# 4033/5033 Assignment: Linear SVM

Sasank Sribhashyam

In this assignment, we will derive the solution for soft-margin linear SVM solution and an interpretation that support vectors are instances either (i) lying on the margin, (ii) falling within the margin or (iii) misclassified. Soft-margin LSVM finds a $\beta$ by solving the following problem

$$\min_{\beta,\beta_0,\epsilon} \frac{1}{2}||\beta||^2 + C\sum_{i=1}^{n} \epsilon_i, \quad \text{s.t. } y_i(x_i^T\beta + \beta_0) \geq 1 - \epsilon_i, \quad \epsilon_i \geq 0,\ i = 1, 2, \ldots, n, \tag{1}$$

where $C$ is a hyper-parameter.

Task 1. Please show that, by applying the Lagrange multiplier technique to solve problem (1),we can derive the following dual objective function (yes, it is the same as hard-margin LSVM).

$$L_D(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j y_i y_j x_i^T x_j. \tag{2}$$

## Primal Problem:

Minimize the following objective function:

$$\text{minimize} \quad \frac{1}{2}\|\beta\|^2 + C\sum_{i=1}^{n} \epsilon_i$$

subject to the constraints:

$$y_i(\mathbf{x}_i^T\beta + \beta_0) \geq 1 - \epsilon_i, \quad \epsilon_i \geq 0, \quad i = 1, 2, \ldots, n$$

## Lagrangian for the Primal Problem:

The Lagrangian for this problem is given by:

$$\begin{aligned}\mathcal{L}(\beta, \beta_0, \epsilon, \alpha, \mu) = \frac{1}{2}\|\beta\|^2 + C\sum_{i=1}^{n} \epsilon_i \\ - \sum_{i=1}^{n} \alpha_i \left(y_i(\mathbf{x}_i^T\beta + \beta_0) - (1 - \epsilon_i)\right) - \sum_{i=1}^{n} \mu_i\epsilon_i\end{aligned}$$

Here, $\alpha_i$ and $\mu_i$ are the Lagrange multipliers associated with the constraints $y_i(\mathbf{x}_i^T\beta + \beta_0) \geq 1 - \epsilon_i$ and $\epsilon_i \geq 0$, respectively.

## Partial Derivatives and Stationary Points:

Taking partial derivatives and setting them to zero, we get the following equations:

$$\frac{\partial \mathcal{L}}{\partial \beta} = 0 \quad \Rightarrow \quad \beta - \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i = 0$$

$$\frac{\partial \mathcal{L}}{\partial \beta_0} = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} \alpha_i y_i = 0$$

$$\frac{\partial \mathcal{L}}{\partial \epsilon_i} = 0 \quad \Rightarrow \quad C - \alpha_i - \mu_i = 0 \quad \Rightarrow \quad \alpha_i + \mu_i = C$$

## Dual Lagrangian Function:

Substituting these back into the Lagrangian, we get the dual Lagrangian function:

$$\mathcal{L}_D(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

## Dual Objective Function:

Finally, the dual objective function for the soft-margin linear SVM is given by:

$$L_D(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

This is the same as the dual objective function for the hard-margin linear SVM.

Task 2. List all KKT conditions for the critical point of (2) to be the optimal solution to problem (1). You may take reference for in lecture note, formula (24), but need to evaluate all functions and their derivatives.

## KKT Conditions for the Soft-Margin SVM Problem:

1. **Primal Feasibility:**

$$f_i(\beta, \beta_0, \epsilon) = 1 - \epsilon_i - y_i(\mathbf{x}_i^T \beta + \beta_0) \leq 0 \quad \text{for all} \quad i = 1, 2, \ldots, n$$

2. **Dual Feasibility:**

$$\lambda_i \geq 0 \quad \text{for all} \quad i = 1, 2, \ldots, n$$

3. **Complementary Slackness:**

$$\lambda_i f_i(\beta, \beta_0, \epsilon) = 0 \quad \text{for all} \quad i = 1, 2, \ldots, n$$

4. **Stationarity (First-Order Conditions):**

$$J'(\beta, \beta_0, \epsilon) + \sum_{i=1}^{n} \lambda_i f_i'(\beta, \beta_0, \epsilon) + \sum_{j=1}^{p} \nu_j h_j'(\beta, \beta_0, \epsilon) = 0$$

where:

- $J'(\beta, \beta_0, \epsilon)$ is the gradient of the objective function.
- $f_i'(\beta, \beta_0, \epsilon)$ is the gradient of the ith inequality constraint.
- $h_j'(\beta, \beta_0, \epsilon)$ is the gradient of the jth equality constraint.

5. **Equality Constraints:**

$$y_i(\mathbf{x}_i^T \beta + \beta_0) - (1 - \epsilon_i) = 0 \quad \text{for all} \quad i = 1, 2, \ldots, n$$

This condition corresponds to the equality constraints in the soft-margin SVM problem.

These conditions together are known as the Karush-Kuhn-Tucker (KKT) conditions and must be satisfied at an optimal solution to the constrained optimization problem. They reflect the balance between the primal and dual variables, feasibility of the solution, and the relationship between the gradients of the objective and constraint functions.

Task 3. Based on the KKT condition, explain why each of three types of instances (listed at top) will contribute to the optimal model.

## Analysis of Instances in Soft-Margin SVM Optimal Model:

1. **Instances lying on the margin (Support Vectors):**

- These instances satisfy the equation $y_i(\mathbf{x}_i^T \beta + \beta_0) = 1 - \epsilon_i$, where $\epsilon_i$ is the slack variable.
- Since $1 - \epsilon_i = 0$, the complementary slackness condition $\lambda_i \epsilon_i = 0$ implies that $\lambda_i > 0$ for these instances.
- The primal feasibility condition $1 - \epsilon_i - y_i(\mathbf{x}_i^T \beta + \beta_0) \leq 0$ is satisfied with equality $(= 0)$ for these support vectors.
- The gradient of the objective function with respect to these instances $(f_i')$ contributes to the stationarity condition, ensuring that the optimal solution respects the margin.

2. **Instances falling within the margin:**

- For these instances, $1 - \epsilon_i > 0$ because they lie within the margin.
- The complementary slackness condition $\lambda_i \epsilon_i = 0$ implies that $\lambda_i = 0$ for these instances, as $\epsilon_i$ is strictly positive.
- Since $\lambda_i = 0$, these instances do not contribute to the sum in the stationarity condition.
- However, they contribute to the objective function's minimization by introducing a penalty term $C\epsilon_i$ (where $C$ is the hyper-parameter). This term encourages the model to have a larger margin but allows for some misclassification.

3. **Misclassified instances:**

- For these instances, $1 - \epsilon_i < 0$ because they are misclassified and fall on the wrong side of the decision boundary.
- The complementary slackness condition $\lambda_i \epsilon_i = 0$ implies that $\lambda_i > 0$ for these instances, as $\epsilon_i$ is strictly positive.
- These instances contribute to the sum in the stationarity condition, as the gradient of the corresponding inequality constraint ($f_i'$) contributes to the optimization process.
- The penalty term $C\epsilon_i$ in the objective function encourages the minimization of misclassification.

In summary, support vectors lying on the margin have non-zero Lagrange multipliers, contributing directly to the optimal solution. Instances falling within the margin have zero Lagrange multipliers and contribute indirectly through the penalty term. Misclassified instances have non-zero Lagrange multipliers, and their gradients contribute to the optimization process. All three types of instances play a role in achieving the optimal soft-margin SVM model.