Explain Soft margin SVM and solve it with steps as convex optimization.

Answer to Question

Problem Definition

Hard margin SVM can not solve some Linearly inseparable problems, so we introduce a slack variable ξ to tolerate classification errors as a constraint. This is exactly Soft margin SVM works for. We formalize the problem as follow:

$$min\frac{||w||_2^2}{2} + C\sum_{\rangle}\xi_i$$
$$s.t.\ y_i(w^Tx_i+b) \geq 1-\xi_i$$
$$\xi_i \geq 0, i = 1, \ldots, m$$

Since you introduce the slack parameter, a penalty item $C\sum_{\rangle}\xi_i$ is added, where C is a constant. In order to solve it, we follow the following steps:

Step 1: Apply Lagrangian function

$$\mathcal{L}(w, b, \alpha, \lambda, \xi) = \frac{||w||_2^2}{2} + C\sum_i \xi_i + \sum_i \alpha_i[1-\xi_i - y_i(w^Tx_i+b)] + \sum_i(-\lambda_i\xi_i)$$
$$s.t.\ \ \alpha_i \geq 0, \lambda_i \geq 0$$

Step 2: Solve for the variables using MLE

$$\max_{\alpha,\lambda}\min_{w,b,\xi}\mathcal{L}(w, b, \alpha, \lambda, \xi)$$

Solve w, b, ξ by setting derivatives with respect to those parameters as 0, that is

$$w^* = \sum_i \alpha_i y_i x_i$$
$$\sum_i -\alpha_i y_i = 0$$
$$C - \alpha_i - \lambda_i = 0$$

Step 3: Replace the derived variables into the original Lagrangian function

$$\mathcal{L}(\alpha, w^*, b^*, \xi^*) = \max(-\frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j x_i x_j + \sum_i \alpha_i)$$
$$s.t.\quad \sum_i \alpha_i y_i = 0,\ \ 0 <= \alpha_i <= C,\ \ \lambda_i >= 0$$

Step 4: Solve for variable for α_i using SMO algorithms

Here we do not give details. See the last paper summary for better understanding. In this way, we can obtain w, b, also ξ, λ, α. So, we get a optimal function to conduct classification. See SMO summary in the same folder.

Step 5: KKT conditions

For optimal solution, the Lagrange functions should meet the KKT conditions. That is

- Stationarity, $\frac{\partial \mathcal{L}}{\partial w} = \frac{\partial \mathcal{L}}{\partial b} = \frac{\partial \mathcal{L}}{\partial \xi_i} = 0$, yes
- Primal feasibility, $\frac{\partial \mathcal{L}}{\partial \alpha_i} \leq 0$, $\frac{\partial \mathcal{L}}{\partial \lambda_i} \leq 0$, yes

- Dual feasibility: $\alpha_i \geq 0, \lambda_i \geq 0$
- Complementary slackness

$$\alpha_i[1 - \xi - y_i(w^T x_i + b)] = 0$$
$$\lambda_i \xi = 0$$

Several parameter conditions for points in which area can be discussed,

- Points on boundary if $\alpha_i > 0, \lambda_i > 0, \xi_i = 0, 0 < \alpha_i < C$
- Points between upper bound and lower, $\xi_i > 0, \lambda_i = 0, \alpha_i = C$
- Points outside boundary, $\xi_i = 0, \lambda_i = C, \alpha_i = 0$

Here, may draw a graph and explain the reason for the parameters under different regions.