Let us apply the necessary conditions for constrained optimality to the SVM objective

$$\min_{w,\xi_i} \frac{1}{2}\|w\|^2 + C\sum_i \xi_i$$

Lagrange multipliers

s.t. $\forall i$

$$1-\xi_i-y_i(w^T\phi(x_i)+w_0) \le 0 \qquad (\alpha_i)$$

$$-\xi_i \le 0 \qquad (\lambda_i)$$

$$\nabla f(w^*, \xi_i^*) + \sum_i \lambda_i \nabla g_i(w^*, \xi_i^*) = 0$$

gives

$$w + \sum_i \alpha_i(-y_i\phi(x_i)) = 0 \quad (1)$$

$\forall i$

$$C - \alpha_i - \lambda_i = 0 \quad (2)$$

$$\alpha_i(1-\xi_i-y_i(w^T\phi(x_i)+w_0)) = 0 \quad (3)$$

$$-\lambda_i\xi_i = 0 \quad (4)$$

$$\xi_i \geq 0 \quad (5)$$

$$y_i(w^T \phi(x_i) + w_0) \geq 1 - \xi_i \quad (6)$$

$\Rightarrow$ At point of optimality. [NECESSARY CONDITIONS]

$$w = \sum_i \alpha_i y_i \phi(x_i) \quad (1)$$

$$\alpha_i + \lambda_i = C \quad (2)$$

$$\alpha_i(1 - \xi_i - y_i(w^T \phi(x_i) + w_0)) = 0 \quad (3)$$

$$\xi_i \lambda_i = 0 \quad (4)$$

$$\xi_i \geq 0 \quad (5)$$

$$y_i(w^T \phi(x_i) + w_0) \geq 1 - \xi_i \quad (6)$$

Karush Kuhn Tucker [KKT] optimality conditions (necessary) for (w^*, ξ_i^*)

Questions to answer

① How to go from the original optimization problem & KKT conditions to a final solution?

↳ How abt substituting for ω & ξ_i in terms of λ_i's & α_i's

↳ In such a case, can you drop the original (primal) constraints altogether?

② What characteristics of the final solution does this system of equations give us?

$$\omega = \sum_i \alpha_i y_i \phi(x_i)$$

α_i → characterizes contribution of x_i to ω

If $\alpha_i = 0 \rightarrow$ no contribution

$$\Downarrow$$

$$\lambda_i = C \Rightarrow \xi_i = 0$$

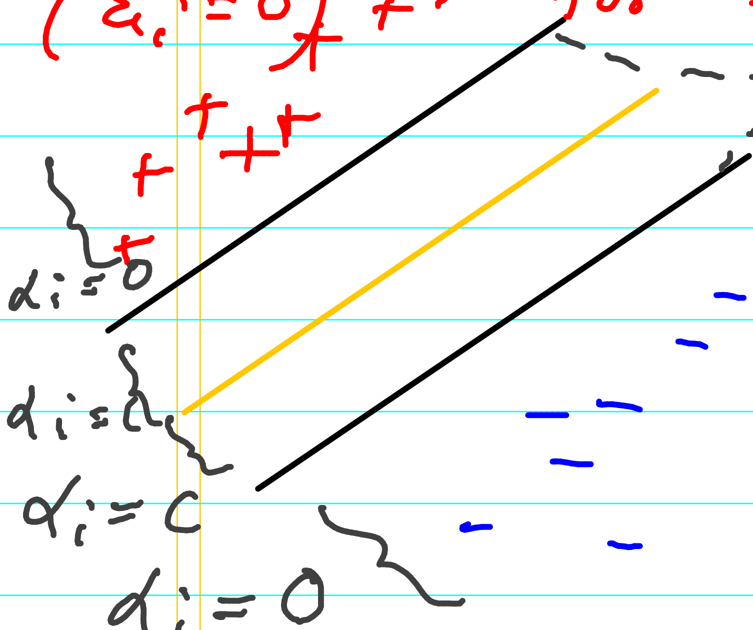
$$\therefore\ y_i(w^T \phi(x_i) + w_0) \geq 1$$

$\rightarrow$ pt x_i is on the desirable side of the margin

desirable side $(\xi_i = 0)$ for "+"

$\alpha_i \in [0, C]$

desirable side for '−' $(\xi_i = 0)$



If $y_i(w^T\phi(x_i) + w_0) > 1$

then $\alpha_i = 0$ & $\xi_i = 0$ & $\lambda_i = C$

(another side of looking at soln)

If $\alpha_i \in (0, C)$

then $\lambda_i \in (0, C) \Rightarrow \xi_i = 0$

& $y_i(w^T\phi(x_i) + w_0) = 1 - \xi_i = 1$

ie the pt must be on a supporting hyperplane

If $y_i(w^T\phi(x_i) + w_0) < 1$

then $\xi_i > 0$ & $\lambda_i = 0 \Rightarrow \alpha_i = C$

↳ If pt is on incorrect side, its α_i must be C

To answer Question ①, we need duality theory.

↳ Addresses the question,

$\boxed{\text{"What to do with inequalities in the primal, once we have KKT conditions \& we could substitute from them in the primal objective"}}$

Primal:

$\min f(x)$

s.t. $g_i(x) \leq 0 \quad (\alpha_i)$

$h_j(x) = 0 \quad (\lambda_i)$

Lagrange fn

$$f(x) + \sum_i g_i(x)\,\alpha_i + \sum_j h_j(x)\,\lambda_j$$

Define the dual function

$$\min_{x} \; f(x) + \sum_i g_i(x)\alpha_i + \sum_j h_j(x)\lambda_j$$

(copying from the notes at

http://www.cse.iitb.ac.in/~cs717/notes/classNotes/BasicsOfConvexOptimization.pdf

Section 4.4.2 (page 288 onwards)

[H/W: Understand the parts & their context pasted below, from the notes]

Consider the general inequality constrained minimization problem in (4.78), restated below.

$$\begin{aligned} \min_{\mathbf{x}\in\mathcal{D}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & g_i(\mathbf{x}) \le 0,\ i = 1, 2, \ldots, m \end{aligned} \qquad (4.80)$$

There are three simple and straightforward steps in forming a dual problem.

1. The first step involves forming the lagrange function by associating a price λ_i, called a lagrange multiplier, with the constraint involving g_i.

$$L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \sum_{i=1}^{n} \lambda_i g_i(\mathbf{x}) = f(\mathbf{x}) + \lambda^T \mathbf{g}(\mathbf{x})$$

2. The second step is the construction of the dual function $L^*(\lambda)$ which is defined as:

$$L^*(\lambda) = \min_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \lambda) = \min_{\mathbf{x} \in \mathcal{D}} f(\mathbf{x}) + \lambda^T \mathbf{g}(\mathbf{x})$$

   What makes the theory of duality constructive is when we can solve for L^* efficiently - either in a closed form or some other 'simple' mechanism. If L^* is not easy to evaluate, the duality theory will be less useful.

3. We finally define the dual problem:

$$\begin{array}{ll} \max\limits_{\lambda \in \Re^m} & L^*(\lambda) \\ \text{subject to} & \lambda \geq \mathbf{0} \end{array} \tag{4.81}$$

It can be immediatly proved that the dual problem is a concave maximization problem.

Theorem 80 *The dual function $L^*(\lambda)$ is concave.*

Theorem 81 *If $p^\star \in \Re$ is the solution to the primal problem in (4.80) and $d^\star \in \Re$ is the solution to the dual problem in (4.81), then*

$$p^\star \geq d^\star$$

In general, if $\widehat{\mathbf{x}}$ is any feasible solution to the primal problem (4.80) and $\widehat{\lambda}$ is a feasible solution to the dual problem (4.81), then

$$f(\widehat{\mathbf{x}}) \geq L^*(\widehat{\lambda})$$

The weak duality theorem has some important implications. If the primal problem is unbounded below, that is, $p^\star = -\infty$, we must have $d^\star = -\infty$, which means that the Lagrange dual problem is infeasible. Conversely, if the dual problem is unbounded above, that is, $d^\star = \infty$, we must have $p^\star = \infty$, which is equivalent to saying that the primal problem is infeasible. The difference, $p^\star - d^\star$ is called the duality gap.

In many hard combinatorial optimization problems with duality gaps, we get good dual solutions, which tell us that we are guaranteed of being some k % within the optimal solution to the primal, for some satisfactorily low values of k. This is one of the powerful uses of duality theory; constructing bounds for optimization problems.

Under what conditions can one assert that $d^\star = p^\star$? The condition $d^\star = p^\star$ is called *strong duality* and it does not hold in general. It usually holds for convex problems

Theorem 82 *If the function f is convex, g_i are convex and h_j are affine, then KKT conditions in 4.88 are necessary and sufficient conditions for zero duality gap.*