1. Computing the gradient directly

$$
\begin{aligned}
L =& \Big(- \sum_{i=2}^{m} \alpha_i (y^i y^1) + \sum_{i=2}^{m} \alpha_i \Big) \\
& - \frac{1}{2} \Bigg(\Big(- \sum_{i=2}^{m} \alpha_i (y^i y^1) \Big)^2 (x^1.x^1) + 2 \Big(- \sum_{i=2}^{m} \alpha_i (y^i y^1) \Big) y^1 \sum_{i=2}^{m} \alpha_i y^i (x^i.x^1) \\
& + \sum_{i=2}^{m} \sum_{j=2}^{m} \alpha_i y^i (x^i.x^j) y^j \alpha_j \Bigg) \\
=& - \sum_{i=2}^{m} \alpha_i (y^i y^1) + \sum_{i=2}^{m} \alpha_i \\
& - \frac{1}{2} \Bigg((x^1.x^1) \Big(\sum_{i=2}^{m} \alpha_i (y^i y^1) \Big)^2 - 2 y^1 \Big(\sum_{i=2}^{m} \alpha_i (y^i y^1) \Big) \sum_{i=2}^{m} \alpha_i y^i (x^i.x^1) \\
& + \sum_{i=2}^{m} \sum_{j=2}^{m} \alpha_i y^i (x^i.x^j) y^j \alpha_j \Bigg)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial L}{\partial \alpha_k} =& - (y^k y^1) + 1 \\
& - \frac{1}{2} \Bigg((x^1.x^1) 2 \Big(\sum_{i=2}^{m} \alpha_i (y^i y^1) \Big) (y^k y^1) \\
& - 2 y^1 \Big((y^k y^1) (\sum_{i=2}^{m} \alpha_i y^i (x^i.x^1)) + (\sum_{i=2}^{m} \alpha_i (y^i y^1)) (y^k (x^k.x^1)) \Big) \\
& + 2 (\sum_{i=2}^{m} y^k (x^k.x^i) y^i \alpha_i) \Bigg) \\
=& 1 - y^k y^1 - \Bigg((x^1.x^1) (\sum_{i=2}^{m} \alpha_i y^i) \\
& - \Big(\sum_{i=2}^{m} \alpha_i y^i (x^i.x^1 + x^k.x^1) \Big) \\
& + \sum_{i=2}^{m} (x^i.x^k) y^i \alpha_i \Bigg) y^k
\end{aligned}
$$

2. Another (easier) approch to compute the gradient

Notice that we have $\alpha_1 = -\sum_{i=2}^{m} \alpha_i (y^i y^1)$. We can consider we initially have $m-1$ parameter $\alpha_i, i = 2, ..., m$, and then we construct α_1 using α_i. Finally, we use all α's to compute the objective function.

Namely, we can consider $\alpha_i, i = 2, ..., m$ are input layer. The hidden layer are $m-1$ nodes directly representing α_i, and 1 more node representing α_1. In this case, we can use the idea of ***Backpropagation*** to compute the gradient of α_i:

$$(\frac{\partial L}{\partial \alpha_k})' = \frac{\partial L}{\partial \alpha_k} - y^k y^1 \frac{\partial L}{\partial \alpha_0}$$

Notice that $\frac{\partial L}{\partial \alpha k} = 1 - \sum_{i=1}^{m} (x^i . x^k) y^i \alpha_i y^k$. If we substitute α_1, we will have the exact same formula as the first approach.