EE3001 Machine Learning Fall 2025

Lecture 13. Support Vector Machine II

Lecturer: Xiaojun Chang

Last Update: December 12, 2025

1 The Primal Problem

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

$$\min_x f(x) \tag{1}$$

subject to

$$g(x) \leq 0, \qquad h(x) = 0, \qquad x \in X.$$

We denote the feasible set of (1) by

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

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

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

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

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

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

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

Then, we have

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

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, and there exists (λ^*, μ^*) with $\lambda^* \geq 0$ such that

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

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

$$\hat{x} \in \arg\min_{x \in X} L(x, \lambda^*, \mu^*), \tag{6}$$

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

$$\hat{x} \in \arg\min_{x \in D_0} f(x)?$$

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 the 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

(See Figure 1 for an illustration.)

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

$$(\lambda^*, \mu^*) = (\lambda_1^*, \ldots, \lambda_m^*, \mu_1^*, \ldots, \mu_p^*)$$

is said to be a geometric multiplier vector (or simply geometric multiplier) for the primal problem if

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

and

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

Remark 3. Notice that, Eq. (7) is a requirement of the geometric multiplier instead of a definition of f^*. Recall that

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

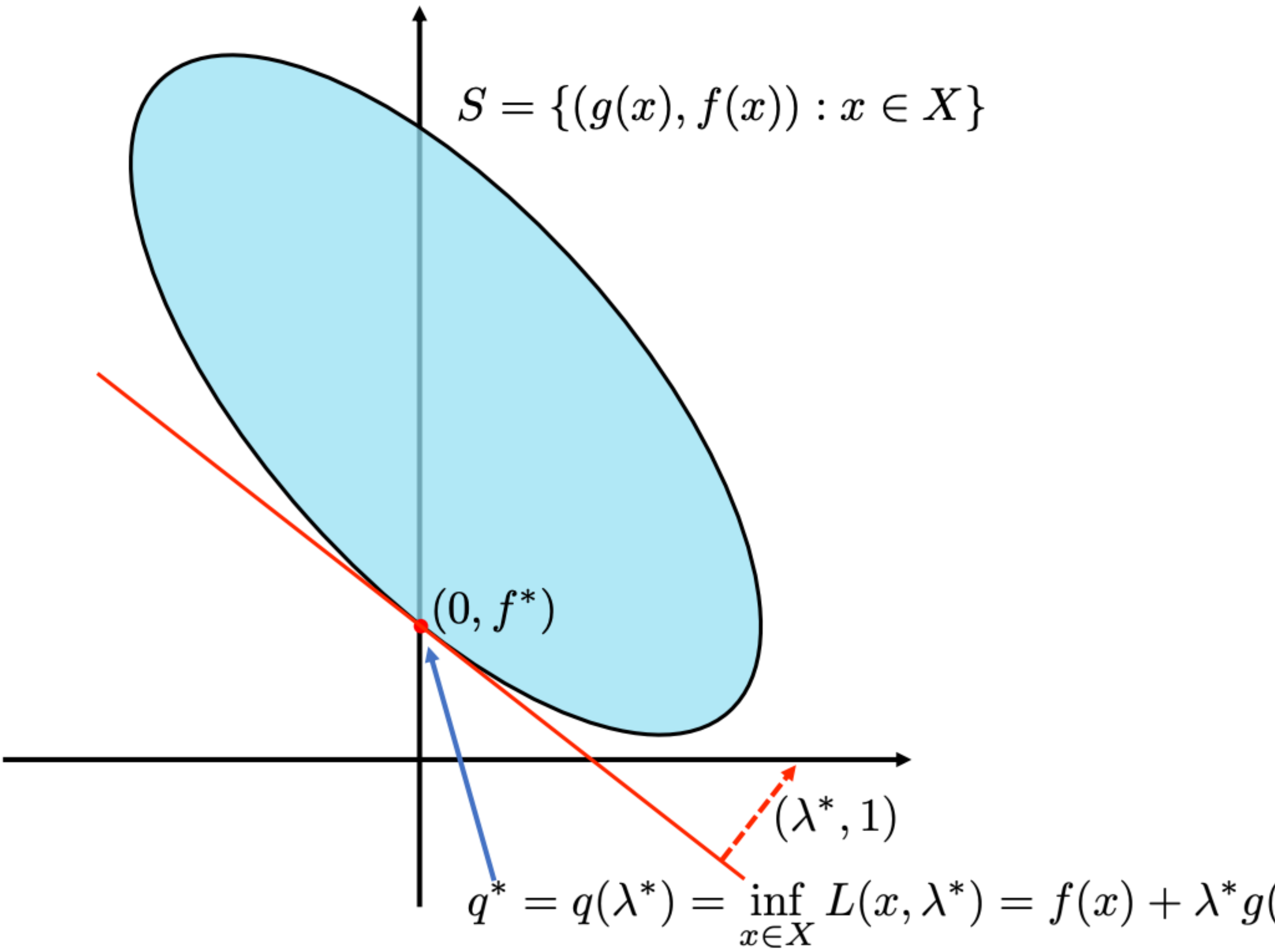

Figure 1: Illustration of the geometric multipliers.

Remark 4. The RHS of Eq. (7) 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

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

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

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

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

$$q^* = q(\lambda^*, \mu^*), \tag{9}$$

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. □

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.

$$\min f(x) = x$$

subject to

$$g(x) = x^2 \leq 0, \qquad x \in X = \mathbb{R}.$$

2.3 The Complementary Slackness

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

Example 2. Consider an optimization problem as follows.

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

subject to

$$g(x) = x \leq 0.$$

We can see that, the geometric multiplier λ^* is 0, and the corresponding Lagrangian is

$$L(x, \lambda^*) = f(x).$$

Thus,

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

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

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

Example 3. Consider an optimization problem as follows.

$$\min f(x) = \begin{cases} -x, & x \leq 0, \\ 0, & x > 0, \end{cases}$$

subject to

$$g(x) = x \leq 0.$$

We can see that, the geometric multiplier λ^* is 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,

$$\arg\min_{x \in \mathbb{R}} L(x, \lambda^*) = \{x : x \leq 0\}.$$

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 (6), which is the so-called *complementary slackness*.

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

$$x^* \text{ is feasible}, \tag{10}$$

$$x^* \in \arg\min_{x \in X} L(x, \lambda^*, \mu^*), \tag{11}$$

$$\lambda_i^* g_i(x^*) = 0, \quad i = 1, \ldots, m. \tag{12}$$

Proof.

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

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

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

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

This leads to (11) and

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

As x^* is feasible, that is, $g(x^*) \leq 0$ and $h(x^*) = 0$, we have Eq. (12).

2. ($\Leftarrow$) Suppose that x^* is feasible and (11) and (12) hold.

In view of (11) and the fact that (λ^*, μ^*) is the geometric multiplier, we have

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

Moreover, the feasibility of x^* and (12) imply that

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

Combining the above two equations leads to

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

which implies that x^* is a global minimum of the primal problem.

The proof is complete. $\square$

Remark 7. Complementary slackness in (12) implies that

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

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 x^* and (λ^*, μ^*) is an optimal solution and geometric multiplier pair if and only if

$$x^* \in X,\ g(x^*) \leq 0,\ h(x^*) = 0, \qquad \text{(Primal Feasibility)}, \tag{13}$$

$$\lambda^* \geq 0, \qquad \text{(Dual Feasibility)}, \tag{14}$$

$$x^* \in \arg\min_{x \in X} L(x, \lambda^*, \mu^*), \qquad \text{(Lagrangian Optimality)}, \tag{15}$$

$$\lambda_i^* g_i(x^*) = 0,\ i = 1, \ldots, m, \qquad \text{(Complementary Slackness)}. \tag{16}$$

Proof.

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

   Moreover,
   $$f(x^*) = f^* = \inf_{x \in X} L(x, \lambda^*, \mu^*) \leq L(x^*, \lambda^*, \mu^*) \leq f(x^*),$$
   which implies the Lagrangian optimality and the complementary slackness.

2. $\Leftarrow$ Suppose that the conditions in (13) to (16) hold. Then

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

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

The proof is complete. $\square$

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

$$L(x^*, \lambda, \mu) \leq L(x^*, \lambda^*, \mu^*) \leq L(x, \lambda^*, \mu^*), \quad \forall x \in X,\ \lambda \geq 0,\ \mu \in \mathbb{R}^p. \tag{17}$$

Proof.

1. $\Rightarrow$ As the pair x^* and (λ^*, μ^*) is an optimal solution–geometric multiplier pair, we have (13) to (16) hold. Clearly, we can see that $x^* \in X$, $\lambda^* \geq 0$, and the Lagrangian optimality in (15) implies that

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

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

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

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

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

   which completes the proof.

2. $\Leftarrow$ In view of Theorem 2, it suffices to show that (13) and (16) hold. The left half of the saddle point property of the Lagrangian in (17) implies that

$$L(x^*, \lambda, \mu) \leq L(x^*, \lambda^*, \mu^*), \quad \forall \lambda \geq 0,$$

$$\Rightarrow \; f(x^*) + \langle \lambda, g(x^*) \rangle + \langle \mu, h(x^*) \rangle \leq L(x^*, \lambda^*, \mu^*), \quad \forall \lambda \geq 0.$$

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

$$g(x^*) \leq 0, \qquad h(x^*) = 0,$$

   i.e., the primal feasibility (13) holds (otherwise $L(x^*, \lambda, \mu)$ can not be upper bounded).

   To show that the complementary slackness in (16) holds, we combine the primal feasibility of x^* and left half of (17)

$$f(x^*) + \langle \lambda, g(x^*) \rangle \leq f(x^*) + \langle \lambda^*, g(x^*) \rangle, \quad \forall \lambda \geq 0,$$

$$\lambda \to 0 \Longrightarrow \langle \lambda^*, g(x^*) \rangle = \sum_{i=1}^{m} \lambda_i^* g_i(x^*) \geq 0.$$

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

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

   All together, we have

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

   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. *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 $\hat{x} \in X$ such that $g(\hat{x}) < 0$ and $h(\hat{x}) = 0$, and $0 \in \operatorname{int} h(X)$, where $h(X) = \{h(x) : x \in X\}$. Then, the duality gap is zero. Furthermore, if f^* is finite, then there exists at least one geometric multiplier.*

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

Proposition 5 (Linear and Quadratic Programming Duality). Consider the primal problem. Suppose that f is convex quadratic, X 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

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i, \tag{18}$$

subject to

$$y_i(\langle w, x_i\rangle + b) \geq 1 - \xi_i,\ i \in [n],$$
$$\xi_i \geq 0,\ i \in [n].$$

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

The Lagrangian

To find the dual problem of (18), we first construct the Lagrangian:

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

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

The Dual Function

We next find the dual function:

$$\begin{aligned} q(\alpha, \mu) &= \inf_{w,b,\xi} L(w, b, \xi, \alpha, \mu) \\ &= \inf_{w} \left(\frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i y_i\langle w, x_i\rangle\right) + \inf_{b}\left(-b\sum_{i=1}^{n} \alpha_i y_i\right) + \inf_{\xi} \sum_{i=1}^{n} (C - \alpha_i - \mu_i)\xi_i. \end{aligned} \tag{19}$$

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

$$\nabla_w L(w, b, \xi, \alpha, \mu)\big|_{w=\hat{w}} = 0 \ \Rightarrow\ \hat{w} - \sum_{i=1}^{n} \alpha_i y_i x_i = 0,$$

$$\nabla_b L(w, b, \xi, \alpha, \mu)\big|_{b=\hat{b}} = 0 \ \Rightarrow\ -\sum_{i=1}^{n} \alpha_i y_i = 0,$$

$$\nabla_{\xi_i} L(w, b, \xi, \alpha, \mu)\big|_{\xi_i=\hat{\xi}_i} = 0 \ \Rightarrow\ C - \alpha_i - \mu_i = 0,\ i = 1, \ldots, n.$$

Plugging the above equations into Eq. (19) leads to

$$q(\alpha, \mu) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j y_i y_j\langle x_i, x_j\rangle + \sum_{i=1}^{n} \alpha_i. \tag{20}$$

The Dual Problem

Thus, the dual problem of the soft margin SVM in (18) is

$$\max_{\alpha} -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j\langle x_i, x_j\rangle + \sum_{i=1}^{n}\alpha_i$$

subject to

$$\begin{aligned} \sum_{i=1}^{n}\alpha_i y_i = 0,\\ C - \alpha_i - \mu_i = 0,\\ \alpha_i \geq 0, \quad \mu_i \geq 0,\ i = 1, \ldots, n. \end{aligned}$$

We can remove μ from the problem by noting that

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

which leads to

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j\langle x_i, x_j\rangle - \sum_{i=1}^{n}\alpha_i \\ \text{s.t.} \quad & \sum_{i=1}^{n}\alpha_i y_i = 0, \\ & \alpha_i \in [0, C],\ i = 1, \ldots, n. \end{aligned} \tag{21}$$

Complementary Slackness

Let (w^*, b^*, ξ^*) 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.

$$\alpha_i^*(1 - \xi_i^* - y_i(\langle w^*, x_i\rangle + b^*)) = 0, \quad i = 1, \ldots, n, \tag{22}$$

$$\mu_i^*(-\xi_i^*) = (C - \alpha_i^*)(-\xi_i^*) = 0, \quad i = 1, \ldots, n. \tag{23}$$

By the complementary slackness in (22) and (23), 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 (22) and (23) implies that

   $$y_k(\langle w^*, x_k\rangle + b^*) = 1 - \xi_k^* = 0,$$

   and $\xi_k^* = 0$, respectively. Clearly, we have

   $$y_k(\langle w^*, x_k\rangle + b^*) = 1, \tag{24}$$

   which implies that x_k is a support vector.

2. Suppose that

   $$1 - \xi_k^* - y_k(\langle w^*, x_k\rangle + b^*) < 0.$$

   Then, by (22) and (23), we have $\alpha_k^* = 0$ and $\xi_k^* = 0$, respectively. Thus,

   $$y_k(\langle w^*, x_k\rangle + b^*) > 1,$$

   which implies that 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 (21). 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

$$w^* = \sum_{i=1}^{n} \alpha_i^* y_i x_i,$$

$$b^* = y_k - \langle w^*, x_k \rangle.$$

References

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