We have the SVM objective for classification.

$$\min_{\substack{w\in\mathbb{R}^d\\ b\in\mathbb{R}}} \frac{1}{2}\|w\|_2^2 + \underbrace{C\sum_i \left(1-y_i(\langle x_i,w\rangle+b)\right)_+}_{(*)}$$

Just like Lasso, without constraints the implicit perturbation function is uninformative, resulting in a trivial dual. So we introduce constraints. Note

$$\min_{\xi} C\sum_i \xi_i \iff (*)$$

$$\text{st } \xi_i \geq \max(0, 1-y_i(\langle x_i,w\rangle+b))$$

$$\iff \xi_i \geq 0 \wedge \xi_i \geq 1-y_i(\cdots)$$

Rewriting, we get the constrained soft-SVM objective.

$$\min_{w,\xi} \frac{1}{2}\|w\|_2^2 + C\sum_i \xi_i$$

$$\text{st} \quad \xi \geq 0$$

$$D(Xw + b) \geq 1 - \xi$$

Where $D = \operatorname{diag}(y)$

This objective is Strongly Dual by Slater: choose ξ suff large. We introduce duals ν, α.

$$\max_{\substack{\nu \geq 0 \\ \alpha \geq 0}} \inf_{w,\xi,b} \frac{1}{2}\|w\|_2^2 + C\mathbb{1}^T \xi$$

$$+ \alpha^T(1 - \xi - DXw - D\mathbb{1}b)$$

$$+ \nu^T \xi$$

Splitting the inf across the composite sum over b, ξ, w terms:

ξ term:

$$\inf_{\xi} \xi^T (c\mathbf{1} - \alpha - \nu)$$

$$= \begin{cases} 0 & \text{if } \alpha + \nu = c\mathbf{1} \\ -\infty & \text{o/w} \end{cases}$$

w term:

$$\inf_{w} \frac{1}{2}\|w\|_2^2 - w^T (X^T D \alpha)$$

$$= -\frac{1}{2}\|X^T D \alpha\|_2^2$$

by completing the square.

$$\inf_{b} -\alpha^T D \mathbf{1} b$$

$$= \begin{cases} 0 & \text{if } \alpha^T D \mathbf{1} = 0 \\ -\infty & \text{o/w} \end{cases}$$

Combining conditions we get an unsimplified dual:

$$\max_{\nu, \alpha} \quad -\frac{1}{2}\|X^T D \alpha\|_2^2 + 1^T \alpha$$

$$\text{st} \quad \nu \geq 0 \quad \alpha + \nu = c1$$
$$\alpha \geq 0 \quad \alpha^T D 1 = 0$$

Simplifying:

$$\max_{\alpha} \sum_i \alpha_i - \frac{1}{2} \sum_{ij} y_i y_j \alpha_i \alpha_j \langle x_i, x_j \rangle$$

$$0 \leq \alpha \leq c$$

$$\sum_i \alpha_i y_i = 0$$

Where we're of course obligated that only the Gram entries $\langle x_i, x_j \rangle$ are used from the X data, as Kernel Trick.