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 $\vec{w} \in \mathbb{R}^d$.

Solution: We introduce slack variables ζ_i for $i = 1, \ldots, n$ corresponding to the max term. This leads to rewriting the miimization of f_1 as

$$\begin{aligned} \min \sum_{i=1}^{n} \zeta_i^2 + C\|\vec{w}\|_2^2 & \\ \text{s.t.} \quad & \zeta_i \geq 0 \\ & \zeta_i \geq 1 - y_i\vec{x}_i^\top\vec{w} \quad i = 1, \ldots, n \end{aligned}$$

This minimization probem is clearly a convex QP.

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

Solution: While this problem appears almost identical to the one in part (a), it **cannot be written as an SOCP**; in fact, it is not convex. To see this, we can attempt to follow the same procedure as above. First, we introduce our new variables, to generate equivalent problem

$$p_2^* = \min_{\vec{x},\vec{t}\in\mathbb{R}^n, s\in\mathbb{R}} \quad s - \sum_{i=1}^{n} t_i \tag{2}$$

$$\text{s.t.} \quad \|A\vec{x} - \vec{y}\|_2 = s \tag{3}$$

$$|x_i| = t_i, \quad i = 1, \ldots, n. \tag{4}$$

We then relax our equality constraints to inequalities; again, we can perform this step successfully, but we must be cautious, as the sign on each value t_i in the objective function has changed. If we require that $|x_i| \leq t_i$ as before, then our optimization problem is unbounded below, since we can increase t_i arbitrarily high to generate an arbitrarily low value. Instead, we must require that $|x_i| \geq t_i$, giving us optimization problem

$$p_2^* = \min_{\vec{x},\vec{t}\in\mathbb{R}^n, s\in\mathbb{R}} \quad s - \sum_{i=1}^{n} t_i \tag{5}$$

$$\text{s.t.} \quad \|A\vec{x} - \vec{y}\|_2 \leq s \tag{6}$$

$$|x_i| \geq t_i, \quad i = 1, \ldots, n. \tag{7}$$

While this is a valid (though nonconvex) optimization problem, each of the final constraints $|x_i| \geq t_i$ is not convex:

We therefore cannot linearize them to get an SOCP.

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.

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

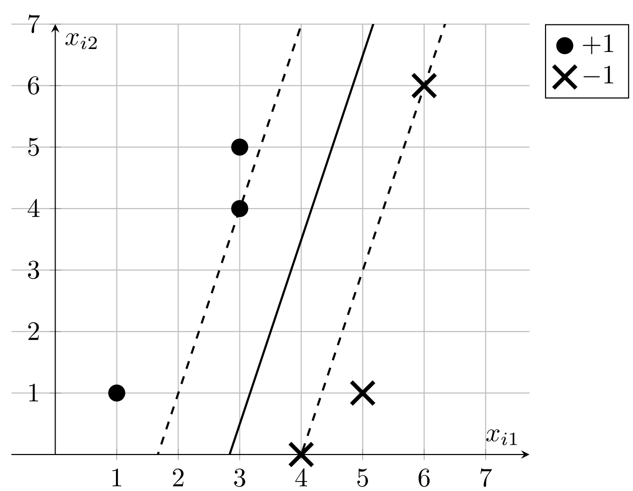

Figure 2: Maximum margin hyperplane and support vectors

HINT: Note that the constraints in the maximum margin support vector machine problem must be satisfied with equality at the 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

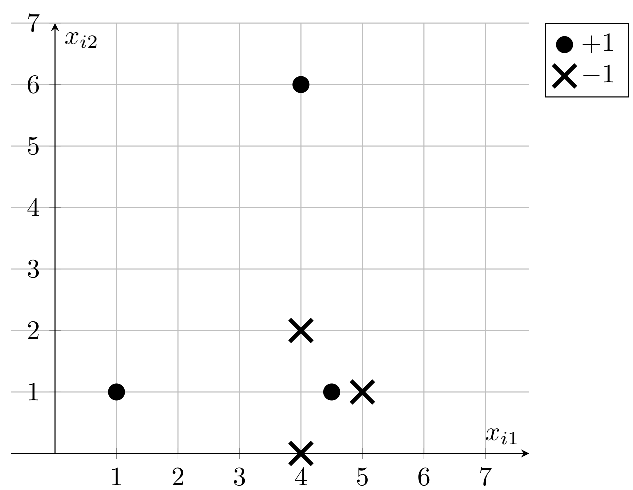

Figure 4: Visual depiction of data points and labels

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:

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

3. Soft-Margin SVM

Consider the soft-margin SVM problem,

$$p^\star(C) = \min_{\vec{w}\in\mathbb{R}^m, b\in\mathbb{R}, \vec{\xi}\in\mathbb{R}^n} \quad \frac{1}{2}\|\vec{w}\|_2^2 + C\sum_{i=1}^n \xi_i \tag{8}$$

$$\text{s.t.} \quad 1-\xi_i - y_i(\vec{x}_i^\top \vec{w} - b) \le 0, \quad i = 1,2,\ldots,n \tag{9}$$

$$-\xi_i \le 0, \quad i = 1,2,\ldots,n, \tag{10}$$

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

$$\mathcal{L}(\vec{w}, b, \vec{\xi}, \vec{\alpha}, \beta) = \frac{1}{2}\|\vec{w}\|_2^2 + C\sum_{i=1}^n \xi_i + \sum_{i=1}^n \alpha_i(1-\xi_i - y_i(\vec{x}_i^\top \vec{w} - b)) - \sum_{i=1}^n \beta_i\xi_i. \tag{11}$$

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,

$$\alpha_i^\star(1-\xi_i^\star - y_i(\vec{x}_i^\top \vec{w}^\star - b^\star)) = 0, \tag{12}$$

which gives,

$$1-\xi_i^\star - y_i(\vec{x}_i^\top \vec{w}^\star - b^\star) = 0 \tag{13}$$

$$\implies y_i(\vec{x}_i^\top \vec{w}^\star - b^\star) = 1 - \xi_i^\star. \tag{14}$$

(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 $\alpha_i^\star \ge 0$ which together gives $\alpha_i^\star > 0$.

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

$$y_i(\vec{x}_i^\top \widetilde{\vec{w}} - \widetilde{b}) > 0. \tag{15}$$

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

Solution: False. Since

$$y_i(\vec{x}_i^\top \widetilde{\vec{w}} - \widetilde{b}) > 0, \tag{16}$$

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

$$y_i(\vec{x}_i^\top \widetilde{\vec{w}} - \widetilde{b}) \ge \epsilon \implies y_i\left(\vec{x}_i^\top \frac{\widetilde{\vec{w}}}{\epsilon} - \frac{\widetilde{b}}{\epsilon}\right) \ge 1. \tag{17}$$

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

