

支持向量机

(Support Vector Machine, SVM)

zyz

October 7, 2025

Contents

1	SVM 的基本思想	2
1.1	分类与超平面	2
1.2	间隔与支持向量	2
2	线性可分情形	2
2.1	硬间隔最大化	2
2.2	对偶问题与 KKT 条件	4
2.3	SVM 情形下的强对偶与 KKT 条件	5
2.4	SMO 算法	6
2.5	硬间隔 SVM 优化问题求解	6
3	近似线性可分情形	7
3.1	引入松弛变量与惩罚系数	7
3.2	软间隔 SVM 问题求解	8
4	线性不可分情形与核函数	10

1 SVM 的基本思想

1.1 分类与超平面

给定训练数据集：

$$\{(\mathbf{x}_i, y_i) \mid \mathbf{x}_i \in \mathbb{R}^n, y_i \in \{+1, -1\}, i = 1, 2, \dots, m\},$$

其中， $\mathbf{x}_i$ 表示 n 维特征向量， y_i 为类别标签。

若数据线性可分，则存在超平面：

$$\mathbf{w}^\top \mathbf{x} + b = 0,$$

可以将样本划分为两类：

$$\begin{cases} \mathbf{w}^\top \mathbf{x} + b \geq 0, & y = +1, \\ \mathbf{w}^\top \mathbf{x} + b < 0, & y = -1. \end{cases}$$

1.2 间隔与支持向量

对于超平面

$$\mathbf{w}^\top \mathbf{x} + b = 0,$$

其几何意义是将样本空间划分为两类。定义一个样本点 $\mathbf{x}_i$ 到超平面的距离为

$$d_i = \frac{|\mathbf{w}^\top \mathbf{x}_i + b|}{\|\mathbf{w}\|}.$$

为了保证所有样本点都能被正确分类，需要满足：

$$y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1, \quad i = 1, 2, \dots, m.$$

此时，距离超平面最近的样本点称为**支持向量**。对应的**间隔 (margin)** 为

$$\gamma = \frac{2}{\|\mathbf{w}\|},$$

其中分子 2 来源于两类支持向量到超平面的距离之和。

2 线性可分情形

2.1 硬间隔最大化

在数据线性可分的情况下，我们希望找到一个能够正确分类所有样本的超平面，并且使得两类样本到该超平面的间隔最大。设分离超平面为

$$\mathbf{w}^\top \mathbf{x} + b = 0,$$

过支持向量的超平面为（超平面参数可成倍扩大/缩小）

$$\begin{cases} \mathbf{w}^\top \mathbf{x}_1 + b = 1, & y = +1, \\ \mathbf{w}^\top \mathbf{x}_2 + b = -1, & y = -1. \end{cases}$$

作差，得

$$\begin{aligned} (\mathbf{w}^\top \mathbf{x}_1 + b) - (\mathbf{w}^\top \mathbf{x}_2 + b) &= 1 - (-1) \\ \Rightarrow \mathbf{w}^\top (\mathbf{x}_1 - \mathbf{x}_2) &= 2. \end{aligned}$$

即

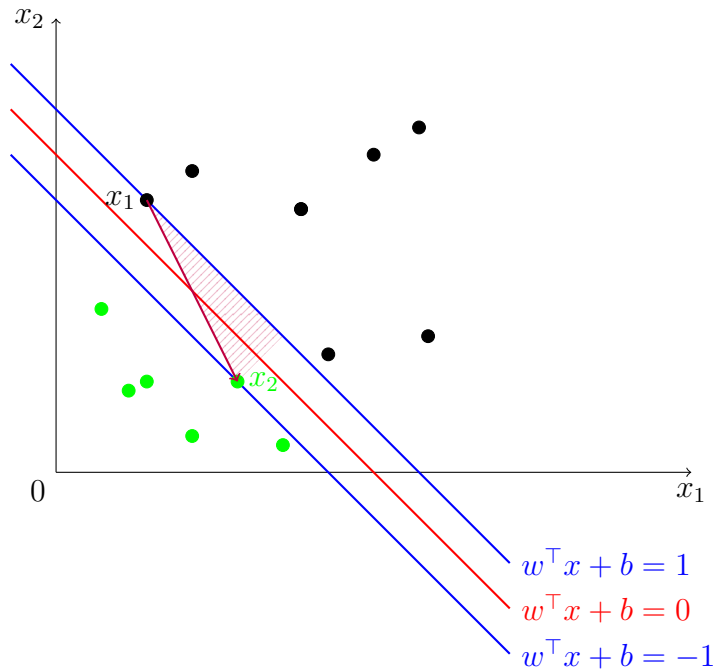
$$\|\mathbf{w}\| \cdot \|\mathbf{x}_1 - \mathbf{x}_2\| \cdot \cos \theta = 2,$$

得到

$$\|\mathbf{w}\| \cdot d = 2,$$

其中 d 为两类支持向量到超平面的距离之和。因此，间隔为

$$\gamma = \frac{2}{\|\mathbf{w}\|}.$$



最大化间隔 γ 等价于最小化 $\|\mathbf{w}\|$ ，同时还要满足所有样本点被正确分类的约束条件

$$y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1, \quad i = 1, 2, \dots, m.$$

因此，硬间隔最大化问题可以表述为以下优化问题：

$$\begin{aligned} \min_{\mathbf{w}, b} \quad & \frac{1}{2} \|\mathbf{w}\|^2 \\ \text{s.t.} \quad & y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1, \quad i = 1, 2, \dots, m. \end{aligned}$$

2.2 对偶问题与 KKT 条件

一般来说，一个带约束的优化问题（称为原问题）可以写成：

原问题：

$$\begin{aligned} \min_{\mathbf{x}} \quad & f_0(\mathbf{x}) \\ \text{s.t.} \quad & g_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \dots, m, \\ & h_j(\mathbf{x}) = 0, \quad j = 1, 2, \dots, p. \end{aligned}$$

为处理约束问题，引入拉格朗日乘子 $\alpha_i \geq 0$ 与 β_j ，定义拉格朗日函数：

$$L(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = f_0(\mathbf{x}) + \sum_{i=1}^m \alpha_i g_i(\mathbf{x}) + \sum_{j=1}^p \beta_j h_j(\mathbf{x}).$$

对偶函数定义为

$$\theta(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}),$$

于是得到对偶问题：

对偶问题：

$$\begin{aligned} \max_{\boldsymbol{\alpha}, \boldsymbol{\beta}} \quad & \theta(\boldsymbol{\alpha}, \boldsymbol{\beta}) \\ \text{s.t.} \quad & \alpha_i \geq 0, \quad i = 1, 2, \dots, m. \end{aligned}$$

原问题与对偶问题之间始终满足弱对偶性：

$$f_0(\mathbf{x}^*) \geq \theta(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*).$$

弱对偶性始终成立，即对偶问题的最优值不超过原问题的最优值。

定义：原问题与对偶问题间的间距为

$$G = f_0(\mathbf{x}^*) - \theta(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*).$$

当 $G = 0$ 时，称原问题与对偶问题之间具有强对偶性。形式化地，有以下定理：

强对偶定理：若 $(\mathbf{x}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ 分别为原问题与对偶问题的最优解，则有

$$\inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) = L(\mathbf{x}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*),$$

即对偶间隙（duality gap）

$$G = f_0(\mathbf{x}^*) - \theta(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) = 0$$

换言之，在最优解处，原问题与对偶问题的目标函数值完全相等。

进一步地，若原问题是凸优化问题，并且满足 **Slater 条件**（即存在一个严格可行解 $\mathbf{x}$ ，使得 $g_i(\mathbf{x}) < 0$ 且 $h_j(\mathbf{x}) = 0$ ）。

$$f_0(\mathbf{x}^*) = \theta(\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*).$$

在强对偶性成立的情况下，原问题与对偶问题的最优解需同时满足 **KKT 条件**。

KKT 条件是原问题与对偶问题同时成立的必要条件（对于凸优化问题也是充分条件）：

KKT 条件：若强对偶性成立，则最优解 $(\mathbf{x}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$ 必须满足以下四个条件：

- 原始可行性： $g_i(\mathbf{x}^*) \leq 0, \quad h_j(\mathbf{x}^*) = 0$
- 对偶可行性： $\alpha_i^* \geq 0$
- 互补松弛性： $\alpha_i^* g_i(\mathbf{x}^*) = 0$
- 梯度最优性：

$$\nabla f_0(\mathbf{x}^*) + \sum_{i=1}^m \alpha_i^* \nabla g_i(\mathbf{x}^*) + \sum_{j=1}^p \beta_j^* \nabla h_j(\mathbf{x}^*) = 0$$

2.3 SVM 情形下的强对偶与 KKT 条件

对于支持向量机（SVM）问题，原函数 $f(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2$ 是凸函数，约束 $g_i(\mathbf{w}, b) = 1 - y_i(\mathbf{w}^\top \mathbf{x}_i + b)$ 是线性的，因此满足凸性与 Slater 条件。由此可知，SVM 问题具有 **强对偶性**，即 $G = 0$ 。

此时，其 KKT 条件具体化为：

$$\begin{cases} \mathbf{w} = \sum_{i=1}^m \alpha_i y_i \mathbf{x}_i, \\ \sum_{i=1}^m \alpha_i y_i = 0, \\ \alpha_i \geq 0, \\ \alpha_i (y_i(\mathbf{w}^\top \mathbf{x}_i + b) - 1) = 0. \end{cases}$$

由互补松弛条件可知：

- 若 $\alpha_i^* = 0$ ，样本点 $\mathbf{x}_i$ 位于间隔外，不影响超平面；
- 若 $\alpha_i^* > 0$ ，则必有 $y_i(\mathbf{w}^{*\top} \mathbf{x}_i + b^*) = 1$ ，该样本为支持向量。

2.4 SMO 算法

<https://zhuanlan.zhihu.com/p/367578887>

SMO 算法步骤：

1. 从所有样本中选择两个拉格朗日乘子 α_i 和 α_j 进行优化；
2. 固定其他乘子，求解关于 α_i 、 α_j 的二次规划子问题；
3. 更新参数 $\mathbf{w}$ 与偏置项 b ；
4. 重复以上步骤，直至所有 KKT 条件满足（或目标函数收敛）。

SMO 通过每次仅更新两个变量来简化二次规划的求解过程，显著提高了 SVM 训练的效率。

2.5 硬间隔 SVM 优化问题求解

硬间隔最大化问题：

$$\begin{aligned} \min_{\mathbf{w}, b} \quad & \frac{1}{2} \|\mathbf{w}\|^2 \\ \text{s.t.} \quad & y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1, \quad i = 1, 2, \dots, m. \end{aligned}$$

引入拉格朗日乘子 $\alpha_i \geq 0$ ，构造拉格朗日函数：

$$\begin{aligned} L(\mathbf{w}, b, \boldsymbol{\alpha}) &= \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^m \alpha_i (y_i(\mathbf{w}^\top \mathbf{x}_i + b) - 1) \\ &= \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^m \alpha_i y_i \mathbf{w}^\top \mathbf{x}_i - b \sum_{i=1}^m \alpha_i y_i + \sum_{i=1}^m \alpha_i. \end{aligned}$$

对 $\mathbf{w}$ 和 b 求偏导并令其为零，得到最优条件：

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^m \alpha_i y_i \mathbf{x}_i = 0, \quad \frac{\partial L}{\partial b} = - \sum_{i=1}^m \alpha_i y_i = 0,$$

即

$$\mathbf{w} = \sum_{i=1}^m \alpha_i y_i \mathbf{x}_i, \quad \sum_{i=1}^m \alpha_i y_i = 0.$$

将其代入原拉格朗日函数，得到最小值：

$$\begin{aligned} L(\mathbf{w}, b, \boldsymbol{\alpha})_{\min} &= \frac{1}{2} \left(\sum_{i=1}^m \alpha_i y_i \mathbf{x}_i \right)^\top \left(\sum_{j=1}^m \alpha_j y_j \mathbf{x}_j \right) - \sum_{i=1}^m \alpha_i y_i \left(\sum_{j=1}^m \alpha_j y_j \mathbf{x}_j^\top \mathbf{x}_i \right) - b \sum_{i=1}^m \alpha_i y_i + \sum_{i=1}^m \alpha_i \\ &= \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j (x_i^\top x_j) - \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j (x_i^\top x_j) + \sum_{i=1}^m \alpha_i - b \sum_{i=1}^m \alpha_i y_i \\ &= \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j (x_i^\top x_j). \end{aligned}$$

因此, SVM 的对偶问题为:

SVM 对偶问题:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j (\mathbf{x}_i^\top \mathbf{x}_j) \\ \text{s.t.} \quad & \alpha_i \geq 0, \quad i = 1, 2, \dots, m, \\ & \sum_{i=1}^m \alpha_i y_i = 0. \end{aligned}$$

该问题为一个凸二次规划问题, 可通过 SMO 算法求解最优拉格朗日乘子 $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_m^*)$ 。

由 KKT 条件可知:

$$\alpha_i^* (y_i (\mathbf{w}^{*\top} \mathbf{x}_i + b^*) - 1) = 0, \quad i = 1, 2, \dots, m.$$

又由最优条件:

$$\mathbf{w}^* = \sum_{i=1}^m \alpha_i^* y_i \mathbf{x}_i, \quad \sum_{i=1}^m \alpha_i^* y_i = 0.$$

当 $\alpha_i^* = 0$ 时, $\mathbf{x}_i$ 对分类结果无影响, 称为非支持向量;

当 $\alpha_i^* > 0$ 时, 根据 KKT 条件,

$$y_i (\mathbf{w}^{*\top} \mathbf{x}_i + b^*) = 1,$$

此时 $\mathbf{x}_i$ 为支持向量。可得:

$$y_i^2 (\mathbf{w}^{*\top} \mathbf{x}_i + b^*) = y_i$$

即

$$\mathbf{w}^{*\top} \mathbf{x}_i + b^* = y_i.$$

由此可得偏置项:

$$b^* = y_i - \mathbf{w}^{*\top} \mathbf{x}_i = y_i - \sum_{j=1}^m \alpha_j^* y_j \mathbf{x}_j^\top \mathbf{x}_i.$$

3 近似线性可分情形

3.1 引入松弛变量与惩罚系数

在实际问题中, 样本往往存在噪声或标注错误, 数据难以做到完全线性可分。此时, 允许少量样本点违反分类约束条件, 引入松弛变量 $\xi_i \geq 0$, 表示样本点 $\mathbf{x}_i$ 对间隔约束的违反程度。

于是，约束条件修改为：

$$y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, m.$$

有如下三种情况：

- 若 $\xi_i = 0$ ，则样本点 $\mathbf{x}_i$ 被正确分类；
- 若 $0 < \xi_i \leq 1$ ，则样本点 $\mathbf{x}_i$ 被正确分类，存在噪声；
- 若 $\xi_i > 1$ ，则样本点 $\mathbf{x}_i$ 被误分类。

同时，在目标函数中加入对松弛变量的惩罚项，以平衡“间隔最大化”与“分类错误最小化”：

$$\min_{\mathbf{w}, b, \xi} \quad \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m \xi_i,$$

其中 $C > 0$ 为**惩罚系数**，用于控制模型复杂度与分类错误之间的权衡：

- 当 C 较大时，对误分类样本惩罚更强，模型趋向于严格分类；
- 当 C 较小时，模型允许更多误差，从而获得更好的泛化能力。

该模型被称为**软间隔支持向量机 (Soft Margin SVM)**。

软间隔最大化问题：

$$\begin{aligned} \min_{\mathbf{w}, b, \xi} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m \xi_i \\ \text{s.t.} \quad & y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i, \quad i = 1, 2, \dots, m, \\ & \xi_i \geq 0, \quad i = 1, 2, \dots, m. \end{aligned}$$

3.2 软间隔 SVM 问题求解

利用拉格朗日乘数法，引入拉格朗日乘子：对约束 $y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i$ 引入 $\alpha_i \geq 0$ ，对 $\xi_i \geq 0$ 引入 $\beta_i \geq 0$ 。构造拉格朗日函数：

$$\begin{aligned} L(\mathbf{w}, b, \xi, \alpha, \beta) &= \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m \xi_i - \sum_{i=1}^m \alpha_i (y_i(\mathbf{w}^\top \mathbf{x}_i + b) - 1 + \xi_i) - \sum_{i=1}^m \beta_i \xi_i \\ &= \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m \xi_i - \sum_{i=1}^m \alpha_i y_i \mathbf{w}^\top \mathbf{x}_i - b \sum_{i=1}^m \alpha_i y_i + \sum_{i=1}^m \alpha_i - \sum_{i=1}^m \alpha_i \xi_i - \sum_{i=1}^m \beta_i \xi_i \\ &= \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^m \alpha_i y_i \mathbf{w}^\top \mathbf{x}_i - b \sum_{i=1}^m \alpha_i y_i + \sum_{i=1}^m (C - \alpha_i - \beta_i) \xi_i + \sum_{i=1}^m \alpha_i. \end{aligned}$$

对 $\mathbf{w}, b, \xi_i$ 分别求偏导并令零（求对偶函数的极小化），得到必要条件：

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^m \alpha_i y_i \mathbf{x}_i = 0 \quad \Rightarrow \quad \mathbf{w} = \sum_{i=1}^m \alpha_i y_i \mathbf{x}_i,$$

$$\frac{\partial L}{\partial b} = - \sum_{i=1}^m \alpha_i y_i = 0 \quad \Rightarrow \quad \sum_{i=1}^m \alpha_i y_i = 0,$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \beta_i = 0 \quad \Rightarrow \quad \alpha_i + \beta_i = C.$$

将 $\mathbf{w} = \sum_{i=1}^m \alpha_i y_i \mathbf{x}_i$ 和 $\sum_{i=1}^m \alpha_i y_i = 0$ 代入拉格朗日函数：

$$\begin{aligned} L_{\min}(\alpha, \beta) &= \frac{1}{2} \left\| \sum_{i=1}^m \alpha_i y_i \mathbf{x}_i \right\|^2 - \sum_{i=1}^m \alpha_i y_i \left(\sum_{j=1}^m \alpha_j y_j \mathbf{x}_j^\top \mathbf{x}_i \right) + \sum_{i=1}^m \alpha_i + \sum_{i=1}^m (C - \alpha_i - \beta_i) \xi_i \\ &= \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j (\mathbf{x}_i^\top \mathbf{x}_j) - \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j (\mathbf{x}_i^\top \mathbf{x}_j) + \sum_{i=1}^m \alpha_i + \sum_{i=1}^m (C - \alpha_i - \beta_i) \xi_i \\ &= \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j (\mathbf{x}_i^\top \mathbf{x}_j). \end{aligned}$$

由偏导条件可知 $0 \leq \alpha_i \leq C$ ，并且

$$\sum_{i=1}^m \alpha_i y_i = 0.$$

因此得到软间隔 SVM 的对偶问题：

软间隔 SVM 对偶问题：

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j (\mathbf{x}_i^\top \mathbf{x}_j), \\ \text{s.t.} \quad & 0 \leq \alpha_i \leq C, \quad i = 1, \dots, m, \\ & \sum_{i=1}^m \alpha_i y_i = 0. \end{aligned}$$

其中，约束条件 $0 \leq \alpha_i \leq C$, $i = 1, \dots, m$ 由约束式 $C - \alpha_i - \beta_i = 0$ 与 $\beta_i \geq 0$ 推得。

由 SMO 算法可求解该二次规划问题，得到最优拉格朗日乘子 $\alpha^* = (\alpha_1^*, \alpha_2^*, \dots, \alpha_m^*)$ 。根据 KKT 条件，有：

$$\begin{cases} \alpha_i^* (y_i (\mathbf{w}^{*\top} \mathbf{x}_i + b^*) - 1 + \xi_i^*) = 0, \\ \beta_i^* \xi_i^* = 0, \\ y_i (\mathbf{w}^{*\top} \mathbf{x}_i + b^*) \geq 1 - \xi_i^*, \\ \xi_i^* \geq 0, \quad \alpha_i^*, \beta_i^* \geq 0. \end{cases}$$

为求偏置项 b^* ，通常选择满足 $0 < \alpha_k^* < C$ 的样本点 $\mathbf{x}_k$ ，即所谓的支持向量。对于该样本，由 KKT 条件第一式知， $\alpha_k^* > 0$ ，因此必须满足：

$$y_k(\mathbf{w}^{*\top} \mathbf{x}_k + b^*) - 1 + \xi_k^* = 0,$$

$$\Rightarrow y_k(\mathbf{w}^{*\top} \mathbf{x}_k + b^*) = 1 - \xi_k^*.$$

又因 $\xi_k^* \geq 0$ ，可得偏置项为：

$$b^* = y_k - \mathbf{w}^{*\top} \mathbf{x}_k - \xi_k^*,$$

其中，

$$\mathbf{w}^* = \sum_{i=1}^m \alpha_i^* y_i \mathbf{x}_i, \quad \xi_k^* = \max(0, 1 - y_k(\mathbf{w}^{*\top} \mathbf{x}_k + b^*)).$$

4 线性不可分情形与核函数

针对软间隔 SVM 的对偶问题：

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j (\mathbf{x}_i^\top \mathbf{x}_j), \\ \text{s.t.} \quad & 0 \leq \alpha_i \leq C, \quad i = 1, \dots, m, \\ & \sum_{i=1}^m \alpha_i y_i = 0. \end{aligned}$$

注意到，数据点 $\mathbf{x}_i$ 仅以内积形式出现在目标函数中。因此，可以通过非线性映射 $\phi: \mathbb{R}^n \rightarrow \mathbb{R}^N$ ，将原始数据映射到高维特征空间：

$$\mathbf{x}_i \mapsto \phi(\mathbf{x}_i).$$

在高维空间中，数据可能变得线性可分，从而可以应用线性 SVM 方法进行分类。映射后的对偶问题为：

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j (\phi(\mathbf{x}_i)^\top \phi(\mathbf{x}_j)), \\ \text{s.t.} \quad & 0 \leq \alpha_i \leq C, \quad i = 1, \dots, m, \\ & \sum_{i=1}^m \alpha_i y_i = 0. \end{aligned}$$

然而，显式计算高维映射 $\phi(\mathbf{x})$ 可能计算量巨大，甚至不可行。为此，引入核函数 $K(\mathbf{x}_i, \mathbf{x}_j)$ ，定义为：

$$K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^\top \phi(\mathbf{x}_j).$$

核函数允许直接计算映射后数据点的内积，而无需显式进行映射。常用的核函数包括：

- 线性核: $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^\top \mathbf{x}_j$;
- 多项式核: $K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i^\top \mathbf{x}_j + r)^d$, 其中 $\gamma > 0$, $r \geq 0$, $d \in \mathbb{N}$;
- 径向基函数 (RBF) 核 (高斯核): $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$, 其中 $\gamma > 0$ 。
- Sigmoid 核: $K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\kappa \mathbf{x}_i^\top \mathbf{x}_j + c)$, 其中 $\kappa > 0$, $c < 0$ 。

为什么可以升维使用核函数?

根据 Mercer 定理, 任意对称正定函数都可以作为核函数, 任意核函数可以写成内积形式。

数学上可证明, 维数升高后, 数据变得线性可分的概率更大。