmatrix} \end{aligned} \tag{37}$$

所以

$$\begin{aligned} g(\boldsymbol{x}) &= \boldsymbol{\omega}^\top \Phi(\boldsymbol{x}) - b \\ &= x_1 x_2 - b \end{aligned} \tag{38}$$

又 $\boldsymbol{x}_i$, $i = 1, 2, 3, 4$ 都是支持向量, 则有

$$y_i g(\boldsymbol{x}_i) - 1 = 0, \quad \forall\ i = 1, 2, 3, 4 \tag{39}$$

从而解得

$$b = 0 \tag{40}$$

因此, 判别函数为

$$g(\boldsymbol{x}) = x_1 x_2 \tag{41}$$

决策超平面为

$$g(\boldsymbol{x}) = x_1 x_2 = 0 \tag{42}$$

3.5 Plot the data points on the subspace of $(\sqrt{2}x_1, \sqrt{2}x_1x_2)$. Demonstrate the decision hyperplane and the margin in your plot.

解: 映射到子空间 $(\sqrt{2}x_1, \sqrt{2}x_1x_2)$ 的四个数据点, 决策超平面以及间隔如图 2 所示.

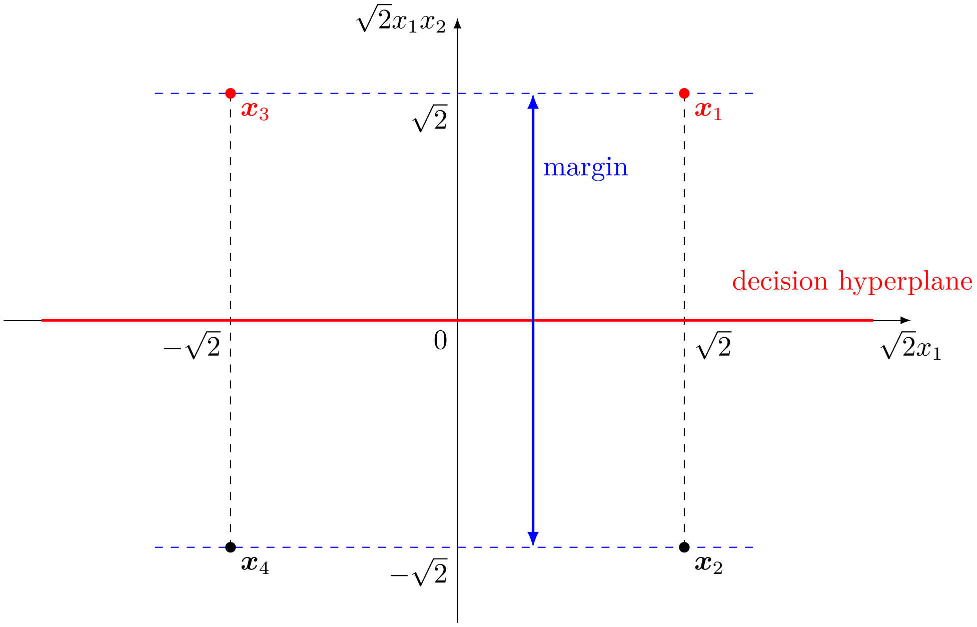

图 2: XOR 问题映射到子空间 $(\sqrt{2}x_1, \sqrt{2}x_1x_2)$, 决策超平面以及间隔

Problem 4

Write a program to complete the following task adopting the SVM algorithm (you could use some toolkits or source code). Train a SVM classifier with data from ω_1 and ω_2 in the following table. Preprocess each training pattern to form a new vector having components $1, x_1, x_2, x_1^2, x_1x_2$ and x_2^2.

	ω_1		ω_2	
sample	x_1	x_2	x_1	x_2
1	-3.0	-2.9	-2.0	-8.4
2	0.5	8.7	-8.9	0.2
3	2.9	2.1	-4.2	-7.7
4	-0.1	5.2	-8.5	-3.2
5	-4.0	2.2	-6.7	-4.0
6	-1.3	3.7	-0.5	-9.2
7	-3.4	6.2	-5.3	-6.7
8	-4.1	3.4	-8.7	-6.4
9	-5.1	1.6	-7.1	-9.7
10	1.9	5.1	-8.0	-6.3

Hint: You needn't program the SVM algorithm by yourself, you can just use some toolkits or source code such as libsvm for MATLAB or scikit-learn for python. You should declare the toolkit you used in your project.

4.1 Train you classifier with just the first point in the ω_1 and ω_2 and find the separating hyperplane and the margin.

解: 使用 MATLAB 内置 `fitcsvm` 函数, 对两个类别的第一个点训练 SVM 得到分类边界如图 3 中蓝色实线所示, 分类超平面的表达式为

$$g(\boldsymbol{x}) = w_1 + w_2x_1 + w_3x_2 + w_4x_1^2 + w_5x_1x_2 + w_6x_2^2 + b = 0 \tag{43}$$

其中 w_i, $i = 1, 2, \ldots, 6$ 的数值见表 1 中第一行, 且间隔 (margin) 为

$$\text{margin} = 63.1228 \tag{44}$$

4.2 Repeat 4.1 with the first two points in each category (four points total), the first three points and so on, until the transformed patterns cannot be linearly separated in the transformed space.

解: 依次使用前 2 至前 10 个点训练 SVM 得到分类边界如图 4 所示, 分类超平面的参数以及间隔如表 1 所示. 由图可知依次生成的训练集数据在转换后的空间上均是线性可分的, 随着训练点数的增多, 间隔逐渐减小, 但是如果新增的点不是支持向量, 则对分类面没有影响.

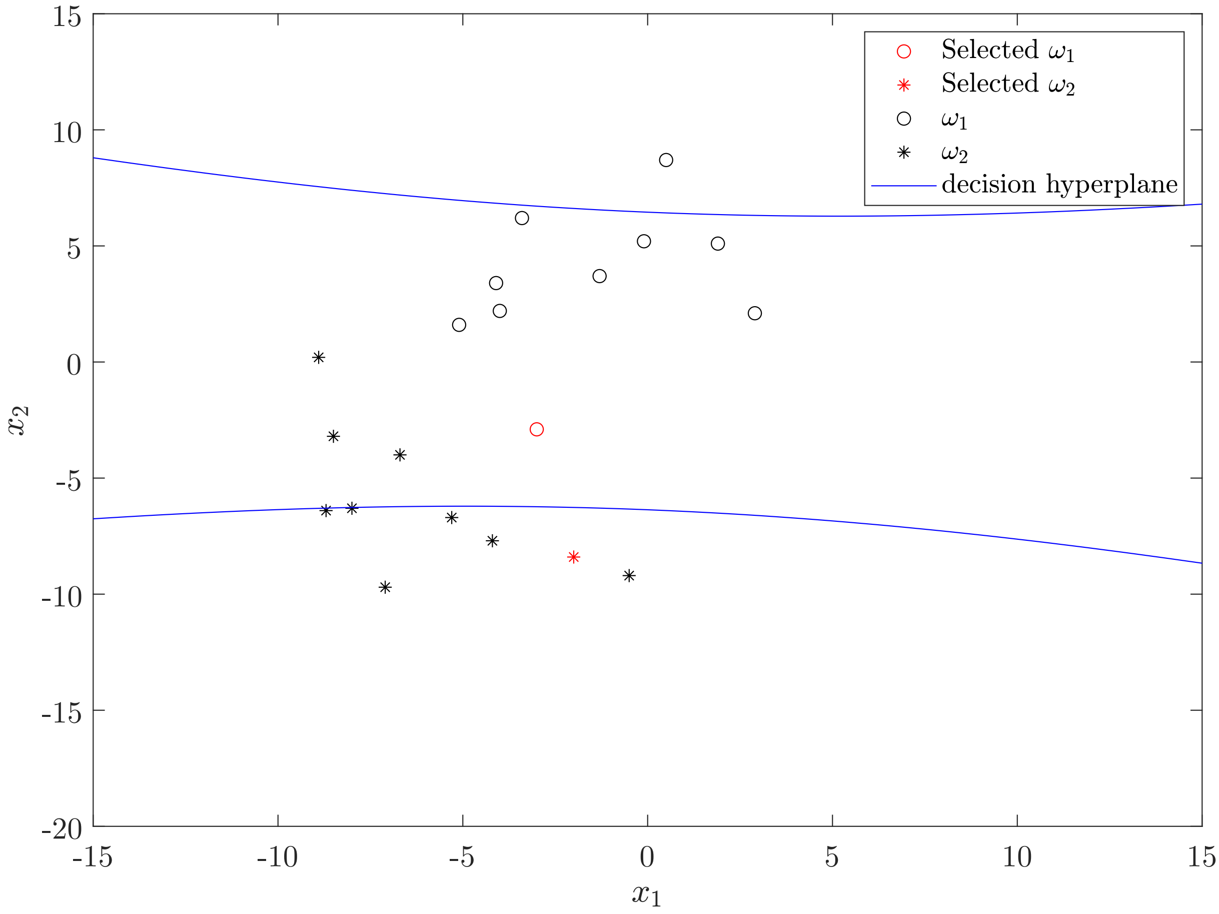

图 3: 使用两类第一个点训练 SVM 得到分类边界

表 1: 分别使用前 1 至 10 个点训练 SVM 得到分类超平面参数和间隔

# of Points	w_1	w_2	w_3	w_4	w_5	w_6	b	margin
1	0	-0.0005	0.0028	0.0025	-0.0041	-0.0312	1.2816	63.1228
2	0	0.0121	0.0739	-0.0413	-0.0534	-0.0222	2.2739	19.3674
3	0	0.0121	0.0739	-0.0413	-0.0534	-0.0222	2.2739	19.3674
4	0	0.0121	0.0739	-0.0413	-0.0534	-0.0222	2.2739	19.3674
5	0	0.0121	0.0739	-0.0413	-0.0534	-0.0222	2.2740	19.3686
6	0	0.0111	0.0989	-0.0379	-0.0210	-0.0239	2.0453	17.9957
7	0	0.0111	0.0989	-0.0379	-0.0210	-0.0239	2.0453	17.9957
8	0	0.0111	0.0989	-0.0379	-0.0210	-0.0239	2.0453	17.9957
9	0	0.0111	0.0989	-0.0379	-0.0210	-0.0239	2.0451	17.9976
10	0	0.0111	0.0989	-0.0379	-0.0210	-0.0239	2.0451	17.9976

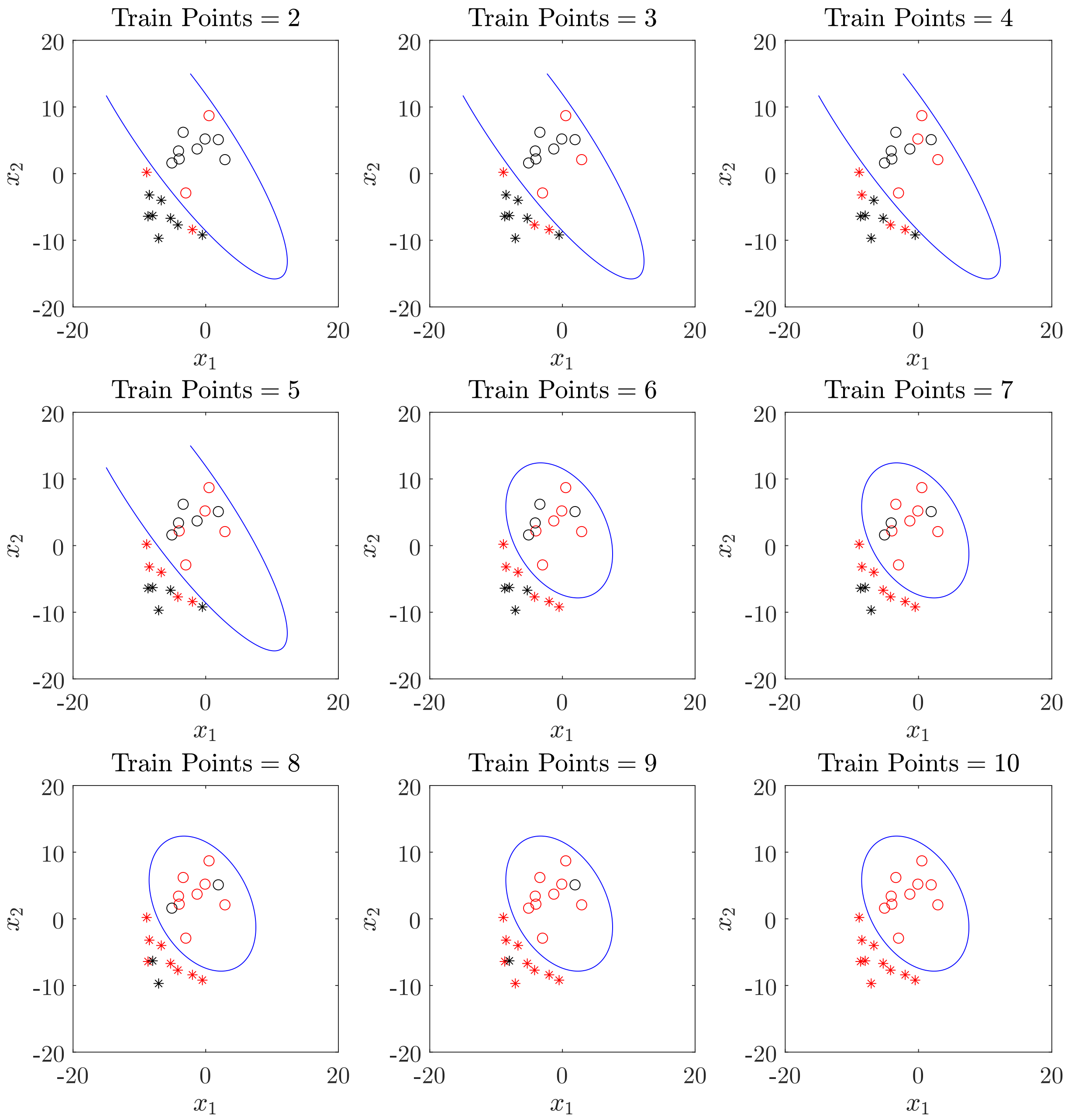

图 4: 依次使用前 2 至前 10 个点训练 SVM 得到分类边界