

Lecture 14. Support Vector Machine II

Lecturer: Jie Wang

Date: Dec 1, 2021

1 The Primal Problem

Recall from the last lecture that, we are interested in the problems that take the form of

$$\begin{aligned} \min_{\mathbf{x}} f(\mathbf{x}) \\ \text{s.t. } \mathbf{g}(\mathbf{x}) \leq 0, \\ \mathbf{h}(\mathbf{x}) = 0, \\ \mathbf{x} \in X. \end{aligned} \quad (1)$$

We denote the *feasible set* of (1) by

$$D_0 = \{\mathbf{x} : \mathbf{g}(\mathbf{x}) \leq 0, \mathbf{h}(\mathbf{x}) = 0, \mathbf{x} \in X\}. \quad (2)$$

Each element in D_0 is called a *feasible solution*. The *optimal function value* is

$$f^* = \inf_{\mathbf{x} \in D_0} f(\mathbf{x}). \quad (3)$$

Assumption 1. Feasibility and Boundedness *The feasible set is nonempty and the objective function is bounded from below, that is,*

$$-\infty < f^* = \inf_{\mathbf{x} \in D_0} f(\mathbf{x}) < \infty.$$

2 The Lagrangian Dual Problem

2.1 Weak duality

Recall from the last lecture that, for any $\lambda \geq 0$, we have

$$q(\lambda, \mu) \leq f^*.$$

This immediately leads to the result as follows.

Theorem 1. Weak Duality Theorem *We define the *dual optimal value* by*

$$q^* = \sup_{\lambda \geq 0, \mu} q(\lambda, \mu). \quad (4)$$

Then, we have

$$q^* \leq f^*. \quad (5)$$

The optimization problem in (4) is the so-called *Lagrangian dual problem*. As we have shown that the dual function q is concave, the Lagrangian dual problem is indeed equivalent to a *convex optimization problem* (why?).

Theorem 1 implies that, the dual optimal value is a lower bound of the optimal function value f^* . The difference between f^* and q^* is the so-called duality gap.

Definition 1. *Duality gap* is defined by

$$f^* - q^*.$$

Remark 1. Duality gap is a commonly used termination condition for a set of optimization algorithms.

In terms of the duality gap, we naturally have a few questions to ask.

Question 1. When is the duality gap zero, i.e., $q^* = f^*$?

Question 2. Suppose that the duality gap is zero, and there exists (λ^*, μ^*) with $\lambda^* \geq 0$ such that

$$q^* = q(\lambda^*, \mu^*) = \inf_{\mathbf{x} \in X} L(\mathbf{x}, \lambda^*, \mu^*).$$

Then, if $\hat{\mathbf{x}}$ minimizes $L(\mathbf{x}, \lambda^*, \mu^*)$, that is,

$$\hat{\mathbf{x}} \in \underset{\mathbf{x} \in X}{\operatorname{argmin}} L(\mathbf{x}, \lambda^*, \mu^*), \quad (6)$$

can we say that, $\hat{\mathbf{x}}$ is one of the optimal solutions to the primal problem, i.e.,

$$\hat{\mathbf{x}} \in \underset{\mathbf{x} \in D_0}{\operatorname{argmin}} f(\mathbf{x})?$$

All of the subsequent discussions are trying to answer the above questions.

Remark 2. The major motivation for introducing the Lagrangian is to transforming a **constrained** optimization problem with feasible set D_0 to an (almost) **unconstrained** optimization problem with feasible set X , while the optimal function value remains the same.

2.2 The Geometric Multipliers

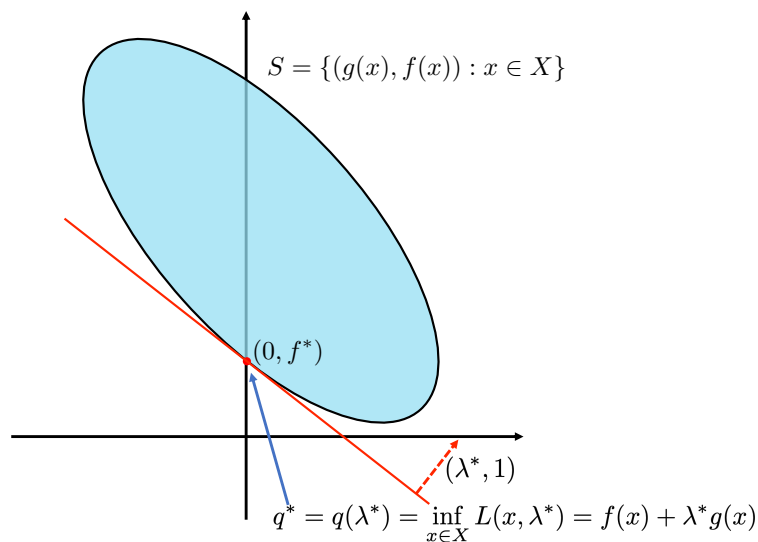


Figure 1: Illustration of the geometric multipliers.

In view of Figure 1, the equality $q^* = f^*$ holds implies that, we can find a hyperplane with the normal vector $(\lambda^*, 1)$ that supports the set S from below intercepts the vertical axis at the level f^* . In this case, we can see that the duality gap is zero. This motivates the concept **geometric multipliers** as follows.

Definition 2. A vector $(\lambda^*, \mu^*) = (\lambda_1^*, \dots, \lambda_m^*, \mu_1^*, \dots, \mu_p^*)$ is said to be a geometric multiplier vector (or simply geometric multiplier) for the primal problem if

$$\lambda_i^* \geq 0, i = 1, \dots, m,$$

and

$$f^* = \inf_{\mathbf{x} \in X} L(\mathbf{x}, \lambda^*, \mu^*). \quad (7)$$

Remark 3. Notice that, Eq. (7) is a requirement of the geometric multiplier instead of a definition of f^* , that is, $f^* = \inf_{\mathbf{x} \in D_0} f(\mathbf{x})$.

The existence of geometric multipliers indeed implies that there is no duality gap. We formalize this result by the proposition as follows.

Proposition 1. Suppose that (λ^*, μ^*) is a geometric multiplier vector of the primal problem. Then, we have the following hold.

1. $q^* = q(\lambda^*, \mu^*)$, that is, (λ^*, μ^*) is one of the dual optimal solutions to the Lagrangian dual problem (4);
2. the duality gap is zero, i.e., $f^* = q^*$.

Proof. Recall that, the Lagrangian dual function is defined by

$$q(\lambda, \mu) = \inf_{\mathbf{x} \in X} L(\mathbf{x}, \lambda, \mu).$$

Thus, the right hand side of Eq. (7) is indeed $q(\lambda^*, \mu^*)$, and we can write the condition in Eq. (7) as

$$f^* = \inf_{\mathbf{x} \in X} L(\mathbf{x}, \lambda^*, \mu^*) = q(\lambda^*, \mu^*). \quad (8)$$

By further noting the weak duality property in (5) and the condition $\lambda \geq 0$ in Definition 2, we can conclude that

$$q^* = q(\lambda^*, \mu^*), \quad (9)$$

that is, the geometric multiplier (λ^*, μ^*) is one of the dual optimal solutions to the Lagrangian dual problem (4). Moreover, combining (8) and (9) immediately leads to $f^* = q^*$, which completes the proof. \square

Remark 4. If we can find a geometric multiplier, then there is no duality gap. However, the converse is not true. That is, if there is no duality gap, we may not be able to find a geometric multiplier. They may not even exist at all.

Example 1. Consider an optimization problem as follows.

$$\begin{aligned} \min \quad & f(x) = x \\ \text{s.t.} \quad & g(x) = x^2 \leq 0, \\ & x \in X = \mathbb{R}. \end{aligned}$$

2.3 The Complementary Slackness

If a geometric multiplier (λ^*, μ^*) is known, we hope that $\hat{\mathbf{x}}$ that minimizes the Lagrangian $L(\mathbf{x}, \lambda^*, \mu^*)$ over $\mathbf{x} \in X$ is one of the optimal solutions to the primal problem as well. However, the vector $\hat{\mathbf{x}} \in \mathbf{argmin}_{\mathbf{x} \in X} L(\mathbf{x}, \lambda^*, \mu^*)$ may not even be in the feasible set D_0 .

Example 2. Consider an optimization problem as follows.

$$\begin{aligned} \min f(x) &= \begin{cases} e^x, & x \leq 0, \\ 1 - x, & x \in [0, 1], \\ 0, & x > 1. \end{cases} \\ \text{s.t. } g(x) &= x \leq 0. \end{aligned}$$

We can see that, the geometric multiplier λ^* is 0, and the corresponding Lagrangian is $L(x, \lambda^*) = f(x)$. Thus,

$$\mathbf{argmin}_{x \in \mathbb{R}} L(x, \lambda^*) = \{x : x \geq 1\}.$$

Clearly, none of the points that minimizes $L(x, \lambda^*)$ is feasible regarding the primal problem.

Example 2 implies that, even we can find a geometric multiplier (λ^*, μ^*) , solving the problem in (6) may not lead us to the optimal solutions to the primal problem. Thus, we need extra conditions to find the desirable optimal solutions from the set in (6).

Proposition 2. Let (λ^*, μ^*) be a geometric multiplier. Then, \mathbf{x}^* is a global minimum of the primal problem if and only if \mathbf{x}^* is feasible and

$$\mathbf{x}^* \in \mathbf{argmin}_{\mathbf{x} \in X} L(\mathbf{x}, \lambda^*, \mu^*), \quad (10)$$

$$\lambda_i^* g_i(\mathbf{x}^*) = 0, \quad i = 1, \dots, m. \quad (11)$$

Proof.

1. (\Rightarrow) Suppose that \mathbf{x}^* is a global minimum of the primal problem. Then, \mathbf{x}^* must be feasible, and thus

$$f(\mathbf{x}^*) \geq L(\mathbf{x}^*, \lambda^*, \mu^*) \geq f^*.$$

The definition of f^* leads to $f^* = f(\mathbf{x}^*)$, which implies that the above inequality is an equality. Thus,

$$f(\mathbf{x}^*) = L(\mathbf{x}^*, \lambda^*, \mu^*) = f^* = \inf_{\mathbf{x} \in X} L(\mathbf{x}, \lambda^*, \mu^*).$$

This leads to (10) and

$$f(\mathbf{x}^*) = L(\mathbf{x}^*) = f(\mathbf{x}^*) + \langle \lambda^*, \mathbf{g}(\mathbf{x}^*) \rangle + \langle \mu^*, \mathbf{h}(\mathbf{x}^*) \rangle.$$

As \mathbf{x}^* is feasible, that is, $\mathbf{g}(\mathbf{x}^*) \leq 0$ and $\mathbf{h}(\mathbf{x}^*) = 0$, we have Eq. (11).

2. (\Leftarrow) Suppose that \mathbf{x}^* is feasible and (10) and (11) hold.

In view of (10) and the fact that (λ^*, μ^*) is the geometric multiplier, we have

$$L(\mathbf{x}^*, \lambda^*, \mu^*) = f^* = \inf_{\mathbf{x} \in D_0} f(\mathbf{x}).$$

Moreover, the feasibility of \mathbf{x}^* and (11) imply that

$$L(\mathbf{x}^*, \lambda^*, \mu^*) = f(\mathbf{x}^*).$$

Combining the above two equations leads to

$$f(\mathbf{x}^*) = \inf_{\mathbf{x} \in D_0} f(\mathbf{x}),$$

which implies that \mathbf{x}^* is a global minimum of the primal problem.

The proof is complete. \square

Remark 5. The condition in (11) is the so-called **complementary slackness**, which implies that

$$\begin{aligned} \lambda_i^* > 0 &\Rightarrow g_i(\mathbf{x}^*) = 0, \\ g_i(\mathbf{x}^*) < 0 &\Rightarrow \lambda_i^* = 0. \end{aligned}$$

Complementary slackness is frequently used in characterizing the optimal solutions.

Question 3. Let (λ^*, μ^*) be a geometric multiplier, \mathbf{x}^* be a vector such that Eq. (10) holds, and \mathbf{x}^* is feasible. Can we say that \mathbf{x}^* is a global minimum of the primal problem?

2.4 Primal and dual optimal solutions

Theorem 2. Optimality Conditions *A pair \mathbf{x}^* and (λ^*, μ^*) is an optimal solution and geometric multiplier pair if and only if*

$$\mathbf{x}^* \in X, \mathbf{g}(\mathbf{x}^*) \leq 0, \mathbf{h}(\mathbf{x}^*) = 0, \quad (\text{Primal Feasibility}), \quad (12)$$

$$\lambda^* \geq 0, \quad (\text{Dual Feasibility}), \quad (13)$$

$$\mathbf{x}^* \in \underset{\mathbf{x} \in X}{\operatorname{argmin}} L(\mathbf{x}, \lambda^*, \mu^*), \quad (\text{Lagrangian Optimality}), \quad (14)$$

$$\lambda_i^* g_i(\mathbf{x}^*) = 0, i = 1, \dots, m, \quad (\text{Complementary Slackness}). \quad (15)$$

Proof.

1. \Rightarrow Suppose that \mathbf{x}^* and (λ^*, μ^*) is an optimal solution and geometric multiplier pair. Then, the primal feasibility and dual feasibility hold.

Moreover,

$$f(\mathbf{x}^*) = f^* = \inf_{\mathbf{x} \in X} L(\mathbf{x}, \lambda^*, \mu^*) \leq L(\mathbf{x}^*, \lambda^*, \mu^*) \leq f(\mathbf{x}^*),$$

which implies the Lagrangian optimality and the complementary slackness.

2. \Leftarrow Suppose that the conditions in (12) to (15) hold. Then

$$f(\mathbf{x}^*) = L(\mathbf{x}^*, \lambda^*, \mu^*) = \min_{\mathbf{x} \in X} L(\mathbf{x}, \lambda^*, \mu^*) \leq \inf_{\mathbf{x} \in D_0} L(\mathbf{x}, \lambda^*, \mu^*) \leq \inf_{\mathbf{x} \in D_0} f(\mathbf{x}) \leq f(\mathbf{x}^*),$$

which implies that \mathbf{x}^* is the optimal solution and (λ^*, μ^*) is the geometric multiplier.

The proof is complete. \square

Proposition 3. Saddle Point Theorem A pair \mathbf{x}^* and (λ^*, μ^*) is an optimal solution-geometric multiplier pair if and only if $\mathbf{x}^* \in X$, $\lambda^* \geq 0$, and $(\mathbf{x}^*, \lambda^*, \mu^*)$ is a saddle point of the Lagrangian, in the sense that

$$L(\mathbf{x}^*, \lambda, \mu) \leq L(\mathbf{x}^*, \lambda^*, \mu^*) \leq L(\mathbf{x}, \lambda^*, \mu^*), \forall \mathbf{x} \in X, \lambda \geq 0, \mu \in \mathbb{R}^p. \quad (16)$$

Proof.

1. \Rightarrow As the pair \mathbf{x}^* and (λ^*, μ^*) is an optimal solution-geometric multiplier pair, we have (12) to (15) hold. Clearly, we can see that $\mathbf{x}^* \in X$, $\lambda^* \geq 0$, and the Lagrangian optimality in (14) implies that

$$L(\mathbf{x}^*, \lambda^*, \mu^*) \leq L(\mathbf{x}, \lambda^*, \mu^*), \forall \mathbf{x} \in X.$$

Moreover, in view of the definition of geometric multiplier, we have

$$f^* = \inf_{\mathbf{x} \in X} L(\mathbf{x}, \lambda^*, \mu^*) = L(\mathbf{x}^*, \lambda^*, \mu^*).$$

Thus, combining the feasibility of \mathbf{x}^* and $\lambda \geq 0$ leads to

$$L(\mathbf{x}^*, \lambda, \mu) = f(\mathbf{x}^*) + \langle \lambda, \mathbf{g}(\mathbf{x}^*) \rangle + \langle \mu, \mathbf{h}(\mathbf{x}^*) \rangle \leq f(\mathbf{x}^*) = L(\mathbf{x}^*, \lambda^*, \mu^*),$$

which completes the proof.

2. \Leftarrow In view of Theorem 2, it suffices to show that (12) and (15) hold. The left half of the saddle point property of the Lagrangian in (16) implies that

$$\begin{aligned} L(\mathbf{x}^*, \lambda, \mu) &\leq L(\mathbf{x}^*, \lambda^*, \mu^*), \forall \lambda \geq 0, \\ \Rightarrow f(\mathbf{x}^*) + \langle \lambda, \mathbf{g}(\mathbf{x}^*) \rangle + \langle \mu, \mathbf{h}(\mathbf{x}^*) \rangle &\leq L(\mathbf{x}^*, \lambda^*, \mu^*), \forall \lambda \geq 0. \end{aligned}$$

In other words, $L(\mathbf{x}^*, \lambda, \mu)$ is upper bounded for any $\lambda \geq 0$. Consequently, we have

$$\mathbf{g}(\mathbf{x}^*) \leq 0, \mathbf{h}(\mathbf{x}^*) = 0,$$

i.e., the primal feasibility (12) holds (otherwise $L(\mathbf{x}^*, \lambda, \mu)$ can not be upper bounded).

To show that the complementary slackness in (15) holds, we combine the primal feasibility of \mathbf{x}^* and left half of (16)

$$\begin{aligned} f(\mathbf{x}^*) + \langle \lambda, \mathbf{g}(\mathbf{x}^*) \rangle &\leq f(\mathbf{x}^*) + \langle \lambda^*, \mathbf{g}(\mathbf{x}^*) \rangle, \forall \lambda \geq 0, \\ \xrightarrow{\lambda \rightarrow 0} \langle \lambda^*, \mathbf{g}(\mathbf{x}^*) \rangle &= \sum_{i=1}^m \lambda_i^* g_i(\mathbf{x}^*) \geq 0. \end{aligned}$$

On the other hand, in view of the facts that $\lambda^* \geq 0$ and $\mathbf{g}(\mathbf{x}^*) \leq 0$, we have

$$\lambda_i^* g_i(\mathbf{x}^*) \leq 0, i = 1, \dots, m.$$

All together, we have

$$\lambda_i^* g_i(\mathbf{x}^*) = 0, i = 1, \dots, m.$$

Thus, the complementary slackness holds and the proof is complete. \square

2.5 Strong duality

We discuss conditions that ensure the duality gap is zero.

Proposition 4. Strong Duality Theorem - Linear Constraints *Consider the primal problem. Suppose that f is convex, D_0 is a polyhedron (that is, $X = \{\mathbf{x} : \langle \mathbf{a}_i, \mathbf{x} \rangle \leq b_i, i = 1, \dots, r\}$), and f^* is finite. Then, there is no duality gap and there exists at least one geometric multiplier.*

Proposition 5. Linear and Quadratic Programming Duality *Consider the primal problem. Suppose that f is convex quadratic, D_0 is a polyhedron, and f^* is finite. Then, the primal and dual problems have optimal solutions, and the duality gap is 0.*

3 The Dual Problem of SVM

The Primal Problem

Recall that the soft margin SVM takes the form of

$$\begin{aligned} \min_{\mathbf{w}, b, \xi} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i, \\ \text{s.t.} \quad & y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq 1 - \xi_i, i \in [n], \\ & \xi_i \geq 0, i \in [n]. \end{aligned} \tag{17}$$

The *primal variables* are \mathbf{w} , b , and ξ . By Proposition (5), the strong duality holds.

The Lagrangian

To find the dual problem of (17), we first construct the Lagrangian:

$$L(\mathbf{w}, b, \xi, \alpha, \mu) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i + \sum_{i=1}^n \alpha_i (1 - \xi_i - y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b)) - \sum_{i=1}^n \mu_i \xi_i,$$

where $\alpha_i, \mu_i \geq 0$, $i = 1, \dots, n$, are the *dual variables*.

The Dual Function

We next find the dual function:

$$\begin{aligned} q(\alpha, \mu) &= \inf_{\mathbf{w}, b, \xi} L(\mathbf{w}, b, \xi, \alpha, \mu) \\ &= \inf_{\mathbf{w}} \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i y_i \langle \mathbf{w}, \mathbf{x}_i \rangle \\ &\quad + \inf_b -b \sum_{i=1}^n \alpha_i y_i \\ &\quad + \inf_{\xi} \sum_{i=1}^n (C - \alpha_i - \mu_i) \xi_i. \end{aligned} \tag{18}$$

For fixed (α, μ) , let $(\hat{\mathbf{w}}, \hat{b}, \hat{\xi})$ be the optimal solution to the above problem. The first order optimal condition implies that

$$\begin{aligned}\nabla_{\mathbf{w}} L(\mathbf{w}, b, \xi, \alpha, \mu)|_{\mathbf{w}=\hat{\mathbf{w}}} &= 0 \Rightarrow \hat{\mathbf{w}} - \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i = 0, \\ \nabla_b L(\mathbf{w}, b, \xi, \alpha, \mu)|_{b=\hat{b}} &= 0 \Rightarrow - \sum_{i=1}^n \alpha_i y_i = 0, \\ \nabla_{\xi_i} L(\mathbf{w}, b, \xi, \alpha, \mu)|_{\xi_i=\hat{\xi}_i} &= 0 \Rightarrow C - \alpha_i - \mu_i = 0, i = 1, \dots, n.\end{aligned}$$

Plugging the above equations into Eq. (18) leads to

$$q(\alpha, \mu) = -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle + \sum_{i=1}^n \alpha_i. \quad (19)$$

The Dual Problem

Thus, the dual problem of the soft margin SVM in (17) is

$$\begin{aligned}\max_{\alpha} \quad & -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle + \sum_{i=1}^n \alpha_i \\ \text{s.t.} \quad & \sum_{i=1}^n \alpha_i y_i = 0, \\ & C - \alpha_i - \mu_i = 0, \\ & \alpha_i \geq 0, \\ & \mu_i \geq 0, i = 1, \dots, n.\end{aligned}$$

We can remove μ from the problem by noting that

$$\mu_i = C - \alpha_i, i = 1, \dots, n,$$

which leads to

$$\begin{aligned}\min_{\alpha} \quad & \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle - \sum_{i=1}^n \alpha_i \\ \text{s.t.} \quad & \sum_{i=1}^n \alpha_i y_i = 0, \\ & \alpha_i \in [0, C], i = 1, \dots, n.\end{aligned} \quad (20)$$

Complementary Slackness

Let $(\mathbf{w}^*, b^*, \xi^*)$ and (α^*, μ^*) be the optimal solutions to the primal and dual problems of SVM, respectively. By Theorem 2, we write the complementary slackness as follows.

$$\alpha_i^* (1 - \xi_i^* - y_i (\langle \mathbf{w}^*, \mathbf{x}_i \rangle + b^*)) = 0, i = 1, \dots, n, \quad (21)$$

$$\mu_i^* (-\xi_i^*) = (C - \alpha_i^*) (-\xi_i^*) = 0, i = 1, \dots, n. \quad (22)$$

By the complementary slackness in (21) and (22), we have several interesting observations.

1. Suppose that one of the entries of α^* , say α_k^* , falls in the interval $(0, C)$. Then, the complementary slackness conditions (21) and (22) implies that

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

respectively. Clearly, we have

$$y_k(\langle \mathbf{w}^*, \mathbf{x}_k \rangle + b^*) = 1, \quad (23)$$

which implies that \mathbf{x}_k is a support vector.

2. Suppose that

$$1 - \xi_k^* - y_k(\langle \mathbf{w}^*, \mathbf{x}_k \rangle + b^*) < 0.$$

Then, by (21) and (22), we have $\alpha_k^* = 0$ and $\xi_k^* = 0$, respectively. Thus,

$$y_k(\langle \mathbf{w}^*, \mathbf{x}_k \rangle + b^*) > 1,$$

which implies that \mathbf{x}_k is correctly classified and outside of the region between the marginal hyperplanes.

Recovering the Primal Optimum from the Dual Optimum

Proposition 6. *Let α^* be one of the optimal solutions to (20). Suppose that α_k^* is one of the entries of α^* and $\alpha_k^* \in (0, C)$, then we can find a primal optimal solution by*

$$\begin{aligned} \mathbf{w}^* &= \sum_{i=1}^n \alpha_i^* y_i \mathbf{x}_i, \\ b^* &= y_k - \langle \mathbf{w}^*, \mathbf{x}_k \rangle. \end{aligned}$$

References