EECS 127/227AT Optimization Models in Engineering UC Berkeley Spring 2026

Homework 14

1. Best Approximation in the Uniform Norm

Let $(x_1, y_1), \cdots, (x_n, y_n) \in \mathbb{R}^2$ be the given data points, and define vectors $\vec{x} = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}^\top$ and $\vec{y} = \begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix}^\top$.

(a) We want to find $a, b \in \mathbb{R}$ that minimizes $\left\|a\vec{x} + b\vec{1} - \vec{y}\right\|_\infty$, where $\vec{1}$ is an n-dimensional vector of ones. Formulate this problem as an LP.

Solution: $\vec{x}$ and $\vec{y}$ are given and the decision variables are a and b, so the problem can be formulated as follows:

$$\min_{a,b} \quad \left\|a\vec{x} + b\vec{1} - \vec{y}\right\|_\infty \tag{1}$$

$$= \min_{a,b} \quad \max_i |ax_i + b - y_i| \tag{2}$$

$$= \min_{a,b,t} \quad t \tag{3}$$

$$\text{s.t.} \quad t \geq \pm(ax_i + b - y_i), \forall i \tag{4}$$

(b) Now we want to find $a, b \in \mathbb{R}$ that minimizes $\left\|a\vec{x} + b\vec{1} - \vec{y}\right\|_1$, where $\vec{1}$ is an n-dimensional vector of ones. Formulate this problem as an LP.

Solution: The problem can be formulated as follows:

$$\min_{a,b} \quad \left\|a\vec{x} + b\vec{1} - \vec{y}\right\|_1 \tag{5}$$

$$= \min_{a,b} \quad \sum_i |ax_i + b - y_i| \tag{6}$$

$$= \min_{a,b,t} \quad \sum_i t_i \tag{7}$$

$$\text{s.t.} \quad t_i \geq \pm(ax_i + b - y_i), \forall i \tag{8}$$

2. Modified SVM

Let $C > 0$. Suppose we have labeled data $(\vec{x}_i, y_i) \in \mathbb{R}^d \times \{-1, 1\}$ for $i = 1, \ldots, n$. For each i, define $\vec{z}_i \doteq y_i \vec{x}_i$. Finally, define $Z \doteq \left[\vec{z}_1, \ldots, \vec{z}_n\right]^\top \in \mathbb{R}^{n \times d}$.

Recall that the soft-margin support vector machine problem can be expressed using slack variables as

$$p_1^\star = \min_{\vec{w}, \vec{s}} \quad \frac{1}{2} \|\vec{w}\|_2^2 + C \sum_{i=1}^n s_i \tag{9}$$
$$\text{s.t.} \quad s_i = \max\{0, 1 - \vec{z}_i^\top \vec{w}\}, \qquad \forall i \in \{1, \ldots, n\}.$$

In this problem we consider a modified SVM program with a squared penalty:

$$p_2^\star = \min_{\vec{w}, \vec{s}} \quad \frac{1}{2} \|\vec{w}\|_2^2 + \frac{C}{2} \sum_{i=1}^n s_i^2 \tag{10}$$
$$\text{s.t.} \quad s_i = \max\{0, 1 - \vec{z}_i^\top \vec{w}\}, \qquad \forall i \in \{1, \ldots, n\}.$$

We will use another representation of this program, namely one with affine constraints:

$$p^\star = \min_{\vec{w}, \vec{s}} \quad \frac{1}{2} \|\vec{w}\|_2^2 + \frac{C}{2} \|\vec{s}\|_2^2 \tag{11}$$
$$\text{s.t.} \quad \vec{s} \geq \vec{0}$$
$$\vec{s} \geq \vec{1} - Z\vec{w},$$

where the inequality constraints are componentwise (as usual).

(a) Choose the smallest class that problem (11) belongs to (LP/QP/SOCP/etc).

Solution: It is a QP – it has a quadratic objective and affine constraints.

(b) Prove that strong duality holds for (11).

Solution: The objective function is a convex quadratic and the constraints are affine (hence convex) in $\vec{w}$ and $\vec{s}$, so the problem is convex. Furthermore, there is a strictly feasible point – we can construct one by picking any $\vec{w}$ and then picking $\vec{s}$ whose components are large enough to fulfill the inequalities. This is always possible since there is no upper bound on the components of $\vec{s}$. Thus Slater's condition holds, so strong duality holds.

(c) Are the KKT conditions for problem (11) necessary, sufficient or both necessary and sufficient for global optimality?

Solution: The objective function is a convex quadratic and the constraints are affine (hence convex) in $\vec{w}$ and $\vec{s}$, so the problem is convex.

Since the problem is convex, all functions involved are continuously differentiable, and strong duality holds, the KKT conditions are both necessary and sufficient for optimality; that is, they are equivalent to optimality conditions.

(d) Let $\vec{\alpha}$ be the dual variable corresponding to the constraint $\vec{s} \geq \vec{0}$. What is the dimension (i.e., number of entries) of $\vec{\alpha}$?

Solution: $\vec{\alpha} \in \mathbb{R}^n$ since $\vec{s} \in \mathbb{R}^n$.

(e) Show that the Lagrangian $L(\vec{w}, \vec{s}, \vec{\alpha}, \vec{\beta})$ of problem (11), where $\vec{\alpha}$ is the dual variable corresponding to the constraint $\vec{s} \geq \vec{0}$, and $\vec{\beta}$ is the dual variable corresponding to the constraint $\vec{s} \geq \vec{1} - Z\vec{w}$, is equal to

$$L(\vec{w}, \vec{s}, \vec{\alpha}, \vec{\beta}) = \frac{1}{2}\|\vec{w}\|_2^2 + \frac{C}{2}\|\vec{s}\|_2^2 - \vec{s}^\top(\vec{\alpha} + \vec{\beta}) - \vec{w}^\top Z^\top \vec{\beta} + \vec{1}^\top \vec{\beta}. \tag{12}$$

Solution: We have

$$\begin{aligned} L(\vec{w}, \vec{s}, \vec{\alpha}, \vec{\beta}) &= \frac{1}{2}\|\vec{w}\|_2^2 + \frac{C}{2}\|\vec{s}\|_2^2 + \vec{\alpha}^\top(-\vec{s}) + \vec{\beta}^\top(\vec{1} - Z\vec{w} - \vec{s}) && (13) \\ &= \frac{1}{2}\|\vec{w}\|_2^2 + \frac{C}{2}\|\vec{s}\|_2^2 - \vec{s}^\top(\vec{\alpha} + \vec{\beta}) - \vec{w}^\top Z^\top \vec{\beta} + \vec{1}^\top \vec{\beta}. && (14) \end{aligned}$$

(f) Write the KKT conditions for problem (11). Show that if $(\vec{w}^\star, \vec{s}^\star, \vec{\alpha}^\star, \vec{\beta}^\star)$ obey the KKT conditions for problem (11), then

$$\vec{w}^\star = Z^\top \vec{\beta}^\star \qquad \text{and} \qquad \vec{s}^\star = \frac{\vec{\alpha}^\star + \vec{\beta}^\star}{C}. \tag{15}$$

HINT: For the first order/stationarity condition on the Lagrangian you will need to consider partial derivatives with respect to both $\vec{w}$ and $\vec{s}$.

Solution: Let $(\vec{w}^\star, \vec{s}^\star, \vec{\alpha}^\star, \vec{\beta}^\star)$ satisfy the KKT conditions. We have:

- Primal feasibility: $\vec{s}^\star \geq 0$ and $\vec{s}^\star \geq \vec{1} - Z\vec{w}^\star$.
- Dual feasibility: $\vec{\alpha}^\star \geq \vec{0}$, $\vec{\beta}^\star \geq \vec{0}$.
- Complementary slackness: $\alpha_i^\star s_i^\star = 0$ and $\beta_i^\star(1 - \vec{z}_i^\top \vec{w}^\star - s_i^\star) = 0$ for each i.
- Stationarity: $\nabla_{\vec{w}} L(\vec{w}^\star, \vec{s}^\star, \vec{\alpha}^\star, \vec{\beta}^\star) = \vec{0}$ and $\nabla_{\vec{s}} L(\vec{w}^\star, \vec{s}^\star, \vec{\alpha}^\star, \vec{\beta}^\star) = \vec{0}$. These become

$$\begin{aligned} \vec{0} &= \vec{w}^\star - Z^\top \vec{\beta}^\star && (16) \\ \vec{0} &= C\vec{s}^\star - (\vec{\alpha}^\star + \vec{\beta}^\star) && (17) \end{aligned}$$

which rearrange to the claimed equalities.

(g) Compute the dual function of problem (11) as

$$g(\vec{\alpha}, \vec{\beta}) \doteq L(\vec{w}^\star(\vec{\alpha}, \vec{\beta}), \vec{s}^\star(\vec{\alpha}, \vec{\beta}), \vec{\alpha}, \vec{\beta}) \tag{18}$$

where from the previous part we have that

$$\vec{w}^\star(\vec{\alpha}, \vec{\beta}) = Z^\top \vec{\beta} \qquad \text{and} \qquad \vec{s}^\star(\vec{\alpha}, \vec{\beta}) = \frac{\vec{\alpha} + \vec{\beta}}{C}. \tag{19}$$

Your final expression for $g(\vec{\alpha}, \vec{\beta})$ should not contain any maximizations, minimizations or terms including $\vec{w}$, $\vec{s}$, $\vec{w}^\star$, or $\vec{s}^\star$. It should only contain $\vec{\alpha}$, $\vec{\beta}$, C, Z, and numerical constants.

Solution: The dual function is

$$\begin{aligned} g(\vec{\alpha}, \vec{\beta}) &= L(\vec{w}^\star(\vec{\alpha}, \vec{\beta}), \vec{s}^\star(\vec{\alpha}, \vec{\beta}), \vec{\alpha}, \vec{\beta}) && (20) \\ &= \frac{1}{2}\left\|\vec{w}^\star(\vec{\alpha}, \vec{\beta})\right\|_2^2 + \frac{C}{2}\left\|\vec{s}^\star(\vec{\alpha}, \vec{\beta})\right\|_2^2 - \vec{s}^\star(\vec{\alpha}, \vec{\beta})^\top(\vec{\alpha} + \vec{\beta}) - \vec{w}^\star(\vec{\alpha}, \vec{\beta})^\top Z^\top \vec{\beta} + \vec{1}^\top \vec{\beta} && (21) \\ &= \frac{1}{2}\left\|Z^\top \vec{\beta}\right\|_2^2 + \frac{C}{2}\left\|\frac{\vec{\alpha} + \vec{\beta}}{C}\right\|_2^2 - \left(\frac{\vec{\alpha} + \vec{\beta}}{C}\right)^\top (\vec{\alpha} + \vec{\beta}) - \vec{\beta}^\top Z Z^\top \vec{\beta} + \vec{1}^\top \vec{\beta} && (22) \\ &= -\frac{1}{2}\vec{\beta}^\top Z Z^\top \vec{\beta} - \frac{1}{2C}\left\|\vec{\alpha} + \vec{\beta}\right\|_2^2 + \vec{1}^\top \vec{\beta}. && (23) \end{aligned}$$

(h) Let $\vec{\alpha}^\star$ and $\vec{\beta}^\star$ be optimal dual variables that solve the problem

$$d^\star \doteq \max_{\vec{\alpha},\vec{\beta}\geq\vec{0}} g(\vec{\alpha},\vec{\beta}). \tag{24}$$

It turns out that $\vec{\alpha}^\star$ can also be obtained by solving the quadratic program:

$$\begin{aligned} \min_{\vec{\alpha}} \quad & \left\|\vec{\alpha}+\vec{\beta}^\star\right\|_2^2 \\ \text{s.t.} \quad & \vec{\alpha} \geq \vec{0}. \end{aligned} \tag{25}$$

Solve this quadratic program (25) directly and find $\vec{\alpha}^\star$.

HINT: The duality or KKT approaches are not recommended. Consider $\vec{\alpha} = \begin{bmatrix}\alpha_1 & \cdots & \alpha_n\end{bmatrix}^\top$*, and use the components of* $\vec{\alpha}$ *to decompose the problem into* n *separate scalar problems. Solve each one by checking critical points; that is, points where the gradient is* 0*, the boundary of the feasible set, and* $\pm\infty$*.*

Solution: We have that

$$\left\|\vec{\alpha}+\vec{\beta}^\star\right\|_2^2 = \sum_{i=1}^{n}(\alpha_i+\beta_i^\star)^2. \tag{26}$$

Also, the $\vec{\alpha} \geq \vec{0}$ constraint is n separate constraints of the form $\alpha_i \geq 0$. Thus, we can solve for each α_i separately as

$$\alpha_i^\star \in \operatorname*{argmin}_{\alpha_i\geq 0}(\alpha_i+\beta_i^\star)^2. \tag{27}$$

This problem is convex and so we can solve it by checking the critical points.

- The gradient (w.r.t. α_i) is 0 if and only if $\alpha_i = -\beta_i^\star$. If $\beta_i^\star > 0$ then this solution is infeasible, and if $\beta_i^\star = 0$ then $\alpha_i = 0$.
- The constraint boundary is $\alpha_i = 0$; this solution is feasible with objective value $(\beta_i^\star)^2$.
- The limit $\alpha_i \to +\infty$ makes the objective value arbitrarily large, much larger than $(\beta_i^\star)^2$. The limit $\alpha_i \to -\infty$ makes the solution infeasible.

Thus the optimal solution for each scalar problem is $\alpha_i^\star = 0$. Thus $\vec{\alpha}^\star = \vec{0}$.

(i) Let $\beta^\star$ be a solution to the dual problem. Characterize the pairs $(\vec{x}_i, y_i)$ which are "support vectors", i.e., contribute to the optimal weight vector $\vec{w}^\star$, in terms of $\beta^\star$.

Solution: We have that $\vec{w}^\star = \sum_{i=1}^{n}\beta_i^\star\vec{z}_i$. If $\beta_i^\star > 0$ then $\vec{z}_i$ contributes to $\vec{w}^\star$, so $(\vec{x}_i, y_i)$ is a support vector. Otherwise $\beta_i^\star = 0$, then $\vec{z}_i$ does not contribute to $\vec{w}^\star$, so $(\vec{x}_i, y_i)$ is not a support vector.

3. **Soft-Margin SVM**

Consider the soft-margin SVM problem,

$$p^\star(C) = \min_{\vec{w}\in\mathbb{R}^m, b\in\mathbb{R}, \vec{\xi}\in\mathbb{R}^n} \quad \frac{1}{2}\|\vec{w}\|_2^2 + C\sum_{i=1}^n \xi_i \tag{28}$$

$$\text{s.t.} \quad 1-\xi_i - y_i(\vec{x}_i^\top \vec{w} - b) \leq 0, \quad i = 1,2,\ldots,n \tag{29}$$

$$-\xi_i \leq 0, \quad i = 1,2,\ldots,n, \tag{30}$$

where $\vec{x}_i \in \mathbb{R}^m$ refers to the i^{th} training data point, $y_i \in \{-1,1\}$ is its label, and $C \in \mathbb{R}_+$ (i.e. $C > 0$) is a hyperparameter. Let α_i denote the dual variable corresponding to the inequality $1-\xi_i - y_i(\vec{x}_i^\top \vec{w} - b) \leq 0$ and let β_i denote the dual variable corresponding to the inequality $-\xi_i \leq 0$. The Lagrangian is then given by

$$\mathcal{L}(\vec{w}, b, \vec{\xi}, \vec{\alpha}, \beta) = \frac{1}{2}\|\vec{w}\|_2^2 + C\sum_{i=1}^n \xi_i + \sum_{i=1}^n \alpha_i(1-\xi_i - y_i(\vec{x}_i^\top \vec{w} - b)) - \sum_{i=1}^n \beta_i\xi_i. \tag{31}$$

Suppose $\vec{w}^\star, b^\star, \vec{\xi}^\star, \vec{\alpha}^\star, \beta^\star$ satisfy the KKT conditions. Classify the following statements as true or false and justify your answers mathematically.

(a) Suppose the optimal solution $\vec{w}^\star, b^\star$ changes when the training point $\vec{x}_i$ is removed. Then originally, we necessarily have $y_i(\vec{x}_i^\top \vec{w}^\star - b^\star) = 1 - \xi_i^\star$.

Solution: True. Since optimal $\vec{w}^\star$ changes if we remove point $\vec{x}_i$ we have $\alpha_i^\star \neq 0$. By complementary slackness we have,

$$\alpha_i^\star(1-\xi_i^\star - y_i(\vec{x}_i^\top \vec{w}^\star - b^\star)) = 0, \tag{32}$$

which gives,

$$1-\xi_i^\star - y_i(\vec{x}_i^\top \vec{w}^\star - b^\star) = 0 \tag{33}$$

$$\implies y_i(\vec{x}_i^\top \vec{w}^\star - b^\star) = 1-\xi_i^\star. \tag{34}$$

(b) Suppose the optimal solution $\vec{w}^\star, b^\star$ changes when the training point $\vec{x}_i$ is removed. Then originally, we necessarily have $\alpha_i^\star > 0$.

Solution: True. Since optimal $\vec{w}^\star$ changes if we remove point $\vec{x}_i$ we have $\alpha_i^\star \neq 0$. Further by dual feasibility we have $\alpha_i^\star \geq 0$ which together gives $\alpha_i^\star > 0$.

(c) Suppose the data points are strictly linearly separable, i.e. there exist $\tilde{\vec{w}}$ and $\tilde{b}$ such that for all i,

$$y_i(\vec{x}_i^\top \tilde{\vec{w}} - \tilde{b}) > 0. \tag{35}$$

Then $p^\star(C) \to \infty$ as $C \to \infty$.

Solution: False. Since

$$y_i(\vec{x}_i^\top \tilde{\vec{w}} - \tilde{b}) > 0, \tag{36}$$

we have for sufficiently small $\epsilon > 0$,

$$y_i(\vec{x}_i^\top \tilde{\vec{w}} - \tilde{b}) \geq \epsilon \implies y_i\left(\vec{x}_i^\top \frac{\tilde{\vec{w}}}{\epsilon} - \frac{\tilde{b}}{\epsilon}\right) \geq 1. \tag{37}$$

Thus, $\bar{\vec{w}} = \frac{\tilde{\vec{w}}}{\epsilon}, \bar{b} = \frac{\tilde{b}}{\epsilon}, \bar{\vec{\xi}} = 0$ is a feasible point with objective value $\frac{1}{2}\left\|\bar{\vec{w}}\right\|_2^2 < \infty$ irrespective of value of C.

4. Support Vector Machine Concepts

Recall the maximum margin support vector machine problem:

$$\begin{aligned} \min_{\vec{w}\in\mathbb{R}^k, b\in\mathbb{R}} \quad & \frac{1}{2}\|\vec{w}\|_2^2 \\ \text{s.t.} \quad & y_i(\vec{w}^\top \vec{x}_i + b) \geq 1 \quad \forall i \in \{1,\dots,n\}, \end{aligned}$$

where the data points $(\vec{x}_i, y_i)$, with features $\vec{x}_i \in \mathbb{R}^k$ and labels $y_i \in \{+1, -1\}$ for $i \in \{1,\dots,n\}$, are given.

(a) Consider the pairs of features $\vec{x}_i \in \mathbb{R}^2$ and labels $y_i \in \{+1,-1\}$ given in Figure 1. The maximum margin hyperplane for this data along with the support vectors are depicted in Figure 2. Find the vector $\vec{w}$ and scalar b that solve this problem. *HINT: Note that the constraints in the maximum margin support vector machine problem must be satisfied with equality at the support vectors.*

Index i	Features $(x_{i1}, x_{i2}) \in \mathbb{R}^2$	Label $y_i \in \{+1,-1\}$
1	$(1,1)$	$+1$
2	$(3,4)$	$+1$
3	$(3,5)$	$+1$
4	$(4,0)$	-1
5	$(5,1)$	-1
6	$(6,6)$	-1

Figure 1: Data points and their labels

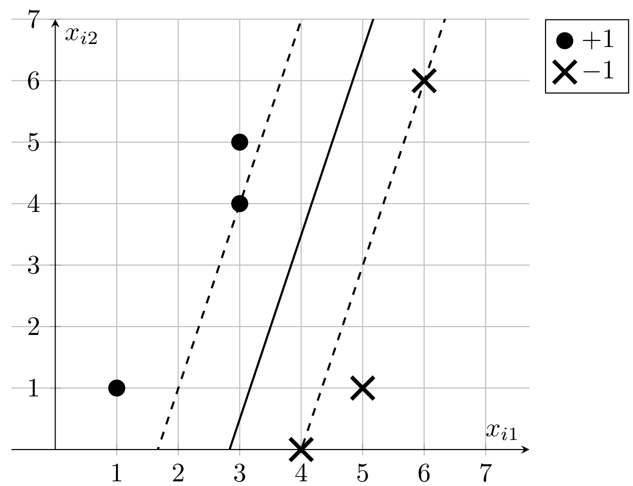

Figure 2: Maximum margin hyperplane and support vectors

HINT: You are likely to find at least one of these two calculations to be useful:

$$\begin{bmatrix} 3 & 4 & 1 \\ 4 & 0 & 1 \\ 6 & 6 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} -3/7 & 1/7 & 2/7 \\ 1/7 & -3/14 & 1/14 \\ 12/7 & 3/7 & -8/7 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 1 & 1 \\ 3 & 5 & 1 \\ 5 & 1 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} -1/4 & 0 & 1/4 \\ -1/8 & 1/4 & -1/8 \\ 11/8 & -1/4 & -1/8 \end{bmatrix}.$$

Solution: Using the hint, we note that the constraints in the maximum margin support vector machine problem are satisfied with equality at the support vectors. The support vectors are given in Figure 2 by $(3,4)^\top$, which is classified as $+1$, and $(4,0)^\top$, $(6,6)^\top$, which are classified as -1. This gives rise to the following equations in terms of the variables w, b:

$$\begin{aligned} 1((3,4)^\top w + b) &= 1, \\ -1((4,0)^\top w + b) &= 1, \\ -1((6,6)^\top w + b) &= 1. \end{aligned}$$

Putting these equations in matrix form gives us:

$$\begin{bmatrix} 3 & 4 & 1 \\ 4 & 0 & 1 \\ 6 & 6 & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}.$$

The inverse of the matrix on the left hand side is provided to us in the hint, which gives the solution:

$$\begin{bmatrix} w_1 \\ w_2 \\ b \end{bmatrix} = \begin{bmatrix} 3 & 4 & 1 \\ 4 & 0 & 1 \\ 6 & 6 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix} = \begin{bmatrix} -3/7 & 1/7 & 2/7 \\ 1/7 & -3/14 & 1/14 \\ 12/7 & 3/7 & -8/7 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}.$$

This gives $w_1^* = \frac{-6}{7}$, $w_2^* = \frac{2}{7}$, $b^* = \frac{17}{7}$.

(b) Now, consider the pairs of features $\vec{x}_i \in \mathbb{R}^2$ and labels $y_i \in \{+1, -1\}$ given in Figure 3, and depicted visually in Figure 4:

Index i	Features $(x_{i1}, x_{i2}) \in \mathbb{R}^2$	Label $y_i \in \{+1, -1\}$
1	$(1, 1)$	$+1$
2	$(4.5, 1)$	$+1$
3	$(4, 6)$	$+1$
4	$(4, 0)$	-1
5	$(4, 2)$	-1
6	$(5, 1)$	-1

Figure 3: Data points and their labels

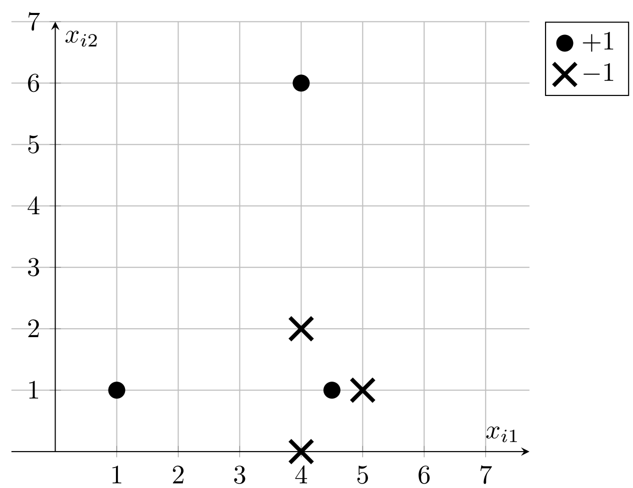

Figure 4: Visual depiction of data points and labels

If possible, find a separating hyperplane that solves the maximum margin support vector machine problem with this data, or provide a justification why such a hyperplane cannot be found.

Solution: Such a hyperplane cannot be found because the data are not linearly separable. This is because the point $(4.5, 1)$, which is classified as $+1$, can be written as a convex combination of the points classified as -1:

$$\begin{bmatrix} 4.5 \\ 1 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 4 \\ 0 \end{bmatrix} + \frac{1}{4} \begin{bmatrix} 4 \\ 2 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 5 \\ 1 \end{bmatrix}.$$

5. Homework Process

With whom did you work on this homework? List the names and SIDs of your group members.

NOTE: If you didn't work with anyone, you can put "none" as your answer.

