EE3001 Machine Learning **Fall 2025**

Lecture 13. Support Vector Machine II

Lecturer: Jie Wang Date: Dec 5, 2025

Last Update: December 7, 2025

1 The Primal Problem

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

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

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

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

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

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

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 $\lambda \ge 0$, we have

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

This immediately leads to the result as follows.

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

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

Then, we have

$$q^* \le f^*. \tag{5}$$

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, can we find **a set of** (λ^*, μ^*) with $\lambda^* \geq 0$ such that

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

Question 3. Suppose that the duality gap is zero, and there exists (λ^*, μ^*) with $\lambda^* \geq 0$ such that

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

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

$$\hat{\mathbf{x}} \in \mathop{\mathbf{argmin}}_{\mathbf{x} \in X} L(\mathbf{x}, \lambda^*, \mu^*), \tag{6}$$

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

$$\hat{\mathbf{x}} \in \mathop{\mathbf{argmin}}_{\mathbf{x} \in D_0} f(\mathbf{x})?$$

Question 4. Suppose that the duality gap is zero, and there exists (λ^*, μ^*) with $\lambda^* \geq 0$ such that

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

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

$$\hat{\mathbf{x}} \in \mathop{\mathbf{argmin}}_{\mathbf{x} \in X} L(\mathbf{x}, \lambda^*, \mu^*), \tag{7}$$

and $\hat{\mathbf{x}}$ **is feasible**, i.e., $\hat{\mathbf{x}} \in D_0$, can we say that, $\hat{\mathbf{x}}$ is one of the optimal solutions to the primal problem, i.e.,

$$\hat{\mathbf{x}} \in \mathop{\mathbf{argmin}}_{\mathbf{x} \in D_0} f(\mathbf{x})?$$

All of the subsequent discussions are trying to answer the above questions. The major tool is the geometric treatment we introduced last lecture, that is, we treat the Lagrangian as a linear function over the space where the set $S = (g_1(\mathbf{x}), \ldots, g_m(\mathbf{x}), h_1(\mathbf{x}), \ldots, h_p(\mathbf{x}), f(\mathbf{x}))$ lies in.

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

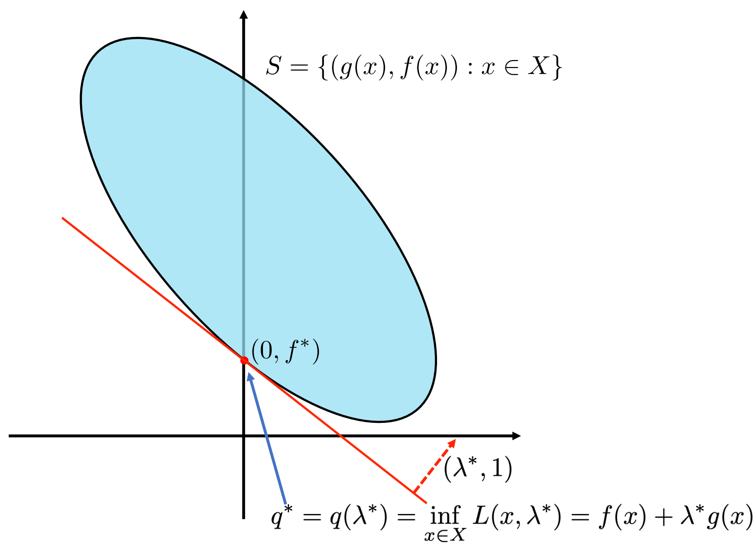

Figure 1: Illustration of the geometric multipliers.

2.2 The Geometric Multipliers

In view of Figure 1, the equality $q^* = f^*$ holds implies that, we may be able to (but not necessarily, see ***Remark*** 5) 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

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

Remark *3.* Notice that, Eq. (8) is a requirement of the geometric multiplier instead of a definition of f^*. Recall that,

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

Remark *4.* The RHS of Eq. (8) is indeed $q(\lambda^*, \mu^*)$. Therefore, the existence of a geometric multiplier (λ^*, μ^*) implies that we can find a feasible solution (λ^*, μ^*) of the dual problem such that $f^* = q(\lambda^*, \mu^*)$.

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. (8) is indeed $q(\lambda^*, \mu^*)$, and we can write the condition in Eq. (8) as

$$f^* = \inf_{\mathbf{x} \in X} L(\mathbf{x}, \lambda^*, \mu^*) = q(\lambda^*, \mu^*). \tag{9}$$

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

$$q^* = q(\lambda^*, \mu^*), \tag{10}$$

that is, the geometric multiplier (λ^*, μ^*) is one of the dual optimal solutions to the Lagrangian dual problem (4). Moreover, combining (9) and (10) immediately leads to $f^* = q^*$, which completes the proof. □

Remark 5. 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.

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

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,

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

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

What if $\hat{\mathbf{x}} \in \mathbf{argmin}_{\mathbf{x}\in X} L(\mathbf{x}, \lambda^*, \mu^*)$ is a feasible solution to the primal problem? Can we conclude that such a $\hat{\mathbf{x}}$ is an optimal solution to the primal problem? The answer is still no.

Example 3. Consider an optimization problem as follows.

$$\min f(x) = \begin{cases} -x, \; x \leq 0, \\ 0, \quad x > 0. \end{cases}$$
$$\text{s.t.} \;\; g(x) = x \leq 0.$$

We can see that, the geometric multiplier λ^* is not unique and it can take any value from $[0, 1]$. Let us consider the case in which $\lambda^* = 1$ and the corresponding Lagrangian is

$$L(x, \lambda^*) = f(x) + g(x) = \begin{cases} 0, \; x \leq 0, \\ x, \; x > 0. \end{cases}$$

Thus,

$$\mathop{\mathbf{argmin}}_{x\in\mathbb{R}} L(x, \lambda^*) = \{x : x \leq 0\}.$$

However, it is easy to see that only $x^* = 0$ is the optimal solution to the problem.

***Remark* 6.** Notice that, Example 3 also provides us an example that the geometric multiplier may not be unique. Indeed, for Example 3, the geometric multiplier is $\lambda^* \in [0, 1]$.

Thus, we need extra conditions to find the desirable optimal solutions from the set in (7), which is the so-called ***complementary slackness***.

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

$$\mathbf{x}^* \text{ is feasible}, \tag{11}$$
$$\mathbf{x}^* \in \mathop{\mathbf{argmin}}_{\mathbf{x}\in X} L(\mathbf{x}, \lambda^*, \mu^*), \tag{12}$$
$$\lambda_i^* g_i(\mathbf{x}^*) = 0, \; i = 1, \ldots, m. \tag{13}$$

Proof.

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

$$f(\mathbf{x}^*) \geq f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}^*) + \sum_{i=1}^{p} \mu_i h_i(\mathbf{x}^*) = L(\mathbf{x}^*, \lambda^*, \mu^*) \geq \inf_{\mathbf{x}\in X} L(\mathbf{x}, \lambda^*, \mu^*) = f^*.$$

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 (12) and

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

As $\mathbf{x}^*$ is feasible, that is, $\mathbf{g}(\mathbf{x}^*) \leq 0$ and $\mathbf{h}(\mathbf{x}^*) = 0$, we have Eq. (13).

2. (⇐) Suppose that $\mathbf{x}^*$ is feasible and (12) and (13) hold.

   In view of (12) 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 (13) imply that

$$L(\mathbf{x}^*, \lambda^*, \mu^*) = f(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}^*) + \sum_{i=1}^{p} \mu_i h_i(\mathbf{x}^*) = f(\mathbf{x}^*).$$

   Combining the above two equations leads to

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

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

The proof is complete. □

Remark* 7.** ***Complementary slackness in (13) implies that

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

Complementary slackness is frequently used in characterizing the optimal solutions.

2.4 Primal and dual optimal solutions

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

$$\mathbf{x}^* \in X,\ \mathbf{g}(\mathbf{x}^*) \leq 0, \mathbf{h}(\mathbf{x}^*) = 0, \quad \text{(Primal Feasibility)}, \tag{14}$$
$$\lambda^* \geq 0, \quad \text{(Dual Feasibility)}, \tag{15}$$
$$\mathbf{x}^* \in \mathop{\mathbf{argmin}}_{\mathbf{x}\in X} L(\mathbf{x}, \lambda^*, \mu^*), \quad \text{(Lagrangian Optimality)}, \tag{16}$$
$$\lambda_i^* g_i(\mathbf{x}^*) = 0,\ i = 1, \ldots, m, \quad \text{(Complementary Slackness)}. \tag{17}$$

Proof.

1. ⇒ 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. ⇐ Suppose that the conditions in (14) to (17) 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. □

Proposition 3. Saddle Point Theorem (Optional) *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*

$$L(\mathbf{x}^*, \lambda, \mu) \leq L(\mathbf{x}^*, \lambda^*, \mu^*) \leq L(\mathbf{x}, \lambda^*, \mu^*), \forall \mathbf{x} \in X, \lambda \geq 0, \mu \in \mathbb{R}^p. \tag{18}$$

Proof.

1. $\Rightarrow$ As the pair $\mathbf{x}^*$ and (λ^*, μ^*) is an optimal solution-geometric multiplier pair, we have (14) to (17) hold. Clearly, we can see that $\mathbf{x}^* \in X$, $\lambda^* \geq 0$, and the Lagrangian optimality in (16) 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 (14) and (17) hold. The left half of the saddle point property of the Lagrangian in (18) 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 (14) holds (otherwise $L(\mathbf{x}^*, \lambda, \mu)$ can not be upper bounded).

   To show that the complementary slackness in (17) holds, we combine the primal feasibility of $\mathbf{x}^*$ and left half of (18)

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

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

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

   All together, we have

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

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

□

2.5 Strong duality

We discuss conditions that ensure the duality gap is zero.

Theorem 3. [1] *Suppose that the primal problem in (1) is a convex optimization problem, that is, f and g_i, $i = 1, \ldots, m$ are convex, h_i, $i = 1, \ldots, p$ are affine, and X is a convex set. If there exists an $\hat{\mathbf{x}} \in X$ such that $\mathbf{g}(\hat{\mathbf{x}}) < \mathbf{0}$ and $\mathbf{h}(\hat{\mathbf{x}}) = \mathbf{0}$, and $\mathbf{0} \in \textbf{int } \mathbf{h}(X)$, where $\mathbf{h}(X) = \{\mathbf{h}(\mathbf{x}) : \mathbf{x} \in X\}$. Then, the duality gap is zero. Furthermore, if f^* is finite, then there exists at least one geometric multiplier.*

Proposition 4. [2] **Strong Duality Theorem - Linear Constraints** *Consider the primal problem. Suppose that f is convex, D_0 is a polyhedron (that is, $D_0 = \{\mathbf{x} : \langle \mathbf{a}_i, \mathbf{x} \rangle \leq b_i, i = 1, \ldots, r\}$), and f^* is finite. Then, there is no duality gap and there exists at least one geometric multiplier.*

Proposition 5. [2] **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

$$\begin{aligned} &\min_{\mathbf{w},b,\xi} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n} \xi_i, \\ &\text{s.t. } y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq 1 - \xi_i, i \in [n], \\ &\qquad \xi_i \geq 0,\ i \in [n]. \end{aligned} \tag{19}$$

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

The Lagrangian

To find the dual problem of (19), 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 $\alpha_i, \mu_i \geq 0$, $i = 1, \ldots, n$, are the *dual variables*.

The Dual Function

We next find the dual function:

$$
\begin{aligned}
q(\alpha, \mu) &= \inf_{\mathbf{w}, b, \xi} L(\mathbf{w}, b, \xi, \alpha, \mu) \quad (20)\\
&= \inf_{\mathbf{w}} \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \alpha_i y_i \langle \mathbf{w}, \mathbf{x}_i \rangle \\
&+ \inf_{b} -b \sum_{i=1}^{n} \alpha_i y_i \\
&+ \inf_{\xi} \sum_{i=1}^{n} (C - \alpha_i - \mu_i)\xi_i.
\end{aligned}
$$

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

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

Plugging the above equations into Eq. (20) leads to

$$
q(\alpha, \mu) = -\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle + \sum_{i=1}^{n} \alpha_i. \quad (21)
$$

The Dual Problem

Thus, the dual problem of the soft margin SVM in (19) is

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

We can remove μ from the problem by noting that

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

which leads to

$$\min_{\alpha} \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j\rangle - \sum_{i=1}^{n}\alpha_i \tag{22}$$
$$\text{s.t.} \quad \sum_{i=1}^{n}\alpha_i y_i = 0,$$
$$\alpha_i \in [0, C], i = 1, \ldots, n.$$

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.

$$\alpha_i^*(1 - \xi_i^* - y_i(\langle \mathbf{w}^*, \mathbf{x}_i\rangle + b^*)) = 0,\ i = 1, \ldots, n, \tag{23}$$
$$\mu_i^*(-\xi_i^*) = (C - \alpha_i^*)(-\xi_i^*) = 0,\ i = 1, \ldots, n. \tag{24}$$

By the complementary slackness in (23) and (24), 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 (23) and (24) implies that

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

   respectively. Clearly, we have

   $$y_k(\langle \mathbf{w}^*, \mathbf{x}_k\rangle + b^*) = 1, \tag{25}$$

   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 (23) and (24), 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 (22). Then, we have*

$$\mathbf{w}^* = \sum_{i=1}^{n} \alpha^* y_i \mathbf{x}_i.$$

If further α_k^* is one of the entries of α^* and $\alpha_k^* \in (0, C)$, then we have

$$b^* = y_k - \langle \mathbf{w}^*, \mathbf{x}_k\rangle.$$

References

[1] M. Bazaraa, H. Sherali, and C. Shetty. *Nonlinear Programming.* Wiley-Interscience, 2006.

[2] D. P. Bertsekas. *Nonlinear Programming, 3ed.* Athena Scientific, 2016.