CS189 作业：硬间隔支持向量机（理论题）中文题面与证明

背景与记号

线性 SVM 的判别函数（原文式 (1)）为

$$r(x) = \begin{cases} +1, & w^\top x + \alpha \geq 0, \\ -1, & \text{otherwise.} \end{cases}$$

原始（primal）硬间隔 SVM（原文式 (2)）：

$$\min_{w,\alpha} \ \|w\|^2 \quad \text{s.t.} \quad y_i\big(X_i^\top w + \alpha\big) \geq 1, \ i = 1, \ldots, n.$$

其拉格朗日鞍点形式（原文式 (3)）：

$$\max_{\lambda_i \geq 0} \ \min_{w,\alpha} \ \|w\|^2 - \sum_{i=1}^{n} \lambda_i \big(y_i(X_i^\top w + \alpha) - 1\big).$$

问题 1. （*a*）把式 *(3)* 化为对偶优化（原文式 *(4)*），并解释新等式约束的来源。

证明. 设

$$\mathcal{L}(w, \alpha, \lambda) = \|w\|^2 - \sum_{i=1}^{n} \lambda_i \big(y_i(X_i^\top w + \alpha) - 1\big), \qquad \lambda_i \geq 0.$$

对内层最小化的一阶条件（站立性）为

$$\frac{\partial \mathcal{L}}{\partial w} = 2w - \sum_{i=1}^{n} \lambda_i y_i X_i = 0 \ \Rightarrow \ w^* = \frac{1}{2} \sum_{i=1}^{n} \lambda_i y_i X_i,$$
$$\frac{\partial \mathcal{L}}{\partial \alpha} = -\sum_{i=1}^{n} \lambda_i y_i = 0 \ \Rightarrow \ \sum_{i=1}^{n} \lambda_i y_i = 0.$$

将 w^* 与等式约束代回 $\mathcal{L}$，注意

$$\|w^*\|^2 = \Big(\tfrac{1}{2} \sum_i \lambda_i y_i X_i\Big)^\top \Big(\tfrac{1}{2} \sum_j \lambda_j y_j X_j\Big) = \frac{1}{4} \sum_{i,j} \lambda_i \lambda_j y_i y_j \, X_i^\top X_j,$$

并且 $-\alpha\sum_i \lambda_i y_i = 0$，故对偶函数为

$$g(\lambda) = \sum_{i=1}^n \lambda_i - \frac{1}{4}\sum_{i=1}^n\sum_{j=1}^n \lambda_i\lambda_j y_i y_j\, X_i^\top X_j.$$

弱对偶性保证 $g(\lambda)$ 是原始问题最优值的下界，故对偶问题为

$$\max_{\lambda\geq 0}\ g(\lambda) \quad s.t. \quad \sum_{i=1}^n \lambda_i y_i = 0.$$

新等式约束正是由 $\partial\mathcal{L}/\partial\alpha = 0$ 得到。证毕。

问题 2. （b）已知能最优解式 *(3)* 的 λ_i^* 与 α^*。证明由式 *(1)* 给出的判别函数可写为

$$r(x) = \operatorname{sign}\left(\alpha^* + \frac{1}{2}\sum_{i=1}^n \lambda_i^* y_i\, X_i^\top x\right).$$

证明. 由 *(a)* 的平稳性条件，$w^* = \frac{1}{2}\sum_i \lambda_i^* y_i X_i$。代回

$$r(x) = \operatorname{sign}(w^{*\top}x + \alpha^*) = \operatorname{sign}\left(\alpha^* + \tfrac{1}{2}\sum_{i=1}^n \lambda_i^* y_i\, X_i^\top x\right).$$

证毕。

问题 3. （c）利用互补松弛条件 $\lambda_i^*\big(y_i(X_i^\top w^* + \alpha^*) - 1\big) = 0$，解释当 $\lambda_i^* > 0$ 时样本点 X_i 的几何意义。

证明. 若 $\lambda_i^* > 0$，则必有 $y_i(X_i^\top w^* + \alpha^*) = 1$，即该点恰落在两条间隔超平面之一 $\{x:\ w^{*\top}x + \alpha^* = \pm 1\}$ 上，因而是支持向量。证毕。

问题 4. （d）说明在计算 $r(x)$ 时，只需使用 $\lambda_i^* > 0$ 的样本点。

证明. 由 *(b)*

$$w^{*\top}x + \alpha^* = \alpha^* + \frac{1}{2}\sum_{i=1}^n \lambda_i^* y_i\, X_i^\top x.$$

若某 i 使 $\lambda_i^* = 0$，其对和式无贡献；只有 $\lambda_i^* > 0$（支持向量）才影响判别值，故仅需支持向量即可计算 $r(x)$。证毕。

问题 5. （e）给定近似参数 $w = [-0.4528, -0.5190]^\top$，$\alpha = 0.1471$。写出二维平面上决策边界 $w^\top x + \alpha = 0$ 及两条边距 $w^\top x + \alpha = \pm 1$ 的解析式，并说明支持向量所在位置。

证明. 令 $x = (x_1, x_2)^\top$。决策边界为

$$x_2 = -\frac{w_1}{w_2}x_1 - \frac{\alpha}{w_2} \approx -0.8723\, x_1 - 0.2834,$$

两条边距为

$$x_2 = -\frac{w_1}{w_2}x_1 - \frac{\alpha \pm 1}{w_2}.$$

支持向量满足 $w^\top X_i + \alpha = \pm 1$，即恰位于上述两条平行直线上。证毕。

问题 6.（f）在线性可分情形下，用原始形式证明：最优解的每个类别（$+1$ 与 -1）至少各有一个支持向量。

证明. 反证法。设最优解 (w^*, α^*) 下，正类无支持向量。则存在 $\varepsilon > 0$，对所有 $y_i = +1$ 有 $w^{*\top} X_i + \alpha^* \geq 1 + \varepsilon$。取

$$\tilde{\alpha} = \alpha^* - \varepsilon/2.$$

则正类满足 $w^{*\top} X_i + \tilde{\alpha} \geq 1 + \varepsilon/2 > 1$。对负类 $y_i = -1$，有

$$y_i(w^{*\top} X_i + \tilde{\alpha}) = -(w^{*\top} X_i + \alpha^*) + \varepsilon/2 \geq 1 + \varepsilon/2 > 1.$$

于是存在 $\delta = \varepsilon/2 > 0$ 使所有样本均满足 $y_i(w^{*\top} X_i + \tilde{\alpha}) \geq 1 + \delta$。对 $(w^*, \tilde{\alpha})$ 同比缩放

$$w' = \frac{w^*}{1 + \delta/2}, \qquad \alpha' = \frac{\tilde{\alpha}}{1 + \delta/2},$$

则 $y_i(w'^\top X_i + \alpha') \geq \frac{1+\delta}{1+\delta/2} > 1$，故 (w', α') 仍可行，但目标 $\|w'\|^2 = \|w^*\|^2/(1+\delta/2)^2 < \|w^*\|^2$，与最优性矛盾。故正类必须存在支持向量；对负类同理，结论成立。证毕。