Part I: Linear SVM Dual Form

Primal Form

$$\min \frac{1}{2}\|w\|^2 \quad \text{s.t. } (w^T x_i + b)y_i \geq 1 \quad \forall i$$

Penalty $\rightarrow$ Lagrange

This is how we'd turn the constraints into a form we could eventually solve for:

$$\min_{w,b} \frac{1}{2}\|w\|^2 + \text{penalty term}$$

For data $\{(x_i, y_i)\}$, use the penalty:

$$\begin{cases} 0 & \text{if } (w^T x_i + b)y_i \geq 1 \\ \infty & \text{otherwise} \end{cases} = \max_{\alpha_i \geq 0} \alpha_i [1 - (w^T x_i + b)y_i]$$

So the minimization becomes:

$$\min_{w,b} \left\{ \frac{1}{2}\|w\|^2 + \sum_{i=1}^{n} \max_{\alpha_i \geq 0} \alpha_i [1 - (w^T x_i + b)y_i] \right\}$$

Introducing Lagrange multipliers α_i:

$$\min_{w,b} \max_{\alpha_i \geq 0} \left\{ \frac{1}{2}\|w\|^2 + \sum_{i=1}^{n} \alpha_i [1 - (w^T x_i + b)y_i] \right\}$$

Swap min and max:

$$\max_{\alpha_i \geq 0} \min_{w,b} \left\{ \frac{1}{2}\|w\|^2 + \sum_{i=1}^{n} \alpha_i [1 - (w^T x_i + b)y_i] \right\}$$

$$= \max_{\alpha_i \geq 0} \min_{w,b} J(w, b; \alpha)$$

Exercise 1

Now take the derivative with J with respect to w and b:

$$\frac{\partial J}{\partial w} = w - \sum_{i=1}^{n} \alpha_i x_i y_i = 0$$

$$\frac{\partial J}{\partial b} = -\sum_{i=1}^{n} \alpha_i y_i = 0$$

Set the gradients to 0 for optimization:

$$\sum_{i=1}^{n} \alpha_i x_i y_i = 0 \rightarrow w = \sum_{i=1}^{n} \alpha_i x_i y_i$$

$$-\sum_{i=1}^{n} \alpha_i y_i = 0 \rightarrow \sum_{i=1}^{n} \alpha_i y_i = 0$$

Exercise 2

Substitute w back into J:

$$
\begin{align}
w = \sum_{i=1}^{n} \alpha_i x_i y_i, \quad \sum_{i=1}^{n} \alpha_i y_i = 0 \tag{1}\\
J(w, b, \alpha) &= \frac{1}{2}\|w\|^2 + \sum_i \alpha_i [1 - (w x_i + b) y_i] \tag{2}\\
&= \frac{1}{2}\|w\|^2 + \sum_i (\alpha_i - \alpha_i (w x_i + b) y_i) \tag{3}\\
&= \frac{1}{2}\|w\|^2 + \sum_i \alpha_i - \sum_i \alpha_i y_i (w x_i + b) \tag{4}\\
&= \frac{1}{2}\|w\|^2 + \sum_i \alpha_i - \sum_i \alpha_i y_i (w \cdot x_i) - \sum_i \alpha_i y_i b \tag{5}\\
&= \frac{1}{2}\|w\|^2 + \sum_i \alpha_i - \sum_i \alpha_i y_i (w \cdot x_i) \tag{6}\\
&= \frac{1}{2}\left\|\sum_i \alpha_i x_i y_i\right\|^2 + \sum_i \alpha_i - \sum_i \alpha_i y_i \left(\sum_j \alpha_j x_j y_j \cdot x_i\right) \tag{7}\\
&= \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) + \sum_i \alpha_i - \sum_{i,j} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \tag{8}\\
&= \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \tag{9}
\end{align}
$$

Part II: Kernel SVM

Note how in the dual:

$$L = \max_{\alpha_i \geq 0} \left\{ \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j}^{n} y_i y_j \alpha_i \alpha_j (x_i \cdot x_j) \right\}$$

The dot product can be projected into higher dimension:

$$(x_i \cdot x_j) \to \phi(x_i) \cdot \phi(x_j)$$

Instead of computing $\phi(x)$, use the kernel trick:

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$$

Exercise 3

If $k(a, b) = (a \cdot b)^3 = (a_1 b_1 + a_2 b_2)^3$, is this a valid kernel? Can it be written as $\phi(a) \cdot \phi(b)$?

$$(a_1 b_1 + a_2 b_2)^3 = a_1^3 b_1^3 + 3a_1^2 a_2 b_1^2 b_2 + 3a_1 a_2^2 b_1 b_2^2 + a_2^3 b_2^3$$

$$= (a_1^3, \sqrt{3}a_1^2 a_2, \sqrt{3}a_1 a_2^2, a_2^3) \cdot (b_1^3, \sqrt{3}b_1^2 b_2, \sqrt{3}b_1 b_2^2, b_2^3)$$

$$= \phi(a) \cdot \phi(b)$$