Homework 14

This homework is due at 11 PM on May 8th, 2026.

Submission Format: Your homework submission should consist of a single PDF file that contains all of your answers (any handwritten answers should be scanned).

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.

   (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.

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 $Z \doteq \begin{bmatrix} \vec{z}_1, \ldots, \vec{z}_n \end{bmatrix}^\top \in \mathbb{R}^{n \times d}$.

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

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

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

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

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

$$\begin{aligned} p^\star = \min_{\vec{w}, \vec{s}} \quad & \frac{1}{2} \|\vec{w}\|_2^2 + \frac{C}{2} \|\vec{s}\|_2^2 \\ \text{s.t.} \quad & \vec{s} \geq \vec{0} \\ & \vec{s} \geq \vec{1} - Z\vec{w}, \end{aligned} \tag{3}$$

where the inequality constraints are componentwise (as usual).

(a) Choose the smallest class that problem (3) belongs to (LP/QP/SOCP/etc).

(b) Prove that strong duality holds for (3).

(c) Are the KKT conditions for problem (3) necessary, sufficient or both necessary and sufficient for global optimality?

(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 $\vec{\alpha}$?

(e) Show that the Lagrangian $L(\vec{w}, \vec{s}, \vec{\alpha}, \vec{\beta})$ of problem (3), 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

$$L(\vec{w}, \vec{s}, \vec{\alpha}, \vec{\beta}) = \frac{1}{2} \|\vec{w}\|_2^2 + \frac{C}{2} \|\vec{s}\|_2^2 - \vec{s}^\top (\vec{\alpha} + \vec{\beta}) - \vec{w}^\top Z^\top \vec{\beta} + \vec{1}^\top \vec{\beta}. \tag{4}$$

(f) Write the KKT conditions for problem (3). Show that if $(\vec{w}^\star, \vec{s}^\star, \vec{\alpha}^\star, \vec{\beta}^\star)$ obey the KKT conditions for problem (3), then

$$\vec{w}^\star = Z^\top \vec{\beta}^\star \qquad \text{and} \qquad \vec{s}^\star = \frac{\vec{\alpha}^\star + \vec{\beta}^\star}{C}. \tag{5}$$

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

(g) Compute the dual function of problem (3) as

$$g(\vec{\alpha}, \vec{\beta}) \doteq L(\vec{w}^\star(\vec{\alpha}, \vec{\beta}), \vec{s}^\star(\vec{\alpha}, \vec{\beta}), \vec{\alpha}, \vec{\beta}) \tag{6}$$

where from the previous part we have that

$$\vec{w}^\star(\vec{\alpha}, \vec{\beta}) = Z^\top \vec{\beta} \qquad \text{and} \qquad \vec{s}^\star(\vec{\alpha}, \vec{\beta}) = \frac{\vec{\alpha} + \vec{\beta}}{C}. \tag{7}$$

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.

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

$$d^\star \doteq \max_{\vec{\alpha}, \vec{\beta} \geq \vec{0}} g(\vec{\alpha}, \vec{\beta}). \tag{8}$$

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

$$\begin{aligned} \min_{\vec{\alpha}} \quad & \left\| \vec{\alpha} + \vec{\beta}^\star \right\|_2^2 \\ \text{s.t.} \quad & \vec{\alpha} \geq \vec{0}. \end{aligned} \tag{9}$$

Solve this quadratic program (9) 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* $\pm\infty$*.*

(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 $\beta^\star$.

3. Soft-Margin SVM

Consider the soft-margin SVM problem,

$$p^\star(C) = \min_{\vec{w}\in\mathbb{R}^m, b\in\mathbb{R}, \vec{\xi}\in\mathbb{R}^n} \quad \frac{1}{2}\left\|\vec{w}\right\|_2^2 + C\sum_{i=1}^{n}\xi_i \tag{10}$$

$$\text{s.t.} \quad 1-\xi_i - y_i(\vec{x}_i^\top \vec{w} - b) \le 0, \quad i = 1,2,\dots,n \tag{11}$$

$$-\xi_i \le 0, \quad i = 1,2,\dots,n, \tag{12}$$

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 $1-\xi_i - y_i(\vec{x}_i^\top \vec{w} - b) \le 0$ and let β_i denote the dual variable corresponding to the inequality $-\xi_i \le 0$. The Lagrangian is then given by

$$\mathcal{L}(\vec{w}, b, \vec{\xi}, \vec{\alpha}, \beta) = \frac{1}{2}\left\|\vec{w}\right\|_2^2 + C\sum_{i=1}^{n}\xi_i + \sum_{i=1}^{n}\alpha_i(1-\xi_i - y_i(\vec{x}_i^\top \vec{w} - b)) - \sum_{i=1}^{n}\beta_i\xi_i. \tag{13}$$

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

(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 $\alpha_i^\star > 0$.

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

$$y_i(\vec{x}_i^\top \vec{\widetilde{w}} - \widetilde{b}) > 0. \tag{14}$$

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

4. Support Vector Machine Concepts

Recall the maximum margin support vector machine problem:

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

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 $i \in \{1, \ldots, n\}$, 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

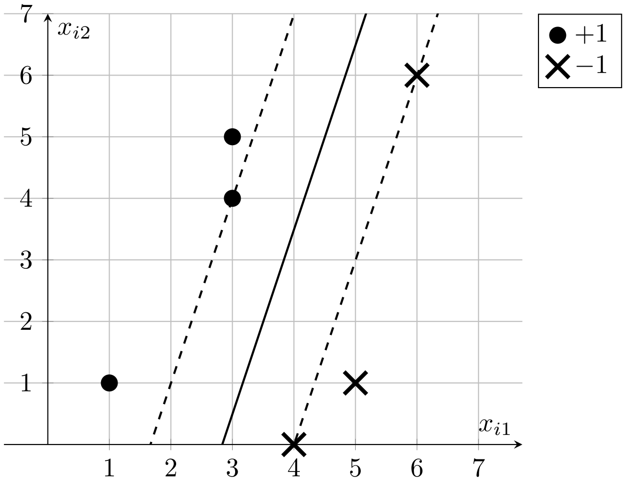

Figure 2: Maximum margin hyperplane and support vectors

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

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

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

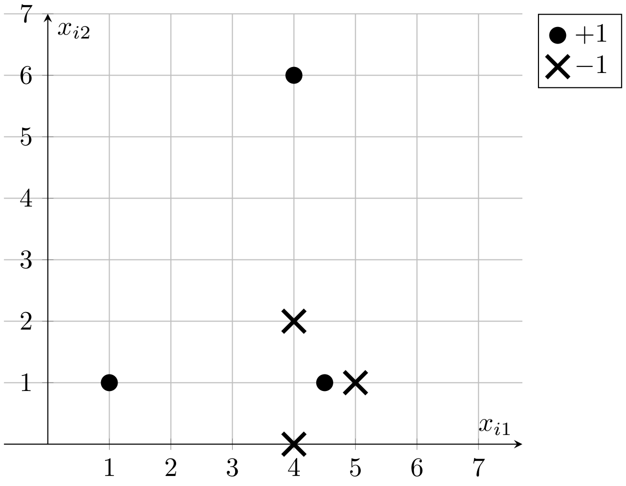

Figure 4: Visual depiction of data points and labels

(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:

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.

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.