Lagrange Optimization
Minimize with constraints

$$\min_{w} f(w), \text{s.t. } g_i(w) \leq 0, \;\; i = 1, \ldots, k, \;\; \text{and } \; h_i(w) = 0, \;\; i = 1, \ldots, l$$

Lagrange function: $L_p(w, \alpha, \beta) = f(w) + \sum_{i=1}^{k} \alpha_i g_i(w) + \sum_{i=1}^{l} \beta_i h_i(w)$

SVM Optimization Function - Linearly Separable Case
Cost function: $\Phi(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w}$
Constraint: $y_n(\mathbf{w}^T\mathbf{x}_i + b) \geq 1$ for $i = 1, 2, ..., N$
Optimization Problem: $\min \frac{1}{2}\|\mathbf{w}\|_2^2$, s.t. $y_n(\mathbf{w}^T\mathbf{x} + w_0) \geq 1$, $n = 1, \ldots, N$

Step 1: Formulate Lagrangian function (primal problem)

$$L_p = \frac{1}{2}\|\mathbf{w}\|_2^2 - \sum_{n=1}^{N} \alpha_n \left[y_n(\mathbf{w}^T\mathbf{x}_n + w_0) - 1 \right]$$

(The "-" sign is because the sign of inequality in the constrain is now flipped)
Step 2: Set gradient of Lagrangian to 0 wrt primal variables $\mathbf{w}, w_0$

$$\frac{\partial L_p}{\partial \mathbf{w}} = 0 \;\; \Rightarrow \;\; \mathbf{w} = \sum_{n=1}^{N} \alpha_n y_n \mathbf{x_n}$$

$$\frac{\partial L_p}{\partial w_0} = 0 \;\; \Rightarrow \;\; \sum_{n=1}^{N} \alpha_n y_n = 0$$

Step 3: Substitute the primal variables $\mathbf{w}, w_0$ into the Lagrangian and express in terms of dual variables α_n

$$\begin{aligned} L_d &= \frac{1}{2}\|\mathbf{w}\|_2^2 - \mathbf{w}^T \sum_{n=1}^{N} \alpha_n y_n \mathbf{x_n} - w_0 \sum_{n=1}^{N} \alpha_n y_n + \sum_{n=1}^{N} \alpha_n \\ &= -\frac{1}{2}\mathbf{w}^T\mathbf{w} + \sum_{n=1}^{N} \alpha_n \\ &= -\frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{M} \alpha_n \alpha_m y_n y_m \mathbf{x_n}^T \mathbf{x_m} + \sum_{n=1}^{N} \alpha_n \end{aligned}$$

Step 4: Set gradient of Lagrangian to 0 wrt to dual variables (dual problem)

$$\max_{\alpha_n} L_d = \max_{\alpha_n} \left\{ -\frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m \mathbf{x_n}^T \mathbf{x_m} + \sum_{n=1}^{N} \alpha_n \right\}$$

$$\text{s.t.} \sum_{n=1}^{N} \alpha_n y_n = 0 \;\; \text{and} \;\; \alpha_n \geq 0, \;\; \text{for} \;\; n = 1, \ldots, N$$

Dual problem whose cost depends on data size N, rather than data dimensionality D. Most of the α_n will vanish. (See outline in slides for solving the problem) **Step 5:** Substitute the multipliers to the weight

SVM Optimization Function - Non-Linearly Separable Case
Cost function: $\min_{\mathbf{w},\boldsymbol{\xi}} \left[\frac{1}{2}\|\mathbf{w}\|_2^2 + C\sum_{n=1}^{N}\xi_n\right]$
Constraint: $y_n(\mathbf{w}^T\mathbf{x}+w_0) \geq 1-\xi_n$ and $\xi_n \geq 0$, $\forall n$

Step 1: Formulate Lagrangian function (primal problem)

$$L_p = \frac{1}{2}\|\mathbf{w}\|_2^2 + C\sum_{n=1}^{N}\xi_n - \sum_{n=1}^{N}\alpha_n\left[y_n(\mathbf{w}^T\mathbf{x}_n+w_0)-1+\xi_n\right] - \sum_{n=1}^{N}\mu_n\xi_n$$

Step 2: Set gradient of Lagrangian to 0 wrt primal variables $\mathbf{w}, w_0$

$$\frac{\partial L_p}{\partial \mathbf{w}} = 0 \quad \Rightarrow \quad \mathbf{w} = \sum_{n=1}^{N}\alpha_n y_n \mathbf{x_n}$$

$$\frac{\partial L_p}{\partial w_0} = 0 \quad \Rightarrow \quad \sum_{n=1}^{N}\alpha_n y_n = 0$$

$$\frac{\partial L_p}{\partial \xi_n} = 0 \quad \Rightarrow \quad C-\alpha_n-\mu_n = 0$$

Step 3: Substitute the primal variables $\mathbf{w}, w_0$ into the Lagrangian and express in terms of dual variables α_n

$$\begin{aligned}
L_d = \frac{1}{2}\|\mathbf{w}\|_2^2 + \underbrace{C\sum_{n=1}^{N}\xi_n}\; & -\mathbf{w}^T\sum_{n=1}^{N}\alpha_n y_n \mathbf{x_n} \\
& - w_0\sum_{n=1}^{N}\alpha_n y_n + \sum_{n=1}^{N}\alpha_n - \underbrace{\sum_{n=1}^{N}\alpha_n\xi_n} - \underbrace{\sum_{n=1}^{N}\mu_n\xi_n} \\
& = \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m y_n y_m \mathbf{x_n}^T\mathbf{x_m} + \sum_{n=1}^{N}\xi_n\left(C-\alpha_n-\mu_n\right) + \sum_{n=1}^{N}\alpha_n \\
& = \sum_{n=1}^{N}\alpha_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m y_n y_m \mathbf{x_n}^T\mathbf{x_m}
\end{aligned}$$

Step 4: Set gradient of Lagrangian to 0 wrt to dual variables (dual problem)

$$\max_{\alpha_n} L_d = \max_{\alpha_n}\left\{-\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\alpha_n\alpha_m y_n y_m \mathbf{x_n}^T\mathbf{x_m} + \sum_{n=1}^{N}\alpha_n\right\}$$

$$\text{s.t.} \sum_{n=1}^{N}\alpha_n y_n = 0 \;\text{ and }\; \alpha_n \geq 0, \;\text{ for }\; n=1,\ldots,N$$

Dual problem solved with approximate methods.
Step 5: Substitute the multipliers to the weight