机器学习纸质课程作业

一、 证明题

(1) 证明$\mathbf{f(\omega)=\frac{1}{2}\|\omega\|^2}$是凸函数，即对$\forall\omega_1,\omega_2,\forall\lambda\in[0,1]$，都有

$$\mathbf{f(\lambda\omega_1+(1-\lambda)\omega_2)\le\lambda f(\omega_1)+(1-\lambda)f(\omega_2)}$$

证明：

目标不等式:

我们需要证明

$$\frac{1}{2}\|\lambda\omega_1+(1-\lambda)\omega_2\|^2\le\lambda\cdot\frac{1}{2}\|\omega_1\|^2+(1-\lambda)\cdot\frac{1}{2}\|\omega_2\|^2$$

两边同时乘以 2，等价于

$$\|\lambda\omega_1+(1-\lambda)\omega_2\|^2\le\lambda\|\omega_1\|^2+(1-\lambda)\|\omega_2\|^2$$

展开左边项:

令 $\omega=\lambda\omega_1+(1-\lambda)\omega_2$，则

$$\|\omega\|^2=\langle\omega,\omega\rangle=\langle\lambda\omega_1+(1-\lambda)\omega_2,\lambda\omega_1+(1-\lambda)\omega_2\rangle$$

其中 $\langle\cdot,\cdot\rangle$ 表示点积。展开点积:

$$\begin{aligned}\|\omega\|^2&=\lambda^2\langle\omega_1,\omega_1\rangle+2\lambda(1-\lambda)\langle\omega_1,\omega_2\rangle+(1-\lambda)^2\langle\omega_2,\omega_2\rangle\\&=\lambda^2\|\omega_1\|^2+2\lambda(1-\lambda)\langle\omega_1,\omega_2\rangle+(1-\lambda)^2\|\omega_2\|^2\end{aligned}$$

计算差值:

$$\lambda\|\omega_1\|^2+(1-\lambda)\|\omega_2\|^2-\|\lambda\omega_1+(1-\lambda)\omega_2\|^2$$

代入展开式:

$$=\lambda\|\omega_1\|^2+(1-\lambda)\|\omega_2\|^2-[\lambda^2\|\omega_1\|^2+2\lambda(1-\lambda)\langle\omega_1,\omega_2\rangle+(1-\lambda)^2\|\omega_2\|^2]$$

简化:

$$=(\lambda-\lambda^2)\|\omega_1\|^2+[(1-\lambda)-(1-\lambda)^2]\|\omega_2\|^2-2\lambda(1-\lambda)\langle\omega_1,\omega_2\rangle$$

注意 $\lambda-\lambda^2=\lambda(1-\lambda)$ 和 $(1-\lambda)-(1-\lambda)^2=(1-\lambda)[1-(1-\lambda)]=(1-\lambda)\lambda=\lambda(1-\lambda)$，所以:

$$\begin{aligned}&=\lambda(1-\lambda)\|\omega_1\|^2+\lambda(1-\lambda)\|\omega_2\|^2-2\lambda(1-\lambda)\langle\omega_1,\omega_2\rangle\\&=\lambda(1-\lambda)[\|\omega_1\|^2+\|\omega_2\|^2-2\langle\omega_1,\omega_2\rangle]\end{aligned}$$

括号中的项为 $\|\omega_1-\omega_2\|^2$:

$$=\lambda(1-\lambda)\|\omega_1-\omega_2\|^2$$

分析差值:

由于 $\lambda\in[0,1]$，有 $\lambda(1-\lambda)\ge0$，且 $\|\omega_1-\omega_2\|^2\ge0$（范数平方非负），因此

$$\lambda(1-\lambda)\|\omega_1-\omega_2\|^2\ge0$$

这意味着

$$\lambda \| \omega_1 \|^2 + (1-\lambda) \| \omega_2 \|^2 - \| \lambda\omega_1 + (1-\lambda)\omega_2 \|^2 \geq 0$$

即

$$\| \lambda\omega_1 + (1-\lambda)\omega_2 \|^2 \leq \lambda \| \omega_1 \|^2 + (1-\lambda) \| \omega_2 \|^2$$

回到函数 f:

将上述不等式乘以 $\frac{1}{2}$，得到

$$\frac{1}{2} \| \lambda\omega_1 + (1-\lambda)\omega_2 \|^2 \leq \frac{1}{2} [\lambda \| \omega_1 \|^2 + (1-\lambda) \| \omega_2 \|^2] = \lambda \cdot \frac{1}{2} \| \omega_1 \|^2 + (1-\lambda) \cdot \frac{1}{2} \| \omega_2 \|^2$$

也就是

$$f(\lambda\omega_1 + (1-\lambda)\omega_2) \leq \lambda f(\omega_1) + (1-\lambda) f(\omega_2)$$

结论：

因此，函数 $f(\omega) = \frac{1}{2} \| \omega \|^2$ 满足凸函数的定义，是凸函数。

（2） 证明两个凸函数的和仍然是凸函数，因此支持向量机的目标函数

$$\frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^{N} \delta_i$$

是凸函数。

证明一：两个凸函数之和是凸的

设 f与 g在同一凸集 D上凸。取任意 $x, y \in D$与任意 $\lambda \in [0,1]$。由凸性分别有

$$f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y),$$
$$g(\lambda x + (1-\lambda) y) \leq \lambda g(x) + (1-\lambda) g(y).$$

对两式逐项相加得到

$$f(\lambda x + (1-\lambda) y) + g(\lambda x + (1-\lambda) y) \leq \lambda (f(x) + g(x)) + (1-\lambda)(f(y) + g(y)).$$

即

$$h(\lambda x + (1-\lambda) y) \leq \lambda h(x) + (1-\lambda) h(y),$$

因此 $h = f + g$在 D上凸。

证明二：正标量乘保凸性

若 f凸且 $C \geq 0$，则对所有 $x, y \in D, \lambda \in [0,1]$:

$$Cf(\lambda x + (1-\lambda)y) \le C(\lambda f(x) + (1-\lambda)f(y)) = \lambda(Cf(x)) + (1-\lambda)(Cf(y)),$$

所以 Cf也凸（若 $C < 0$则会反向，不再保持凸性）。

综上

$$\frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{N}\delta_i$$

是凸函数。

(3)　请证明：$\boldsymbol{\theta(\alpha,\beta)}$是一个凹函数,即对任意$\boldsymbol{(\alpha_1,\beta_1,\alpha_2,\beta_2)}$，$\boldsymbol{\lambda \in [0,1]}$，有：

$$\boldsymbol{\theta(\lambda\alpha_1 + (1-\lambda)\alpha_2, \lambda\beta_1 + (1-\lambda)\beta_2) \ge \lambda\theta(\alpha_1,\beta_1) + (1-\lambda)\theta(\alpha_2,\beta_2)}$$

（提示，请先证明：$\boldsymbol{\inf_{\omega\in\Omega}[a(\omega)+b(\omega)] \ \ge \inf_{\omega\in\Omega}[a(\omega)] + \inf_{\omega\in\Omega}[b(\omega)]}$）

应用该定理可以直接得到支持向量机对偶问题的目标函数

$$\boldsymbol{\sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j K(x_i,x_j)}$$

是凹函数，因此最大化此目标函数有唯一解。

1) 先证明提示的不等式

令 $a(\omega)$与 $b(\omega)$为定义在同一集合 Ω上的任意实值函数。我们要证

$$\inf_{\omega\in\Omega}[a(\omega)+b(\omega)] \ \ge \ \inf_{\omega\in\Omega} a(\omega) \ + \ \inf_{\omega\in\Omega} b(\omega).$$

证明：对任意固定的 $\omega \in \Omega$，都有
$a(\omega) \ge \inf_{\tilde{\omega}} a(\tilde{\omega})$且 $b(\omega) \ge \inf_{\tilde{\omega}} b(\tilde{\omega})$。因此

$$a(\omega) + b(\omega) \ge \inf_{\tilde{\omega}} a(\tilde{\omega}) + \inf_{\tilde{\omega}} b(\tilde{\omega}).$$

对所有 ω求下确界时，左边的下确界不会小于右边的常数，即

$$\inf_{\omega}[a(\omega)+b(\omega)] \ge \inf_{\tilde{\omega}} a(\tilde{\omega}) + \inf_{\tilde{\omega}} b(\tilde{\omega}).$$

证毕。

2) 证明 $\boldsymbol{\theta(\alpha,\beta) = \inf_{\omega\in\Omega} L(\omega,\alpha,\beta)}$是凹函数

回忆拉格朗日函数的形式（对任意固定 ω）：

$$L(\omega,\alpha,\beta) = f(\omega) + \alpha^{\mathsf{T}} g(\omega) + \beta^{\mathsf{T}} h(\omega).$$

对于固定的 ω，$L(\omega,\alpha,\beta)$ 是关于 (α,β)的**仿射函数**（即线性加常数）。仿射函数既是凸的也是凹的；对任意两组乘子 (α_1,β_1)与 (α_2,β_2)和任意 $\lambda\in[0,1]$，

$$L(\omega,\lambda(\alpha_1,\beta_1)+(1-\lambda)(\alpha_2,\beta_2))=\lambda L(\omega,\alpha_1,\beta_1)+(1-\lambda)L(\omega,\alpha_2,\beta_2).$$

因此

$$\begin{aligned}\theta(\lambda(\alpha_1,\beta_1)+(1-\lambda)(\alpha_2,\beta_2)) &= \inf_{\omega\in\Omega} L(\omega,\lambda(\alpha_1,\beta_1)+(1-\lambda)(\alpha_2,\beta_2)) \\ &= \inf_{\omega\in\Omega}[\lambda L(\omega,\alpha_1,\beta_1)+(1-\lambda)L(\omega,\alpha_2,\beta_2)].\end{aligned}$$

用第 1) 节的推广结论（对带系数的线性组合取 inf 的不等式），我们得到

$$\inf_{\omega}[\lambda L(\omega,\alpha_1,\beta_1)+(1-\lambda)L(\omega,\alpha_2,\beta_2)]\geq\lambda\inf_{\omega} L(\omega,\alpha_1,\beta_1)+(1-\lambda)\inf_{\omega} L(\omega,\alpha_2,\beta_2).$$

也就是

$$\theta(\lambda(\alpha_1,\beta_1)+(1-\lambda)(\alpha_2,\beta_2))\geq\lambda\,\theta(\alpha_1,\beta_1)+(1-\lambda)\,\theta(\alpha_2,\beta_2).$$

这正是 θ关于 (α,β)的**凹性**定义。证毕。

3) 应用到支持向量机对偶问题

在支持向量机的对偶问题中，目标函数为：

$$\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j K(x_i,x_j)$$

这个函数可以看作是一个特殊的 $\theta(\alpha)$函数（没有 β 参数），其中：

- 线性项$\sum_{i=1}^{N}\alpha_i$对应 $f(\omega)$部分
- 二次型$-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j K(x_i,x_j)$对应约束部分

由于：

（1） 线性函数$\sum_{i=1}^{N}\alpha_i$是凸的也是凹的；

（2） 二次型$-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j K(x_i,x_j)$是凹函数（因为核矩阵$K(x_i,x_j)$是半正定的，其负二次型是凹的）；

（3） 凹函数的和仍然是凹函数。

因此，整个目标函数是凹函数。

由于目标函数是凹函数且约束条件是线性的，这个最大化问题是一个凸优化问题，有唯一全局最优解。

结论： 我们证明了 θ(α,β)是凹函数，并应用这个结论说明了支持向量机对偶问题的目标函数是凹函数，因此最大化此目标函数有唯一解。

(4) 选做题

(a) 在二维特征空间中不存在能被线性分类器"打散"的 4 个点

证明:

1. 我们先排除退化情形（有三点共线等）——若有三点共线，更容易找出不可分的标记（例如把共线的一端标为 +，另一端标为 −，中间点同一侧等），因此只需论证一般位置（无三点共线）的情况即可；随机取点时一般位置的情况以概率 1 发生。

2. 对无三点共线的四点，凸包要么是一个四边形（四点位于凸位置，即四个顶点按环形排列），要么是一个三角形（其中一个点落在其余三点构成的三角形内部）。

- 若凸包是三角形（即有一个点在其它三个点的凸包内部），记内部点为 p，其余三个为 a, b, c。把 p标为 +1，而把 a, b, c都标为 −1（三者同号），任何将 p与 a, b, c分开的半空间（由直线给出）都必须把包含 a, b, c的半空间与包含 p的半空间分开。但因为 p在 a, b, c的凸包内部，任何包含 a, b, c的凸半空间的凸包必然也包含 p 因而不可能有单一的半空间把 p与 a, b, c分开。由此该标记不可分。

- 若凸包是凸四边形（四点按环形排列无内点），将四点按环形编号为 v_1, v_2, v_3, v_4。考虑"交替标记"：给 v_1, v_3标为 +1，给 v_2, v_4标为 −1（或者相反）。任何半空间与凸四边形交集的顶点集合在环形上必是**连通的区间**（半空间切割多边形后留下的顶点是相邻的一段），因此不可能得到交替出现的模式（+、−、+、−）。更形式地：对任意直线，四边形顶点被分到直线两侧时，同一侧的顶点在原多边形循环序列上形成一个连块；交替标记要求同号顶点不成连块，故无法用一条直线实现。 因此交替标记不可分。

两个情况已覆盖所有无退化的四点布置，故对任意 4 点总能找到不可分的标签组合，因此平面线性分类器不能"打散"任意 4 点。证毕。

（因此平面线性分类器的 VC 维为 3——存在某些 3 点可被打散，且不存在任意 4 点被打散。）

(b) 在线性分类器（超平面 $\boldsymbol{\omega}^\top \mathbf{x} + \mathbf{b} = \mathbf{0}$）下，$\mathbb{R}^{\mathbf{d}}$ 的 VC 维为 $\mathbf{d} + \mathbf{1}$

要证两件事:
(i) 存在 $d + 1$个点能被线性分类器打散;
(ii) 任意 $d + 2$个点不能被线性分类器打散（即 VC ≤ $d + 1$）。可得 VC = $d + 1$。

(i) 构造能被打散的 $\mathbf{d} + \mathbf{1}$个点

取任意 $d + 1$个**仿射无关**的点 $x_1, \dots, x_{d+1} \in \mathbb{R}^d$。仿射无关意味着这些点的增广矩阵

$$\begin{pmatrix} x_1^\top & 1 \\ x_2^\top & 1 \\ \vdots & \vdots \\ x_{d+1}^\top & 1 \end{pmatrix}$$

是一个 $(d+1)\times(d+1)$的可逆矩阵。现在给任意二分类标签 $y_i\in\{+1,-1\}$（共 2^{d+1}种），我们要找 $(\omega,b)\in\mathbb{R}^d\times\mathbb{R}$使得

$$\operatorname{sign}(\omega^\top x_i+b)=y_i, i=1,\dots,d+1.$$

更强地，我们可以要求确切地拟合实数值（不仅仅是符号），例如解线性方程组

$$\omega^\top x_i+b=y_i, i=1,\dots,d+1.$$

这是 $d+1$个线性方程、$d+1$ 个未知数（ω 的 d个分量和标量 b），系数矩阵就是上面的增广矩阵的转置，因仿射无关而可逆，因此方程组有唯一解。求得的 (ω,b)使得 $\omega^\top x_i+b=y_i$（严格等于 ±1），于是符号完全匹配任意给定的 y_i。因此任意标签都能被正确分类，说明这 $d+1$个点被打散。故 VC ⩾ $d+1$。

(ii) 任意 $d+2$个点不能被打散（VC ≤ $d+1$）

任意 $d+2$个点在 $\mathbb{R}^d$中必有仿射依赖（因为维数为 d，任何 $d+2$个点线性/仿射相关）。由 **Radon 定理**：任意 $d+2$个点可以被划分为两组 A与 B（互不相交），使得这两组点的凸包相交，即 $\operatorname{conv}(A)\cap\operatorname{conv}(B)\neq\emptyset$。

现在取这种 Radon 划分，并把 A的点全部标为 $+1$，把 B的点全部标为 -1。若存在一个超平面把 $+1$ 与 -1 完全分开，则正类（$+1$ 点）的半空间是一个凸集，且包含 A的所有点；因此它也应包含 $\operatorname{conv}(A)$。类似，负类的半空间包含 $\operatorname{conv}(B)$。但因为这两凸包有交点，任何将两个凸包严格分开的超平面都不存在（两个凸包相交表示不能通过超平面把它们分离成不同的开半空间）。因此该标记不可被超平面分离。由此任意 $d+2$个点总存在一种标记不可分，故不能被打散，得 VC ⩽ $d+1$。

结合 (i)、(ii) 得 VC = $d+1$。证毕。

(c) 由 (b) 推出 $\lim\limits_{d\to\infty} P(d,N)=1$

明确题意：

- 在固定的样本数 N下，随机从 $\mathbb{R}^d$中取 N个点（例如从连续分布独立同分布抽取），并对这 N个点随机标记 $+1/-1$（独立均匀），问：随着维度 d增大，点集被线性分类器线性可分（即存在一个超平面把正负两类完全分开）的概率 $P(d,N)$的极限。

证明要点：

- 若 $N\le d+1$，并且点为一般位置（概率 1 情形：从连续分布随机取点，任意 $k\le d+1$个点几乎肯定仿射无关），则这些点仿射无关，从 (b) 我们知道对任意标签都存在一个超平面精确拟合（解线性方程组），因而任意标签组合均线性可分 → **对于这种点的具体位置，任意标签下都可分**。因此当 $d\ge N-1$时，随机取的 N个点几乎必是仿射无关，从而**对于任意标签**都线性可分，故此时 $P(d,N)=1$（严格来说为 1 在概率意义上：除去测度为 0 的退化点集）。

- 因此当 $d\to\infty$（固定 N），存在某个阈值 $d_0=N-1$使得对所有 $d\ge d_0$有 $P(d,N)=1$。所以

$$\lim_{d\to\infty} P(d,N)=1.$$

更直观地：当维度远大于样本数时，样本点几乎总是处于仿射无关的位置（没有线性/仿射关系），从而在 d维空间中总能找到一个超平面来实现任意的分配；因此随机标记几乎总被线性分隔。

二、 请写出以下问题的对偶问题，并化成最简形式

（1） $\mathbf{l_2}$-norm SVM Classification

原问题：

$$\min_{\omega,b,\delta} \frac{1}{2}\parallel \omega \parallel^2 + C\sum_{i=1}^{N}\delta_i^2$$

$$\text{s.t.}\quad y_i(\omega^T\phi(x_i)+b)\ge 1-\delta_i,$$

$$\delta_i \ge 0$$

对偶问题：

引入拉格朗日乘子 $\alpha_i \ge 0$对应于约束 $y_i(\omega^T\phi(x_i)+b)\ge 1-\delta_i$。

拉格朗日函数：

$$L=\frac{1}{2}\parallel \omega \parallel^2 + C\sum_i \delta_i^2 - \sum_i \alpha_i\left[y_i(\omega^T\phi(x_i)+b)-1+\delta_i\right]$$

消去原变量 ω, b, δ_i：

$$\omega=\sum_i \alpha_i y_i \phi(x_i), \sum_i \alpha_i y_i = 0, \delta_i=\frac{\alpha_i}{2C}$$

代回得：

$$\max_{\alpha} \sum_{i=1}^{N}\alpha_i - \frac{1}{2}\sum_{i,j}\alpha_i\alpha_j y_i y_j K(x_i,x_j) - \frac{1}{4C}\sum_{i=1}^{N}\alpha_i^2$$

$$\text{s.t.}\quad \sum_{i=1}^{N}\alpha_i y_i = 0, \alpha_i \ge 0$$

这是 l_2-SVM 的最简对偶形式。

（2） SVM Regression（选做）

原问题：

$$\min_{\omega,b,\delta,\zeta} \frac{1}{2}\parallel \omega \parallel^2 + C\sum_i(\delta_i+\zeta_i)$$

$$\text{s.t.}\quad \omega^T\phi(x_i)+b-y_i \le \varepsilon+\delta_i,$$

$$y_i-\omega^T\phi(x_i)-b \le \varepsilon+\zeta_i,$$

$$\delta_i, \zeta_i \ge 0$$

对偶问题

最大化（Maximize）：

$$\theta(\alpha,\beta,u,v)=\sum_{i=1}^{N}(\beta_i-\alpha_i)y_i-\left[\sum_{i=1}^{N}(\alpha_i+\beta_i)\right]\varepsilon-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\beta_i-\alpha_i)\left(\beta_j-\alpha_j\right)K(X_i,X_j)$$

限制条件(Subject to): ① $0\le\alpha_i\le C$

② $0\le\beta_i\le C$

③ $\sum_{i=1}^{N}\alpha_i=\sum_{i=1}^{N}\beta_i$

(3) Kernel Ridge Regression

原问题:

$$\min_{\omega,\delta}\ \frac{1}{2}\parallel\omega\parallel^2+C\sum_i\delta_i^2$$
$$\text{s.t.}\quad y_i-\omega^T\phi(x_i)=\delta_i$$

对偶问题:

拉格朗日函数，令α_i为约束乘子:

$$L=\frac{1}{2}\parallel\omega\parallel^2+C\sum_i\delta_i^2+\sum_i\alpha_i(y_i-\omega^T\phi(x_i)-\delta_i)$$

求偏导:

$$\omega=\sum_i\alpha_i\phi(x_i),\ \delta_i=\frac{\alpha_i}{2C}$$

代回:

$$\max_{\alpha}\ \sum_i\alpha_iy_i-\frac{1}{2}\sum_{i,j}\alpha_i\alpha_jK(x_i,x_j)-\frac{1}{4C}\sum_i\alpha_i^2$$

无约束对偶形式（线性方程形式）:

$$(K+\frac{1}{2C}I)\alpha=y$$

从而 $\omega=\sum_i\alpha_i\phi(x_i)$。

(4) Entropy Maximization Problem

原问题:

$$\min_{x}\ \sum_i x_i\log\ x_i$$
$$\text{s.t.}\quad \omega^Tx\le b,$$
$$\sum_i x_i=1$$

对偶问题:

拉格朗日函数:

$$L(x,\lambda,\nu)=\sum_i x_i\log\ x_i+\lambda(\omega^Tx-b)+\nu(\sum_i x_i-1)$$

对x_i求导并设为 0:

$$1+\log\ x_i+\lambda\omega_i+\nu=0\ \Longrightarrow\ x_i=e^{-1-\nu-\lambda\omega_i}$$

代回原约束 $\sum_i x_i=1$:

$$e^{-1-\nu}\sum_i e^{-\lambda\omega_i}=1\ \Longrightarrow\ e^{-1-\nu}=\frac{1}{\sum_i e^{-\lambda\omega_i}}$$

代回目标函数得对偶形式:

$$\max_{\lambda\ge 0}-\lambda b-\log\ \left(\sum_i e^{-\lambda\omega_i}\right)$$

三、 包含点集的最小球问题

对偶推导

对第 i个不等式引入拉格朗日乘子 $\alpha_i\ge 0$，对 $\xi_i\ge 0$引入 $\mu_i\ge 0$。拉格朗日函数:

$$L(R,C,\xi,\alpha,\mu)=R^2+B\sum_i\xi_i+\sum_i\alpha_i(\|x_i-C\|^2-R^2-\xi_i)-\sum_i\mu_i\xi_i\,.$$

驻点条件（对 R,C,ξ_i求偏导并令零）给出

$$\begin{aligned}\partial_R&:\ 2R(1-\sum_i\alpha_i)=0\ \Rightarrow\ \sum_i\alpha_i=1,\\ \partial_C&:\ -2\sum_i\alpha_i(x_i-C)=0\ \Rightarrow\ C=\sum_i\alpha_i x_i\,,\\ \partial_{\xi_i}&:\ B-\alpha_i-\mu_i=0\ \Rightarrow\ 0\le\alpha_i\le B.\end{aligned}$$

把 $C=\sum_j\alpha_j x_j$代回 L 并消去其他项（R 项为零），得到对偶目标:

$$\begin{aligned}g(\alpha)&=\sum_{i=1}^N\alpha_i\|x_i\|^2-\|\sum_{i=1}^N\alpha_i x_i\|^2\\ &=\sum_{i=1}^N\alpha_i\|x_i\|^2-\sum_{i=1}^N\sum_{j=1}^N\alpha_i\alpha_j\,x_i^\top x_j\,.\end{aligned}$$

（1） 对偶问题（最简形式）

$$\boxed{\begin{aligned}&\max_{\alpha\in\mathbb{R}^N}\sum_{i=1}^N\alpha_i\|x_i\|^2-\sum_{i=1}^N\sum_{j=1}^N\alpha_i\alpha_j\,x_i^\top x_j\\ &\text{s.t.}\quad\sum_{i=1}^N\alpha_i=1,0\le\alpha_i\le B\ (i=1,\dots,N).\end{aligned}}$$

（这是一个带线性约束与简单界的二次凸规划问题。）

(2)　　核化（取 $K(x_i,x_j)=x_i^\top x_j$ 或一般核）

将内积替换为核函数 K，对偶写成核矩阵形式：

$$\boxed{\begin{aligned}\max_{\alpha}\quad & \sum_{i=1}^{N}\alpha_i K(x_i,x_i)-\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\,\alpha_j K(x_i,x_j)\\ \text{s.t.}\quad & \sum_{i=1}^{N}\alpha_i=1, 0\le\alpha_i\le B.\end{aligned}}$$

用向量 / 矩阵记号（令 K 为核矩阵，$d=[K_{11},\dots,K_{NN}]^\top$，$\alpha$ 为系数向量），目标可写为

$$\max_{\alpha}\alpha^\top d-\alpha^\top K\alpha\quad \text{s.t.}\ \mathbf{1}^\top\alpha=1,\ 0\le\alpha\le B\mathbf{1}.$$

(3)　　用核表示 R

若索引 s 满足 $0<\alpha_s<B$（即对应约束为等号的“支撑点”），则由 KKT 有

$\| x_s-C\|^2=R^2$。用 $C=\sum_j\alpha_j x_j$ 展开，得

$$\begin{aligned}R^2&=\| x_s\|^2-2\sum_{j=1}^{N}\alpha_j\,x_s^\top x_j+\sum_{j=1}^{N}\sum_{k=1}^{N}\alpha_j\alpha_k\,x_j^\top x_k\\ &=K(x_s,x_s)-2\sum_{j=1}^{N}\alpha_j K(x_s,x_j)+\sum_{j=1}^{N}\sum_{k=1}^{N}\alpha_j\alpha_k K(x_j,x_k).\end{aligned}$$

因此 R 的核形式为

$$\boxed{R=\sqrt{K(x_s,x_s)-2\sum_{j}\alpha_j K(x_s,x_j)+\sum_{j,k}\alpha_j\alpha_k K(x_j,x_k)}}$$

（任取一个 s 使 $0<\alpha_s<B$；若有多个支撑点，任一支撑点代入均应给出相同的 R^2）。