Discussion 13

We might not cover all the problems on the worksheet, as discussion worksheets are not always designed to be finished within an hour. However, this is totally fine since they are deliberately made slightly longer so they can serve as resources you can use to practice, reinforce, and build upon concepts covered in lectures and homework.

1. Can we Use Slack Variables?

So far, we've presented slack variables as a method of converting optimization problems to a desired form, and it may seem like we can always use them. In this question, we take a more nuanced look at when slack variables are helpful and when they are not. For each of the following functions, consider the unconstrained optimization problem

$$p_j^* = \min_{\vec{x}\in\mathbb{R}^n} f_j(\vec{x}) \tag{1}$$

If possible, reformulate each problem into an LP/convex QP/SOCP using slack variables. If not possible, explain why.

(a) $f_1(\vec{w}) = \sum_{i=1}^{n} \left(\max\{0, 1 - y_i\vec{x}_i^\top\vec{w}\}\right)^2 + C\|\vec{w}\|_2^2$ for $C > 0$ and for some given vectors $\vec{x}_i \in \mathbb{R}^d$ for $i = 1, \ldots, n$ and $\vec{y} \in \mathbb{R}^n$ and variable $\vec{w} \in \mathbb{R}^d$.

(b) $f_2(\vec{x}) = \|A\vec{x} - \vec{y}\|_2 - \|\vec{x}\|_1$.

2. 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

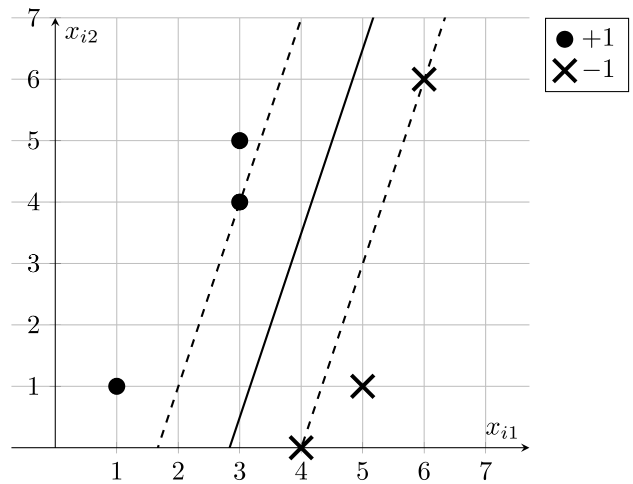

Figure 2: Maximum margin hyperplane and support vectors

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

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

(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.

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

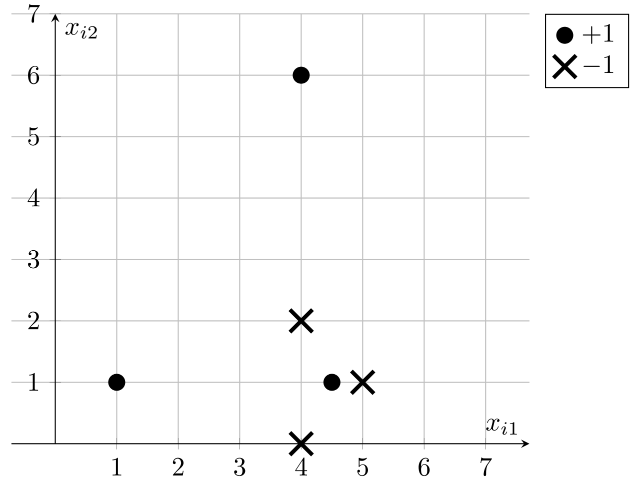

Figure 4: Visual depiction of data points and labels

3. Soft-Margin SVM

Consider the soft-margin SVM problem,

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

$$\text{s.t.} \quad 1 - \xi_i - y_i(\vec{x}_i^\top \vec{w} - b) \leq 0, \quad i = 1, 2, \ldots, n \tag{3}$$

$$-\xi_i \leq 0, \quad i = 1, 2, \ldots, n, \tag{4}$$

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

$$\mathcal{L}(\vec{w}, b, \vec{\xi}, \vec{\alpha}, \beta) = \frac{1}{2}\|\vec{w}\|_2^2 + C\sum_{i=1}^n \xi_i + \sum_{i=1}^n \alpha_i(1 - \xi_i - y_i(\vec{x}_i^\top \vec{w} - b)) - \sum_{i=1}^n \beta_i \xi_i. \tag{5}$$

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 $\widetilde{\vec{w}}$ and $\widetilde{b}$ such that for all i,

$$y_i(\vec{x}_i^\top \widetilde{\vec{w}} - \widetilde{b}) > 0. \tag{6}$$

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