Project 79: mi-SPSVM

Zeglinski Jakub 279736

1 Problem

In Multiple Instance Learning (MIL), data is grouped into bags. The standard assumption:

- Bag is positive $\Leftrightarrow$ contains at least one positive instance
- Bag is negative $\Leftrightarrow$ all instances are negative

Goal: find hyperplane $H(v, \gamma) = \{x \in \mathbb{R}^n : v^T x = \gamma\}$ separating bags.

Notation: m, k = number of positive/negative bags; J_i^+, J_i^- = index sets of instances in bag i; J^+, J^- = current sets treated as positive/negative.

2 Theoretical Background

2.1 SVM and the Margin

For SVM, the margin between classes equals $2/\|v\|$. To maximize margin, we minimize $\|v\|^2$:

$$\min_{v,\gamma,\xi} \frac{1}{2}\|v\|^2 + C\sum_j \xi_j \quad \text{s.t.} \quad \xi_j \geq 1 - y_j(v^T x_j - \gamma),\ \xi_j \geq 0 \tag{1}$$

The hyperplanes $H^+ : v^T x = \gamma + 1$ and $H^- : v^T x = \gamma - 1$ are supporting hyperplanes.

2.2 From SVM to PSVM

Proximal SVM modifies standard SVM:

- Regularization: $\|v\|^2 \to \|(v, \gamma)\|^2$ (includes γ)
- Loss: L_1 (hinge) $\to$ L_2 (squared)
- Constraints: inequality $\to$ equality

The name "semi-proximal" means: L_2 loss with equality for J^+, but L_1 loss with inequality for J^-.

2.3 Quadratic Programming

Standard QP form:

$$\min_z \frac{1}{2} z^T P z + q^T z \quad \text{s.t.} \quad Gz \leq h \tag{2}$$

The matrix P (Hessian) must be symmetric positive semi-definite (PSD). For mi-SPSVM, P comes from the regularization term $\|(v, \gamma)\|^2$ plus the L_2 loss contribution from J^+.

3 Algorithm

3.1 Optimization Problem

For $C = 1$:

$$\boxed{\begin{aligned} \min_{v,\gamma,\xi} \quad & \frac{1}{2}\left\|\begin{pmatrix} v \\ \gamma \end{pmatrix}\right\|^2 + \frac{C}{2}\sum_{j\in J^+}\xi_j^2 + C\sum_{j\in J^-}\xi_j \\ \text{s.t.} \quad & \xi_j = 1 - (v^T x_j - \gamma), \quad j \in J^+ \\ & \xi_j \geq 1 + (v^T x_j - \gamma), \quad j \in J^- \\ & \xi_j \geq 0, \quad j \in J^- \end{aligned}} \tag{3}$$

3.2 QP Formulation

Variables: $z = [v; \gamma; \xi^-] \in \mathbb{R}^{n+1+|J^-|}$

For J^+: substitute $\xi_j = 1 - v^T x_j + \gamma$ into objective. With $A = [X^+, -\mathbf{1}]$:

$$\text{Hessian: } P = \begin{pmatrix} I_{n+1} + C \cdot A^T A & 0 \\ 0 & 0 \end{pmatrix} \tag{4}$$

For J^-: constraints $v^T x_j - \gamma - \xi_j \leq -1$ and $-\xi_j \leq 0$ form $Gz \leq h$.

3.3 Steps

Step 0: $J^+ \leftarrow$ all instances from positive bags, $J^- \leftarrow$ all from negative.

Step 1: Solve QP $\rightarrow$ get (v, γ).

Step 2: For each positive bag i, find witness: $j_i^* = \arg\max_{j \in J_i^+ \cap J^+}(v^T x_j - \gamma)$.
Compute $\bar{J} = \{j \in J^+ \setminus J^* : v^T x_j - \gamma \leq -1\}$. If $\bar{J} = \emptyset$, stop.

Step 3: $J^+ \leftarrow J^+ \setminus \bar{J}, \quad J^- \leftarrow J^- \cup \bar{J}$.

Step 4: Go to Step 1.

3.4 Convergence

The MIL problem is non-convex. This heuristic alternates between a convex QP (Step 1) and label updates (Step 3), finding a local optimum.

Algorithm terminates in finite iterations:

- Note: MIL is non-convex. This heuristic alternates QP (Step 1) and label updates (Step 2-3) to find a local optimum.
- $|J^+|$ decreases by at least 1 each iteration (when $\bar{J} \neq \emptyset$)
- $|J^+| \geq m$ (each bag keeps its witness)
- No cycles: removed instances never return

4 Results

Dataset: 35 instances, 7 bags (3 positive, 4 negative), 2 features.

Output: Converged at iteration 0.

$$v = \begin{pmatrix} -0.00856 \\ 0.00350 \end{pmatrix}, \quad \gamma = -0.6074 \tag{5}$$

Training Correctness:

	Bag 1	Bag 2	Bag 3	Bag 4	Bag 5	Bag 6	Bag 7
Pred	+1	+1	+1	−1	−1	−1	−1
True	+1	+1	+1	−1	−1	−1	−1

Accuracy = 100%

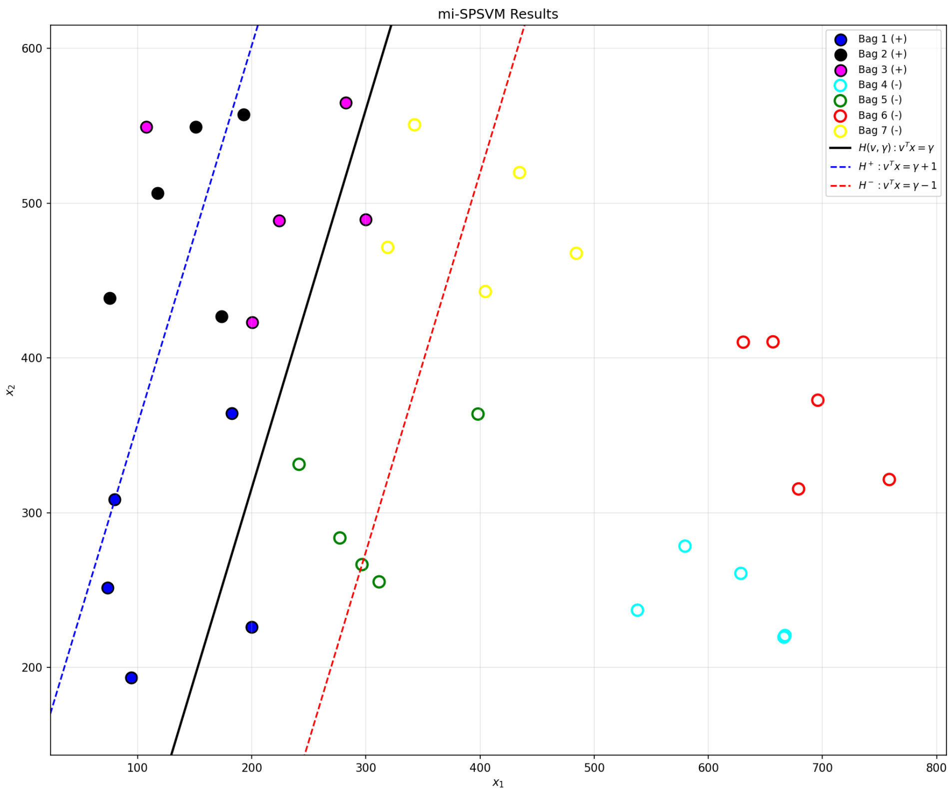

Figure 1: Filled = positive bags, unfilled = negative. H (black), H^+ (blue), H^- (red).

Prediction uses max-aggregation: bag is positive if $\max_j(v^T x_j - \gamma) > 0$.