

Previously, in our investigation of SVMs, we formulated a constrained optimization problem that we can solve to find the optimal parameters for our hyperplane decision boundary. Recall the setup of soft-margin SVMs:

- y_i 's: ± 1 , representing positive or negative class
- x_i 's: feature vectors in \mathbb{R}^d
- ξ_i 's: slack variables representing how much an x_i is allowed to violate the margin
- C : a hyperparameter describing how severely we penalize slack
- The optimization problem for $w \in \mathbb{R}^d$ and $t \in \mathbb{R}$, the parameters of the SVM:

$$\begin{aligned} \min_{w,t,\xi_i} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & y_i(w^T x_i - t) \geq 1 - \xi_i \quad \forall i \\ & \xi_i \geq 0 \quad \forall i \end{aligned}$$

Now, we will investigate the dual of this problem, which will motivate our discussion of kernels. Before we do so, we first have to understand the primal and dual of an optimization problem.

1 The Primal and Dual of an Optimization Problem

All optimization problems can be expressed in the standard form

$$\begin{aligned} \min_x \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0 \quad i = 1, \dots, m \\ & h_j(x) = 0 \quad j = 1, \dots, n \end{aligned} \tag{1}$$

For the purposes of our discussion, assume that $x \in \mathbb{R}^d$. The main components of an optimization problem are:

- The **objective function** $f_0(x)$
- The **inequality constraints**: expressions involving $f_i(x)$
- The **equality constraints**: expressions involving $h_j(x)$

Working with the constraints can be cumbersome and challenging to manipulate, and it would be ideal if we could somehow turn this constrained optimization problem into an unconstrained one. One idea is to re-express the optimization problem into

$$\min_x \mathcal{L}(x) \tag{2}$$

where

$$\mathcal{L}(x) = \begin{cases} f_0(x) & \text{if } f_i(x) \leq 0, \forall i \in [1, m] \text{ and } h_j(x) = 0, \forall j \in [1, n] \\ \infty & \text{otherwise} \end{cases}$$

Note that the unconstrained optimization problem above is equivalent to the original constrained problem. Even though the unconstrained problem considers values that violate the constraints (and therefore are not in the feasible set for the constrained optimization problem), it will effectively ignore them because they are treated as ∞ in a minimization problem.

Even though we are now dealing with an unconstrained problem, it still is difficult to solve the optimization problem, because we still have to deal with all of the casework in the objective function $\mathcal{L}(x)$. In order to solve this issue, we have to introduce dual variables, specifically one set of dual variables for the equality constraints, and one set for the inequality constraints. If we only take into account the dual variables for the equality constraints, the optimization problem now becomes

$$\min_x \max_v \mathcal{L}(x, v) \tag{3}$$

where

$$\mathcal{L}(x, v) = \begin{cases} f_0(x) + \sum_{j=1}^n v_j h_j(x) & \text{if } f_i(x) \leq 0, \forall i \in [1, m] \\ \infty & \text{otherwise} \end{cases}$$

We are still working with an unconstrained optimization problem, except that now, we are optimizing over two sets of variables: the **primal variables** $x \in \mathbb{R}^d$ and the **dual variables** $v \in \mathbb{R}^n$. Also note that the optimization problem has now become a nested one, with an inner optimization problem that maximizes over the dual variables, and an outer optimization problem that minimizes over the primal variables. Let's examine why this optimization problem is equivalent to the original constrained optimization problem:

- Any x that violates the inequality constraints is still treated as ∞ by the outer minimization problem over x and therefore ignored
- For any x that violates the equality constraints (meaning that $\exists j$ s.t. $h_j(x) \neq 0$), the inner maximization problem over v can choose v_j as ∞ if $h_j(x) > 0$ (or v_j as $-\infty$ if $h_j(x) < 0$) to cause the inner maximization blow off to ∞ , therefore being ignored by the outer minimization over x
- For any x that does not violate any of the equality or inequality constraints, the inner maximization problem over v is simply equal to $f_0(x)$

This solution comes at a cost – in an effort to remove the equality constraints, we had to add in dual variables, one for each inequality constraint. With this in mind, let’s try to do the same for the inequality constraints. Adding in dual variable λ_i to represent each inequality constraint, we now have

$$\begin{aligned} \min_x \max_{\lambda, \mathbf{v}} \mathcal{L}(x, \lambda, \mathbf{v}) &= f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{j=1}^n \mathbf{v}_j h_j(x) \\ \text{s.t.} \quad \lambda_i &\geq 0 \quad i = 1, \dots, m \end{aligned} \tag{4}$$

For convenience, we can place the constraints involving λ into the optimization variable.

$$\min_x \max_{\lambda \geq 0, \mathbf{v}} \mathcal{L}(x, \lambda, \mathbf{v}) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{j=1}^n \mathbf{v}_j h_j(x)$$

This optimization problem above is otherwise known as the **primal** (not to be confused with the *primal variables*), and its optimal value is indeed *equivalent* to that of the original constrained optimization problem.

$$p^* = \min_x \max_{\lambda \geq 0, \mathbf{v}} \mathcal{L}(x, \lambda, \mathbf{v})$$

We can verify that this is indeed the case:

- For any x that violates the inequality constraints (meaning that $\exists i \in [1, m]$ s.t. $f_i(x) > 0$), the inner maximization problem over λ can choose λ_i as ∞ to cause the inner maximization blow off to ∞ , therefore being ignored by the outer minimization over x
- For any x that violates the equality constraints (meaning that $\exists j$ s.t. $h_j(x) \neq 0$), the inner maximization problem over \mathbf{v} can choose \mathbf{v}_j as ∞ if $h_j(x) > 0$ (or \mathbf{v}_j as $-\infty$ if $h_j(x) < 0$) to cause the inner maximization blow off to ∞ , therefore being ignored by the outer minimization over x
- For any x that does not violate any of the equality or inequality constraints, in the inner maximization problem over \mathbf{v} , the expression $\sum_{j=1}^n \mathbf{v}_j h_j(x)$ evaluates to 0 no matter what the value of \mathbf{v} is, and in the inner maximization problem over λ , the expression $\sum_{i=1}^m \lambda_i f_i(x)$ can at maximum be 0, because λ_i is constrained to be non-negative, and $f_i(x)$ is non-positive. Therefore, at best, the maximization problem sets $\lambda_i f_i(x) = 0$, and

$$\max_{\lambda \geq 0, \mathbf{v}} \mathcal{L}(x, \lambda, \mathbf{v}) = f_0(x)$$

In its full form, the objective $\mathcal{L}(x, \lambda, \mathbf{v})$ is called the **Lagrangian**, and it takes into account the unconstrained set of primal variables $x \in \mathbb{R}^d$, the constrained set of dual variables $\lambda \in \mathbb{R}^m$ corresponding to the inequality constraints, and the unconstrained set of dual variables $\mathbf{v} \in \mathbb{R}^n$ corresponding to the equality constraints. Note that our dual variables λ_i are in fact constrained, so ultimately we were not able to turn the original optimization problem into an unconstrained one, but our constraints are much simpler than before.

The **dual** of this optimization problem is still over the same optimization objective, except that now we swap the order of the maximization of the dual variables and the minimization of the primal variables.

$$d^* = \max_{\lambda \geq 0, \nu} \min_x \mathcal{L}(x, \lambda, \nu) = \max_{\lambda \geq 0, \nu} g(\lambda, \nu)$$

The dual is effectively a maximization problem (over the dual variables):

$$d^* = \max_{\lambda \geq 0, \nu} g(\lambda, \nu)$$

where

$$g(\lambda, \nu) = \min_x \mathcal{L}(x, \lambda, \nu)$$

The dual is very useful to work with, because now the inner optimization problem over x is an unconstrained problem!

2 Strong Duality and KKT Conditions

It is *always* true that the solution to the primal problem is at least as large as the solution to the dual problem:

$$p^* \geq d^* \tag{5}$$

This condition is known as **weak duality**.

Proof. We know that

$$\forall x, \lambda \geq 0, \nu \quad \max_{\tilde{\lambda} \geq 0, \tilde{\nu}} \mathcal{L}(x, \tilde{\lambda}, \tilde{\nu}) \geq \mathcal{L}(x, \lambda, \nu) \geq \min_{\tilde{x}} \mathcal{L}(\tilde{x}, \lambda, \nu)$$

More compactly,

$$\forall x, \lambda \geq 0, \nu \quad \max_{\tilde{\lambda} \geq 0, \tilde{\nu}} \mathcal{L}(x, \tilde{\lambda}, \tilde{\nu}) \geq \min_{\tilde{x}} \mathcal{L}(\tilde{x}, \lambda, \nu)$$

Since this is true for all $x, \lambda \geq 0, \nu$ this is true in particular when we set

$$x = \arg \min_{\tilde{x}} \max_{\tilde{\lambda} \geq 0, \tilde{\nu}} \mathcal{L}(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$$

and

$$\lambda, \nu = \arg \max_{\tilde{\lambda} \geq 0, \tilde{\nu}} \min_{\tilde{x}} \mathcal{L}(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$$

We therefore know that

$$p^* = \min_{\tilde{x}} \max_{\tilde{\lambda} \geq 0, \tilde{\nu}} \mathcal{L}(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) \geq \max_{\tilde{\lambda} \geq 0, \tilde{\nu}} \min_{\tilde{x}} \mathcal{L}(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) = d^*$$

□

The difference $p^* - d^*$ is known as the **duality gap**. In the case of **strong duality**, the duality gap is 0. That is, we can swap the order of the minimization and maximization and up with the same optimal value:

$$p^* = d^* \quad (6)$$

There are several useful theorems detailing the existence of strong duality, such as **Slater's theorem**, which states that if the primal problem is convex, and there exists an x that can *strictly* meet the inequality constraints and meet the equality constraints, then strong duality holds. Given that strong duality holds, the **Karush-Kuhn-Tucker (KKT) conditions** can help us find the solutions to the dual variables of the optimization problem. The KKT conditions are composed of:

1. Primal feasibility (inequalities)

$$f_i(x) \leq 0, \forall i \in [1, m]$$

2. Primal feasibility (equalities)

$$h_j(x) = 0, \forall j \in [1, n]$$

3. Dual feasibility

$$\lambda_i \geq 0, \forall i \in [1, m]$$

4. Complementary Slackness

$$\lambda_i f_i(x) = 0, \forall i \in [1, m]$$

5. Stationarity

$$\nabla_x f_0(x) + \sum_{i=1}^m \lambda_i \nabla_x f_i(x) + \sum_{j=1}^n \nu_j \nabla_x h_j(x) = 0$$

Let's see how the KKT conditions relate to strong duality.

Theorem 2.1. If x^* and λ^*, ν^* are the primal and dual solutions respectively, with zero duality gap (i.e. strong duality holds), then x^*, λ^*, ν^* also satisfy the KKT conditions.

Proof. KKT conditions 1, 2, 3 are trivially true, because the primal solution x^* must satisfy the primal constraints, and the dual solution λ^*, ν^* must satisfy the dual constraints. Now, let's prove conditions 4 and 5. We know that since strong duality holds, we can say that

$$\begin{aligned} p^* &= f_0(x^*) = g(\lambda^*, \nu^*) = d^* \\ &= \min_x \mathcal{L}(x, \lambda^*, \nu^*) \\ &\leq \mathcal{L}(x^*, \lambda^*, \nu^*) \\ &= f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \sum_{j=1}^n \nu_j^* h_j(x^*) \\ &= f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) \\ &\leq f_0(x^*) \end{aligned}$$

In the fourth step, we can cancel the terms involving $h_j(x^*)$ because we know that the primal solution must satisfy $h_j(x^*) = 0$. In the fifth step, we know that $\lambda_i^* f_i(x^*) \leq 0$, because $\lambda_i^* \geq 0$ in order to satisfy the dual constraints, and $f_i(x^*) \leq 0$ in order to satisfy the primal constraints. Since we established that $f_0(x^*) = \min_x \mathcal{L}(x, \lambda^*, \mathbf{v}^*) \leq \mathcal{L}(x^*, \lambda^*, \mathbf{v}^*) \leq f_0(x^*)$, we know that all of the inequalities hold with equality and therefore $\mathcal{L}(x^*, \lambda^*, \mathbf{v}^*) = \min_x \mathcal{L}(x, \lambda^*, \mathbf{v}^*)$. This implies KKT condition 5 (stationarity), that

$$\nabla_x f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla_x f_i(x^*) + \sum_{j=1}^n \mathbf{v}_j^* \nabla_x h_j(x^*) = 0$$

Finally, note that due to the equality $f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) = f_0(x^*)$, we know that $\sum_{i=1}^m \lambda_i^* f_i(x^*) = 0$. This combined with the fact that $\forall i \lambda_i^* f_i(x^*) \leq 0$, establishes KKT condition 4 (complementary slackness):

$$\lambda_i^* f_i(x^*) = 0, \forall i \in [1, m]$$

□

The theorem above establishes that in the presence of strong duality, if the solutions are optimal, then they satisfy the KKT conditions. Let's prove a statement that is almost (but not quite) the converse, which will be much more helpful for solving optimization problems.

Theorem 2.2. If \bar{x} and $\bar{\lambda}, \bar{\mathbf{v}}$ satisfy the KKT conditions, and the primal problem is convex, then they are the optimal solutions to the primal and dual problems with zero duality gap.

Proof. If \bar{x} and $\bar{\lambda}, \bar{\mathbf{v}}$ satisfy KKT conditions 1, 2, 3 we know that they are at least feasible for the primal and dual problem. From the KKT stationarity condition we know that

$$\nabla_x f_0(\bar{x}) + \sum_{i=1}^m \bar{\lambda}_i \nabla_x f_i(\bar{x}) + \sum_{j=1}^n \bar{\mathbf{v}}_j \nabla_x h_j(\bar{x}) = 0$$

Since the primal problem is convex, we know that $\mathcal{L}(x, \lambda, \mathbf{v})$ is convex in x , and if the gradient of $\mathcal{L}(x, \bar{\lambda}, \bar{\mathbf{v}})$ at \bar{x} is 0, we know that

$$\bar{x} = \arg \min_x \mathcal{L}(x, \bar{\lambda}, \bar{\mathbf{v}}^*)$$

Therefore, we know that the optimal primal values for the primal problem optimize the inner optimization problem of the dual problem, and

$$g(\bar{\lambda}, \bar{\mathbf{v}}) = f_0(\bar{x}) + \sum_{i=1}^m \bar{\lambda}_i f_i(\bar{x}) + \sum_{j=1}^n \bar{\mathbf{v}}_j h_j(\bar{x})$$

By the primal feasibility conditions for $h_j(x)$ and the complementary slackness condition, we know that

$$g(\bar{\lambda}, \bar{\mathbf{v}}) = f_0(\bar{x})$$

Now, all we have to do is to prove that \bar{x} and $\bar{\lambda}, \bar{\mathbf{v}}$ are primal and dual optimal, respectively. Note that since weak duality always holds, we know that

$$p^* \geq d^* = \max_{\lambda \geq 0, \mathbf{v}} g(\lambda, \mathbf{v}) \geq g(\bar{\lambda}, \bar{\mathbf{v}}), \quad \forall \bar{\lambda} \geq 0, \bar{\mathbf{v}}$$

Since we know that $p^* \geq g(\lambda, \nu)$, we can also say that

$$f_0(x) - p^* \leq f_0(x) - g(\lambda, \nu)$$

And if we have that $f_0(\bar{x}) = g(\bar{\lambda}, \bar{\nu})$ as we deduced earlier, then

$$f_0(\bar{x}) - p^* \leq f_0(\bar{x}) - g(\bar{\lambda}, \bar{\nu}) = 0 \implies p^* \geq f_0(\bar{x})$$

Since p^* is the minimum value for the primal problem, we can go further by saying that $p^* \geq f_0(\bar{x})$ holds with equality and

$$p^* = f_0(\bar{x}) = g(\bar{\lambda}, \bar{\nu}) = d^*$$

Therefore, we have proven that \bar{x} and $\bar{\lambda}, \bar{\nu}$ are primal and dual optimal respectively, with zero duality gap. We eventually arrived at the conclusion that strong duality does indeed hold. \square

Let's pause for a second to understand what we've found so far. Given an optimization problem, its primal problem is an optimization problem over the primal variables, and its dual problem is an optimization problem over the dual variables. If strong duality holds, then we can solve the dual problem and arrive at the same optimal value. In order to solve the dual, we have to first solve the unconstrained inner optimization problem over the primal variables and then solve the constrained outer optimization problem over the dual variables. But how do we even know in the first place that strong duality holds? This is where KKT comes into play. If the primal problem is convex and the KKT conditions hold, we can solve for the dual variables easily and also verify strong duality does indeed hold. We shall do just that, in our discussion of SVMs.

3 The Dual of SVMs

Now, let's apply our knowledge of duality to find the dual of the soft-margin SVM optimization problem.

$$\underbrace{\min_{w,t,\xi} f(w,t,\xi)}_{\substack{\min_{w,t,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \\ (1 - \xi_i) - y_i(w^T x_i - t) \leq 0 \text{ and } -\xi_i \leq 0 \\ g(w,t,\xi) \leq 0}}$$

Let's identify the primal and dual variables for the SVM problem. We will have

- Primal variables w, t , and ξ_i
- Dual variables α_i corresponding to each constraint of the form $y_i(w^T x_i - t) \geq 1 - \xi_i$
- Dual variables β_i corresponding to each constraint of the form $\xi_i \geq 0$

For the purposes of notation, note that we are using α and β in place of λ , and there are no dual variables corresponding to v because there are no equality constraints. The lagrangian for the SVM problem is:

$$\begin{aligned}\mathcal{L}(w, t, \xi, \alpha, \beta) &= \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i + \sum_{i=1}^n \alpha_i ((1 - \xi_i) - y_i (w^T x_i - t)) + \sum_{i=1}^n \beta_i (-\xi_i) \\ &= \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i y_i (w^T x_i - t) + \sum_{i=1}^n \alpha_i + \sum_{i=1}^n (C - \alpha_i - \beta_i) \xi_i\end{aligned}\quad (7)$$

Thus, the dual is:

$$\max_{\alpha, \beta \geq 0} g(\alpha, \beta) = \min_{w, t, \xi} \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i y_i (w^T x_i - t) + \sum_{i=1}^n \alpha_i + \sum_{i=1}^n (C - \alpha_i - \beta_i) \xi_i \quad (8)$$

Let's use the KKT conditions to find the optimal dual variables. Verify that the primal problem is convex in the primal variables. We know that from the stationarity conditions, evaluated at the optimal dual values α^* and β^* , and the optimal primal values w^*, t^*, ξ_i^* :

$$\frac{\partial \mathcal{L}}{\partial w_i} = \frac{\partial \mathcal{L}}{\partial t} = \frac{\partial \mathcal{L}}{\partial \xi_i} = 0$$

- $\nabla_w \mathcal{L} = w^* - \sum_{i=1}^n \alpha_i^* y_i x_i = 0 \implies w^* = \sum_{i=1}^n \alpha_i^* y_i x_i$. This tells us that w^* is going to be a weighted combination of the positive-class x_i 's and negative-class x_i 's.
- $\frac{\partial \mathcal{L}}{\partial t} = \sum_{i=1}^n \alpha_i^* y_i = 0$. This tells us that the weights α_i^* will be equally distributed among positive- and negative- class training points.
- $\frac{\partial \mathcal{L}}{\partial \xi_i} = C - \alpha_i^* - \beta_i^* = 0 \implies 0 \leq \alpha_i^* \leq C$. This tells us that the weights α_i^* are restricted to being less than the hyperparameter C .

Verify that the other KKT also hold, establishing strong duality. Using these observations, we can eliminate some terms of the dual problem.

$$\begin{aligned}\mathcal{L}(w, t, \xi, \alpha^*, \beta^*) &= \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i^* y_i (w^T x_i - t) + \sum_{i=1}^n \alpha_i^* + \sum_{i=1}^n (C - \alpha_i^* - \beta_i^*) \xi_i \\ &= \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i^* y_i (w^T x_i) + \underbrace{t \sum_{i=1}^n \alpha_i^* y_i}_{=0} + \underbrace{\sum_{i=1}^n \alpha_i^* + \sum_{i=1}^n (C - \alpha_i^* - \beta_i^*) \xi_i}_{=0} \\ &= \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i^* y_i (w^T x_i) + \sum_{i=1}^n \alpha_i^*\end{aligned}$$

We can also rewrite all the optimal primal variables w^*, t^*, ξ_i^* in terms of the optimal dual variables

α_i^* :

$$\begin{aligned}
 g(\alpha^*, \beta^*) &= \min_{w, t, \xi} \mathcal{L}(w, t, \xi, \alpha^*, \beta^*) \\
 &= \mathcal{L}(w^*, t^*, \xi^*, \alpha^*, \beta^*) \\
 &= \frac{1}{2} \left\| \sum_{i=1}^n \alpha_i^* y_i x_i \right\|^2 - \sum_{i=1}^n \alpha_i^* y_i \left(\left(\sum_{j=1}^n \alpha_j^* y_j x_j \right)^T x_i \right) + \sum_{i=1}^n \alpha_i^* \\
 &= \frac{1}{2} \left\| \sum_{i=1}^n \alpha_i^* y_i x_i \right\|^2 - \sum_{i=1}^n (\alpha_i^* y_i x_i^T \left(\sum_{j=1}^n \alpha_j^* y_j x_j \right)) + \sum_{i=1}^n \alpha_i^* \\
 &= \alpha^{*T} \mathbf{1} - \frac{1}{2} \alpha^{*T} Q \alpha^*
 \end{aligned}$$

where $Q_{ij} = y_i(x_i^T x_j)y_j$ (and $Q = (\text{diag } y)XX^T(\text{diag } y)$).

Now, we can write the final form of the dual, which is only in terms of α and x and y (Note that we have eliminated all references to β):

$$\begin{aligned}
 \max_{\alpha} \quad & \alpha^T \mathbf{1} - \frac{1}{2} \alpha^T Q \alpha \\
 \text{s.t.} \quad & \sum_{i=1}^n \alpha_i y_i = 0 \\
 & 0 \leq \alpha_i \leq C \quad i = 1, \dots, n
 \end{aligned} \tag{9}$$

3.1 Geometric intuition

We've seen that the optimal value of the dual problem in terms of α is equivalent to the optimal value of the primal problem in terms of w , t , and ξ . But what do these dual values α_i even mean? That's a good question!

We know that the following KKT conditions are enforced:

- Stationarity

$$C - \alpha_i^* - \beta_i^* = 0$$

- Complementary slackness

$$\alpha_i^* \cdot ((1 - \xi_i^*) - y_i(w^{*T} x_i - t^*)) = 0$$

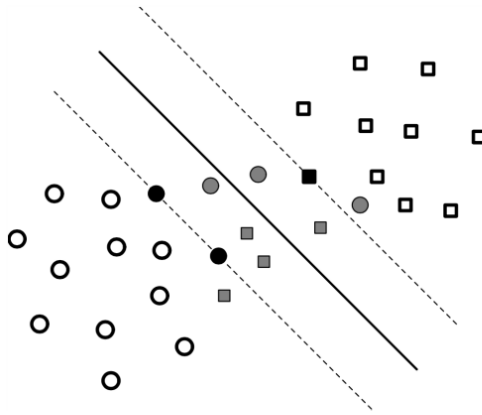
$$\beta_i^* \cdot \xi_i^* = 0$$

Here are some noteworthy relationships between α_i and the properties of the SVMs:

- Case 1: $\alpha_i^* = 0$. In this case, we know $\beta_i^* = C$, which is nonzero, and therefore $\xi_i^* = 0$. That is, if for point i we have that $\alpha_i^* = 0$ by the dual problem, then we know that there is no slack given to this point. Looking at the other complementary slackness condition, this makes sense because if $\alpha_i^* = 0$, then $y_i(w^{*T} x_i - t^*) - (1 - \xi_i^*)$ may be any value, and if we're minimizing the sum of our ξ_i 's, we should have $\xi_i^* = 0$. So, point i lies **on or outside the margin**.

- Case 2: α_i^* is nonzero. If this is the case, then we know $\beta_i^* = C - \alpha_i^* \geq 0$
 - Case 2.1: $\alpha_i^* = C$. If this is the case, then we know $\beta_i^* = 0$, and therefore ξ_i^* may be exactly 0 or nonzero. So, point i lies **on or inside the margin**.
 - Case 2.2: $0 < \alpha_i^* < C$. In this case, then β_i^* is nonzero and $\xi_i^* = 0$. But this is different from Case 1 because with α_i^* nonzero, we can divide by α_i^* in the complementary slackness condition and arrive at the fact that $1 - y_i(w^{*T}x_i - t^*) = 0 \implies y_i(w^{*T}x_i - t^*) = 1$, which means x_i lies exactly on the margin determined by w^* and t^* . So, point i lies **on the margin**. From this condition we can also conclude that

$$w^{*T}x_i - t^* = \frac{1}{y_i} = y_i \implies t^* = w^{*T}x_i - y_i$$



- $\alpha_i = 0 \Rightarrow y_i f(x_i) \geq 1$: on or outside the margin
- $0 < \alpha_i < C \Rightarrow y_i f(x_i) = 1$: on the margin
- $\alpha_i = C \Rightarrow y_i f(x_i) \leq 1$: on or inside the margin
- $\alpha_i = 0 \Leftarrow y_i f(x_i) > 1$: outside the margin
- $\alpha_i = C \Leftarrow y_i f(x_i) < 1$: inside the margin

Finally, let's reconstruct the optimal primal values w^*, t^*, ξ_i^* from the optimal dual values α^* . We know from the stationarity conditions that

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

From the cases that we discussed above, we also know that

$$\xi_i^* = \begin{cases} 1 - y_i(w^{*T}x_i - t^*) & \text{if } \alpha_i^* = C, \\ 0 & \text{otherwise} \end{cases}$$

Finally, from case 2.2 we know that

$$t^* = w^{*T}x_i - y_i \quad (\text{s.t. } 0 < \alpha_i^* < C)$$