Introduction to Machine Learning
Fall 2019
University of Science and Technology of China

Lecturer: Jie Wang — Homework 4
Posted: Nov. 5, 2019 — Due: Nov. 18, 2019
Name: San Zhang — ID: PBXXXXXXXX

Notice, to get the full credits, please show your solutions step by step.

Exercise 1: Support Vector Machine (SVM) for Linearly Separable Cases 40pts

Given the training sample $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$, where $\mathbf{x}_i \in \mathbb{R}^d$ and $y_i \in \{-1, 1\}$. Let

$$\mathcal{D}^+ = \{(\mathbf{x}_i, y_i) \in \mathcal{D} : y_i = 1\}, \quad \mathcal{D}^- = \{(\mathbf{x}_i, y_i) \in \mathcal{D} : y_i = -1\}.$$

Assume that $\mathcal{D}^+$ and $\mathcal{D}^-$ are nonempty and the training sample $\mathcal{D}$ is linearly separable. We have shown in class that SVM can be written as

$$\begin{aligned} \min_{\mathbf{w},b} \ & \frac{1}{2}\|\mathbf{w}\|^2, \\ \text{s.t.} \ & \min_i y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) = 1. \end{aligned} \tag{1}$$

Moreover, we further transform the problem in (1) to

$$\begin{aligned} \min_{\mathbf{w},b} \ & \frac{1}{2}\|\mathbf{w}\|^2, \\ \text{s.t.} \ & y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq 1, i = 1, \ldots, n. \end{aligned} \tag{2}$$

We denote the feasible set of the problem in (2) by

$$\mathcal{F} = \{(\mathbf{w}, b) : y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq 1, i = 1, \ldots, n\}.$$

1. Show that $\mathcal{F}$ is nonempty.

2. Show that the problem in (2) admits an optimal solution.

3. Let $(\mathbf{w}^*, b^*)$ be the optimal solution to problem (2). Show that $\mathbf{w}^* \neq 0$.

4. Show that the problems in (1) and (2) are equivalent, that is, they share the same set of optimal solutions.

5. Let $(\mathbf{w}^*, b^*)$ be the optimal solution to problem (2). Show there exist at least one positive sample and one negative sample, respectively, such that the equality holds. In other words, there exist $i, j \in \{1, 2, \ldots, n\}$ such that

$$\begin{aligned} 1 = y_i =& \langle \mathbf{w}^*, \mathbf{x}_i \rangle + b^*, \\ -1 = y_j =& \langle \mathbf{w}^*, \mathbf{x}_j \rangle + b^*. \end{aligned}$$

6. Show that the optimal solution to problem (2) is unique.

7. Find the dual problem of (2) and the corresponding optimal conditions.

Solution:

1. As the data is linearly separable, there exist some $\mathbf{w}_0, b_0$ such that

$$y_i(\langle \mathbf{w}_0, x_i\rangle + b_0) > 0$$

for $i = 1, 2 \ldots, n$. Suppose $\epsilon = \min_i(y_i(\langle \mathbf{w}_0, x_i\rangle + b_0))$. Let $\mathbf{w}' = \frac{1}{\epsilon}\mathbf{w}_0$, $b' = \frac{1}{\epsilon}b_0$. Then

$$y_i(\langle \mathbf{w}', x_i\rangle + b') \geq 1,\ i = 1, \ldots, n.$$

Therefore, $\mathcal{F}$ is nonempty.

2. Clearly, $\frac{1}{2}\|\mathbf{w}\|^2$ is convex quadratic and $\mathcal{F}$ is a polyhedral. f^* is finite since

$$\begin{cases} \frac{1}{2}\|\mathbf{w}\|^2 \geq 0 & \Rightarrow f^* > -\infty, \\ \mathcal{F} \text{ is nonempty} & \Rightarrow f^* < +\infty. \end{cases}$$

Therefore, the problem in (2) admits an optimal solution.

3. Suppose that $\mathbf{w}^* = 0$. Since $(\mathbf{w}^*, b^*)$ is feasible and $\mathcal{D}^+$ is nonempty, there exists $i \in \mathcal{D}^+$ such that

$$\langle \mathbf{w}^*, \mathbf{x}_i\rangle + b^* \geq 1 \Rightarrow b^* \geq 1.$$

Since $(\mathbf{w}^*, b^*)$ is feasible and $\mathcal{D}^-$ is nonempty, there exists $i \in \mathcal{D}^-$ such that

$$\langle \mathbf{w}^*, \mathbf{x}_i\rangle + b^* \leq -1 \Rightarrow b^* \leq -1,$$

which leads to a contradiction to $b^* \geq 1$.

4. Let $A = \{(\mathbf{w}, b) : \min_i(y_i(\langle \mathbf{w}, \mathbf{x}_i\rangle + b) = 1\}$ and $B = \{(\mathbf{w}, b) : \min_i(y_i(\langle \mathbf{w}, \mathbf{x}_i\rangle + b) > 1\}$. Note that $\mathcal{F} = A \cup B$. We first prove that the set of optimal solutions of (2) belongs to A.

Suppose that there exists $(\mathbf{w}^*, b^*) \in \mathbf{argmin}_{(\mathbf{w},b)\in\mathcal{F}}\ \frac{1}{2}\|\mathbf{w}\|^2$ such that

$$(\mathbf{w}^*, b^*) \in B.$$

Suppose that $\min_i y_i(\langle \mathbf{w}^*, \mathbf{x}_i\rangle + b^*) = 1 + \epsilon$ for some $\epsilon > 0$. Let $\mathbf{w}' = \frac{1}{1+\epsilon}\mathbf{w}^*$, $b' = \frac{1}{1+\epsilon}b^*$. Then

$$\|\mathbf{w}'\|^2 < \|\mathbf{w}^*\|^2$$

and

$$\min_i y_i\langle \mathbf{w}', \mathbf{x}_i\rangle + b' = 1,$$

which leads to a contradiction. Therefore, the optimal solutions to (2) belong to A. Note that $A \cap B = \emptyset$, then problem (2) can be equivalent to the problem as follows.

$$\begin{aligned} \min_{\mathbf{w},b}\ & \frac{1}{2}\|\mathbf{w}\|^2 \\ \text{s.t. }\ & (\mathbf{w}, b) \in A \end{aligned} \tag{3}$$

That is, problem (2) is equivalent to problem (1).

5. It follows from Exercise 1.4 that there exists at least one sample $(\mathbf{x}_k, y_k)$ such that

$$y_k = \langle \mathbf{w}^*, \mathbf{x}_k \rangle + b^*.$$

W.L.O.G, assume $(\mathbf{x}_k, y_k) \in \mathcal{D}^+$. Suppose that

$$\forall\, (\mathbf{x}_i, y_i) \in \mathcal{D}^-,\ \langle \mathbf{w}^*, \mathbf{x}_i \rangle + b^* \leq -1 - \delta$$

for some $\delta > 0$. Let $b = b^* - \frac{\delta}{2}$. We have

$$\forall\, (\mathbf{x}_i, y_i) \in \mathcal{D}^+,\ \langle \mathbf{w}^*, \mathbf{x}_i \rangle + b \geq 1 + \frac{\delta}{2}$$

and

$$\forall\, (\mathbf{x}_i, y_i) \in \mathcal{D}^-,\ \langle \mathbf{w}^*, \mathbf{x}_i \rangle + b \leq -(1 + \frac{\delta}{2}).$$

Let $\mathbf{w}' = \mathbf{w}^*/(1 + \frac{\delta}{2})$, $b' = b/(1 + \frac{\delta}{2})$. Then

$$\|\mathbf{w}'\|^2 < \|\mathbf{w}^*\|^2$$

and

$$\min_i y_i \langle \mathbf{w}', \mathbf{x}_i \rangle + b' \geq 1,$$

which leads to a contradiction.

Therefore, there exist at least one positive sample and one negative sample, respectively, such that the equality in problem (2) holds.

6. Suppose that there exist two different optimal solutions, i.e. $(\mathbf{w}_1, b_1), (\mathbf{w}_2, b_2)$. Then

$$\begin{aligned} \|\mathbf{w}_1\| &= \|\mathbf{w}_2\|, \\ \langle \mathbf{w}_1, \mathbf{w}_2 \rangle &< \|\mathbf{w}_1\|^2. \end{aligned}$$

Let $\mathbf{w}' = \frac{\mathbf{w}_1+\mathbf{w}_2}{2}, b' = \frac{b_1+b_2}{2}$. Then

$$\begin{aligned}\|\mathbf{w}'\|^2 &= \frac{\|\mathbf{w}_1+\mathbf{w}_2\|^2}{4}\\ &= \frac{2\|\mathbf{w}_1\|^2+2\langle\mathbf{w}_1,\mathbf{w}_2\rangle}{4}\\ &< \|\mathbf{w}_1\|^2,\end{aligned}$$

and

$$y_i(\langle\mathbf{w}',\mathbf{x}_i\rangle+b') \geq 1$$

for $i = 1, \ldots, n$.

Therefore, this is contradictory to that $\mathbf{w}_1$ is an optimal solution, i.e. the optimal solution is unique.

7. We first construct the Lagrangian:

$$L(\mathbf{w}, b, \alpha) = \frac{1}{2}\|\mathbf{w}\|^2 + \sum_{i=1}^{n}\alpha_i(1-y_i(\langle\mathbf{w},\mathbf{x}_i\rangle+b)), \tag{4}$$

where $\alpha_i \geq 0$, $i = 1, \ldots, n$.

We next find the dual function:

$$q(\alpha) = \inf_{\mathbf{w},b}(\frac{1}{2}\|\mathbf{w}\|^2 + \sum_{i=1}^{n}\alpha_i(1-y_i(\langle\mathbf{w},\mathbf{x}_i+b))).$$

For fixed α, let $(\mathbf{w}^*, b^*)$ be the optimal solution to the above problem. The first order optimal condition implies that

$$\nabla_{\mathbf{w}}L(\mathbf{w},b)|_{\mathbf{w}=\mathbf{w}^*} = \mathbf{0} \Rightarrow \mathbf{w}^* - \sum_{i=1}^{n}\alpha_i y_i\mathbf{x}_i = \mathbf{0},$$
$$\nabla_b L(\mathbf{w},b)|_{b=b^*} = 0 \Rightarrow -\sum_{i=1}^{n}\alpha_i y_i = 0.$$

Plugging the above equations into Eq. (4) leads to

$$q(\alpha) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j\langle\mathbf{x}_i,\mathbf{x}_j\rangle + \sum_{i=1}^{n}\alpha_i.$$

Thus, the dual problem of (2) is

$$
\begin{aligned}
\max_{\alpha} \quad & -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j\langle \mathbf{x}_i, \mathbf{x}_j\rangle + \sum_{i=1}^{n}\alpha_i, \\
\text{s.t.} \quad & \sum_{i=1}^{n}\alpha_i y_i = 0, \\
& \alpha_i \geq 0,\ i = 1, \ldots, n.
\end{aligned}
$$

Suppose that α^* is a geometric multiplier. The corresponding optimal conditions are

$$
\begin{aligned}
\mathbf{w}^* \in \mathbb{R}^d,\, b^* \in \mathbb{R},\ 1 - y_i(\langle \mathbf{w}^*, \mathbf{x}_i\rangle + b^*) \leq 0, i = 1, \ldots, n, & \qquad \text{(Primal Feasibility)}, \\
\alpha_i^* \geq 0, i = 1, \ldots, n, & \qquad \text{(Dual Feasibility)}, \\
\sum_{i=1}^{n}\alpha_i^* y_i = 0,\, \mathbf{w}^* = \sum_{i=1}^{n}\alpha_i^* y_i \mathbf{x}_i, & \qquad \text{(Lagrangian Optimality)}, \\
\alpha_i^*(1 - y_i(\langle \mathbf{w}^*, \mathbf{x}_i\rangle + b)) = 0, i = 1, \ldots, n. & \qquad \text{(Complementary Slackness)}.
\end{aligned}
$$

■

Exercise 2: Visualization Lemma

Consider the primal problem as follows.

$$\begin{aligned}\min_{\mathbf{x}}\ & f(\mathbf{x}) \\ \text{s.t.}\ & \mathbf{g}(\mathbf{x}) \leq 0, \\ & \mathbf{h}(\mathbf{x}) = 0, \\ & \mathbf{x} \in X,\end{aligned} \tag{5}$$

where $\mathbf{x} \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$, $\mathbf{g} : \mathbb{R}^n \to \mathbb{R}^m$, $\mathbf{h} : \mathbb{R}^n \to \mathbb{R}^p$, and $X \subseteq \mathbb{R}^n$. The functions f, $\mathbf{g}$, and $\mathbf{h}$ are continuously differentiable.

The Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ associated with the problem in (5) takes the form of

$$L(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{i=1}^{p} \mu_i h_i(\mathbf{x}). \tag{6}$$

Let

$$\mathbb{R}^{m+p+1} \supseteq S = \{(\mathbf{g}(\mathbf{x}), \mathbf{h}(\mathbf{x}), f(\mathbf{x})) : \mathbf{x} \in X\}. \tag{7}$$

Show that the results as follows (hint: see Fig. 1).

Lemma 1. Visualization Lemma

1. *The hyperplane with normal $(\lambda, \mu, 1)$ that passes through a vector $(\mathbf{g}(\mathbf{x}), \mathbf{h}(\mathbf{x}), f(\mathbf{x}))$ intercepts the vertical axis $\{(\mathbf{0}, z) : z \in \mathbb{R}\}$ at the level $L(\mathbf{x}, \lambda, \mu)$.*

2. *Among all hyperplanes with normal $(\lambda, \mu, 1)$ that contains in their positive halfspace the set S defined in (7), the highest attained level of interception of the vertical axis is $\inf_{\mathbf{x} \in X} L(\mathbf{x}, \lambda, \mu)$.*

3. *(λ^*, μ^*) is a geometric multiplier if and only if $\lambda^* \geq 0$ and among all hyperplanes with normal $(\lambda^*, \mu^*, 1)$ that contain in their positive halfspace the set S, the highest attained level of of interception of the vertical axis is f^*, where*

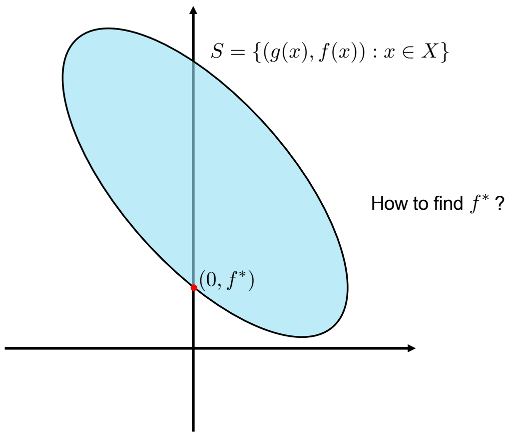
$$f^* = \min\{f(\mathbf{x}) : \mathbf{g}(\mathbf{x}) \leq 0, \mathbf{h}(\mathbf{x}) = 0, \mathbf{x} \in X)\}.$$

Solution: 1. Let $\mathbf{a} \in \mathbb{R}^m$, $\mathbf{b} \in \mathbb{R}^p$ and $z \in \mathbb{R}$. The hyperplane with normal $(\lambda, \mu, 1)$ that passes through a vector $(\mathbf{g}(\mathbf{x}), \mathbf{h}(\mathbf{x}), f(\mathbf{x}))$ can be written as

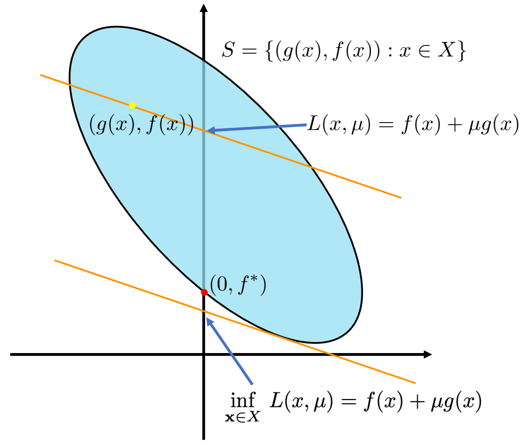
$$\lambda \mathbf{a} + \mu \mathbf{b} + z = f(\mathbf{x}) + \langle \lambda, \mathbf{g}(\mathbf{x}) \rangle + \langle \mu, \mathbf{h}(\mathbf{x}) \rangle.$$

We note that the right hand side of the above equation is indeed $L(\mathbf{x}, \lambda, \mu)$. Thus, we can write the aforementioned hyperplane as

$$\langle \lambda, \mathbf{a} \rangle + \langle \mu, \mathbf{b} \rangle + z = L(\mathbf{x}, \lambda, \mu).$$



(a)



(b)

Figure 1: Illustration of the visualization lemma with one inequality constraint.

Therefore, by setting $\mathbf{a} = \mathbf{0}$, $\mathbf{b} = \mathbf{0}$, we have

$$z = L(\mathbf{x}, \lambda, \mu).$$

2. The hyperplane H with normal $(\lambda, \mu, 1)$ which intercepts the vertical axis at the level c takes the form of

$$\langle \lambda, \mathbf{a} \rangle + \langle \mu, \mathbf{b} \rangle + z = c.$$

Suppose that S is in the positive halfspace of H. This implies that

$$L(\mathbf{x}, \lambda, \mu) = \langle \lambda, \mathbf{g}(\mathbf{x}) \rangle + \langle \mu, \mathbf{h}(\mathbf{x}) \rangle + f(\mathbf{x}) \geq c, \forall\, (\mathbf{g}(\mathbf{x}), \mathbf{h}(\mathbf{x}), f(\mathbf{x})) \in S,$$

which is equivalent to

$$L(\mathbf{x}, \lambda, \mu) \geq c, \forall\, \mathbf{x} \in X.$$

Thus, we have

$$\inf_{\mathbf{x} \in X} L(\mathbf{x}, \lambda, \mu) \geq c.$$

Therefore, the maximum value of c is $\inf_{\mathbf{x} \in X} L(\mathbf{x}, \lambda, \mu)$.

3. $\Rightarrow$: By definition of the geometric multiplier, we have

$$\begin{aligned} &\lambda^\star \geq 0, \\ &f^\star = \inf_{\mathbf{x} \in X} L(\mathbf{x}, \lambda^*, \mu^*). \end{aligned}$$

It follows from Exercise 2.2 that the highest attained level of interception of the vertical axis is

$$\inf_{\mathbf{x}\in X} L(\mathbf{x}, \lambda^*, \mu^*) = f^*.$$

$\Leftarrow$: It follows from Exercise 2.2 that

$$\inf_{\mathbf{x}\in X} L(\mathbf{x}, \lambda^*, \mu^*) = f^*.$$

And $\lambda^* \geq 0$, thus (λ^*, μ^*) is a geometric multiplier.

$\blacksquare$

Exercise 3: Geometric Multiplier

Let (λ^*, μ^*) be a geometric multiplier. Show that $\mathbf{x}^*$ is a global minimum of the primal problem (5) if and only if $\mathbf{x}^*$ is feasible and

$$\mathbf{x}^* \in \underset{\mathbf{x}\in X}{\mathbf{argmin}}\, L(\mathbf{x}, \lambda^*, \mu^*),$$
$$\lambda_i^* g_i(\mathbf{x}^*) = 0,\ i = 1, \ldots, m.$$

Solution: $\Rightarrow$: As $\mathbf{x}^*$ is a global minimum of the primal problem (5) and (λ^*, μ^*) is a geometric multiplier, $\mathbf{x}^*$ is feasible and

$$\begin{aligned}\inf_{\mathbf{x}\in X} L(\mathbf{x}, \lambda^*, \mu^*) &= f(\mathbf{x}^*) \\ &\geq f(\mathbf{x}^*) + \langle \lambda^*, \mathbf{g}(\mathbf{x}^*)\rangle + \langle \mu^*, \mathbf{h}(\mathbf{x}^*)\rangle \\ &= L(\mathbf{x}^*, \lambda^*, \mu^*).\end{aligned}$$

The inequality comes from that $\langle \lambda^*, \mathbf{g}(\mathbf{x}^*)\rangle \leq 0$ and $\langle \mu^*, \mathbf{h}(\mathbf{x}^*)\rangle = 0$ for all feasible $\mathbf{x}^*$. Then, $\inf_{\mathbf{x}\in X} L(\mathbf{x}, \lambda^*, \mu^*) \geq L(\mathbf{x}^*, \lambda^*, \mu^*)$ leads to

$$\begin{aligned}\inf_{\mathbf{x}\in X} L(\mathbf{x}, \lambda^*, \mu^*) &= L(\mathbf{x}^*, \lambda^*, \mu^*), \\ &\Rightarrow \mathbf{x}^* \in \underset{\mathbf{x}\in X}{\mathbf{argmin}}\, L(\mathbf{x}, \lambda^*, \mu^*),\end{aligned}$$

and

$$f(\mathbf{x}^*) = L(\mathbf{x}^*, \lambda^*, \mu^*) = f(\mathbf{x}^*) + \langle \lambda^*, \mathbf{g}(\mathbf{x}^*)\rangle + \langle \mu^*, \mathbf{h}(\mathbf{x}^*)\rangle.$$

As $\mathbf{x}^*$ is feasible, we have

$$\lambda_i^* g_i(\mathbf{x}^*) = 0,\ i = 1, \ldots, m.$$

$\Leftarrow$: As $\mathbf{x}^* \in \mathbf{argmin}_{\mathbf{x}\in X}\, L(\mathbf{x}, \lambda^*, \mu^*)$, it follows from Exercise 2.3 that

$$L(\mathbf{x}^*, \lambda^*, \mu^*) = \min_{\mathbf{x}\in X} L(\mathbf{x}, \lambda^*, \mu^*) = f^*$$

As $\mathbf{x}^*$ is feasible and $\lambda_i^* g_i(\mathbf{x}^*) = 0$ for $i = 1, \ldots, m$, we have

$$L(\mathbf{x}^*, \lambda^*, \mu^*) = f(\mathbf{x}^*) + \langle \lambda^*, \mathbf{g}(\mathbf{x}^*)\rangle + \langle \mu^*, \mathbf{h}(\mathbf{x}^*)\rangle = f(\mathbf{x}^*)$$

Therefore, we have

$$f(\mathbf{x}^*) = L(\mathbf{x}^*, \lambda^*, \mu^*) = f^*.$$

Thus $\mathbf{x}^*$ is a global minimum of the primal problem (5).

■

Exercise 4: Lagrange Dual Problem

Consider the primal problem (5) and the Lagrangian (6). We define the dual function for $(\lambda, \mu) \in \mathbb{R}^{m+p}$ by

$$q(\lambda, \mu) = \inf_{\mathbf{x} \in X} L(\mathbf{x}, \lambda, \mu).$$

The domain of q is

$$\mathbf{dom}\,(q) = \{(\lambda, \mu) : q(\lambda, \mu) > -\infty\}.$$

The dual problem is

$$\begin{aligned} &\sup q(\lambda, \mu), \\ &\text{s.t. } \lambda \geq 0. \end{aligned}$$

1. Show that $\mathbf{dom}\,(q)$ is convex.
2. Show that $-q(\lambda, \mu)$ is a convex function.

Solution:

1. Let $(\lambda_1, \mu_1), (\lambda_2, \mu_2) \in \mathbf{dom}\,(q)$, and $a \in [0, 1]$. Consider

$$(\lambda, \mu) = (a\lambda_1 + (1-a)\lambda_2, a\mu_1 + (1-a)\mu_2).$$

We have

$$\begin{aligned} L(\mathbf{x}, \lambda, \mu) &= f(\mathbf{x}) + \langle a\lambda_1 + (1-a)\lambda_2, \mathbf{g}(\mathbf{x})\rangle + \langle a\mu_1 + (1-a)\mu_2, \mathbf{h}(\mathbf{x}) \\ &= a(f(\mathbf{x}) + \langle \lambda_1, \mathbf{g}(\mathbf{x})\rangle + \langle \mu_1, \mathbf{h}(\mathbf{x})\rangle) + (1-a)(f(\mathbf{x}) + \langle \lambda_2, \mathbf{g}(\mathbf{x})\rangle + \langle \mu_2, \mathbf{h}(\mathbf{x})\rangle) \\ &= aL(\mathbf{x}, \lambda_1, \mu_1) + (1-a)L(\mathbf{x}, \lambda_2, \mu_2). \end{aligned}$$

Therefore,

$$\begin{aligned} q(\lambda, \mu) &= \inf_{\mathbf{x} \in X} L(\mathbf{x}, \lambda, \mu) \\ &= \inf_{\mathbf{x} \in X} aL(\mathbf{x}, \lambda_1, \mu_1) + (1-a)L(\mathbf{x}, \lambda_2, \mu_2) \\ &\geq \inf_{\mathbf{x} \in X} aL(\mathbf{x}, \lambda_1, \mu_1) + \inf_{\mathbf{x} \in X} (1-a)L(\mathbf{x}, \lambda_2, \mu_2) \\ &= aq(\lambda_1, \mu_1) + (1-a)q(\lambda_2, \mu_2) \\ &> -\infty. \end{aligned}$$

Thus, $\mathbf{dom}\,(q)$ is convex.

2. Let $(\lambda_1, \mu_1), (\lambda_2, \mu_2) \in \mathbf{dom}\,(q)$, and $a \in [0, 1]$. Consider

$$(\lambda, \mu) = (a\lambda_1 + (1-a)\lambda_2, a\mu_1 + (1-a)\mu_2).$$

We have

$$
\begin{aligned}
&q(\lambda, \mu) \geq aq(\lambda_1, \mu_1) + (1-a)q(\lambda_2, \mu_2) \\
\Rightarrow &-q(\lambda, \mu) \leq a(-q(\lambda_1, \mu_1)) + (1-a)(-q(\lambda_2, \mu_2)).
\end{aligned}
$$

Therefore, $-q(\lambda, \mu)$ is a convex function.

■

Exercise 5: Duality Gap

Duality gap is defined by

$$f^* - q^*.$$

Show that the following results hold.

1. If there is no duality gap, the set of geometric multipliers is equal to the set of dual optimal solutions.
2. If there is duality gap, the set of geometric multipliers is empty.

Solution:

1. Duality gap is zero implies

$$f^* = q^* \tag{8}$$

We first show that the set of geometric multipliers belongs to the set of dual optimal solutions. Let (λ^*, μ^*) be a geometric multiplier. Thus

$$q(\lambda^*, \mu^*) = \inf_{x \in X} L(\mathbf{x}, \lambda^*, \mu^*) = f^* = q^*.$$

Therefore, (λ^*, μ^*) is a dual optimal solution.

We next show that the set of dual optimal solutions belongs to the set of geometric multipliers. Let (λ^*, μ^*) be a dual optimal solution.

Since

$$\inf_{x \in X} L(\mathbf{x}, \lambda^*, \mu^*) = q(\lambda^*, \mu^*) = q^* = f^*,$$

(λ^*, μ^*) is a geometric multiplier for the primal problem.

Therefore, the set of geometric multipliers is equal to the set of dual optimal solutions.

2. Suppose that $f^* - q^* > 0$. This implies that, for any (λ, μ),

$$q(\lambda, \mu) = \inf_{\mathbf{x} \in X} L(\mathbf{x}, \lambda, \mu) \leq q^* < f^*.$$

Therefore, there is no (λ^*, μ^*), such that $q(\lambda^*, \mu^*) = \inf_{\mathbf{x} \in X} L(\mathbf{x}, \lambda^*, \mu^*) = f^*$. Thus, the set of geometric multipliers is empty.

■

Exercise 6: Optimality Conditions

Show that a pair $\mathbf{x}^*$ and (λ^*, μ^*) is an optimal solution and geometric multiplier pair if and only if

$$
\begin{aligned}
\mathbf{x}^* \in X,\ \mathbf{g}(\mathbf{x}^*) \le 0, \mathbf{h}(\mathbf{x}^*) = 0, &\qquad \text{(Primal Feasibility)}, && (9)\\
\lambda^* \ge 0, &\qquad \text{(Dual Feasibility)}, && (10)\\
\mathbf{x}^* \in \operatorname*{\mathbf{argmin}}_{\mathbf{x}\in X} L(\mathbf{x}, \lambda^*, \mu^*), &\qquad \text{(Lagrangian Optimality)}, && (11)\\
\lambda_i^* g_i(\mathbf{x}^*) = 0,\ i = 1, \ldots, m, &\qquad \text{(Complementary Slackness)}. && (12)
\end{aligned}
$$

Solution: $\Leftarrow$: Let $D = \{\mathbf{x} \in X : \mathbf{g}(\mathbf{x}) \le 0, \mathbf{h}(\mathbf{x}) = 0\}$. We have

$$
\begin{aligned}
f(\mathbf{x}^*) \overset{(9),(12)}{=} L(\mathbf{x}^*, \lambda^*, \mu^*) &\overset{(11)}{=} \inf_{\mathbf{x}\in X} L(\mathbf{x}, \lambda^*, \mu^*)\\
&\le \inf_{\mathbf{x}\in D} L(\mathbf{x}, \lambda^*, \mu^*)\\
&\overset{(9),(10)}{\le} \inf_{\mathbf{x}\in D} f(\mathbf{x})\\
&\le f(\mathbf{x}^*).
\end{aligned}
$$

Thus $f(\mathbf{x}^*) = \inf_{\mathbf{x}\in X} f(\mathbf{x})$, i.e. $\mathbf{x}^*$ is an optimal solution and (λ^*, μ^*) is a geometric multiplier pair.

$\Rightarrow$: Clearly, (9) and (10) hold. It follows from Exercise 2.3 that

$$f(\mathbf{x}^*) = \inf_{\mathbf{x}\in X} L(\mathbf{x}, \lambda^*, \mu^*) \le \inf_{\mathbf{x}\in D} L(\mathbf{x}, \lambda^*, \mu^*).$$

As (9) and (10) hold, we have

$$
\begin{aligned}
L(\mathbf{x}^*, \lambda^*, \mu^*) &= f(\mathbf{x}^*) + \langle \lambda^*, \mathbf{g}(\mathbf{x}^*)\rangle + \langle \mu^*, \mathbf{h}(\mathbf{x}^*)\rangle\\
&\le f(\mathbf{x}^*)\\
&= \inf_{\mathbf{x}\in X} L(\mathbf{x}, \lambda^*, \mu^*)
\end{aligned}
$$

Therefore, $\mathbf{x}^* \in \mathbf{argmin}_{\mathbf{x}\in X} L(\mathbf{x}, \lambda^*, \mu^*)$, and

$$f(\mathbf{x}^*) = L(\mathbf{x}^*, \lambda^*, \mu^*) = f(\mathbf{x}^*) + \langle \lambda^*, \mathbf{g}(\mathbf{x}^*)\rangle + \langle \mu^*, \mathbf{h}(\mathbf{x}^*)\rangle.$$

Thus

$$\lambda_i^* g_i(\mathbf{x}^*) = 0,\ i = 1, \ldots, m.$$

■

Exercise 7: Saddle Point Interpretation

Show that a pair $\mathbf{x}^*$ and (λ^*, μ^*) is an optimal solution-geometric multiplier pair if and only if $\mathbf{x}^* \in X$, $\lambda^* \geq 0$, and $(\mathbf{x}^*, \lambda^*, \mu^*)$ is a saddle point of the Lagrangian, in the sense that

$$L(\mathbf{x}^*, \lambda, \mu) \leq L(\mathbf{x}^*, \lambda^*, \mu^*) \leq L(\mathbf{x}, \lambda^*, \mu^*),\ \forall\, \mathbf{x} \in X,\ \lambda \geq 0. \tag{13}$$

Proof. $\Rightarrow$: It follows from Exercise 6 that

$$\begin{aligned} L(\mathbf{x}^*, \lambda, \mu) &= f(\mathbf{x}^*) + \langle \lambda, \mathbf{g}(\mathbf{x}^*)\rangle + \langle \mu, \mathbf{h}(\mathbf{x}^*)\rangle \\ &\overset{(9),(10)}{\leq} f(\mathbf{x}^*) \\ &\overset{(9),(12)}{=} L(\mathbf{x}^*, \lambda^*, \mu^*) \\ &\overset{(11)}{\leq} L(\mathbf{x}, \lambda^*, \mu^*). \end{aligned}$$

$\Leftarrow$: Clearly, (10) and (11) hold. Next we show that (9) and (12) hold. The saddle point property of the Lagrangian in (13) implies that

$$\begin{aligned} & L(\mathbf{x}^*, \lambda, \mu) \leq L(\mathbf{x}^*, \lambda^*, \mu^*),\ \forall \lambda \geq 0 \\ \Rightarrow & f(\mathbf{x}^*) + \langle \lambda, \mathbf{g}(\mathbf{x}^*)\rangle + \langle \mu, \mathbf{h}(\mathbf{x}^*)\rangle \leq L(\mathbf{x}^*, \lambda^*, \mu^*), \forall \lambda \geq 0 \end{aligned}$$

That is, $L(\mathbf{x}^*, \lambda, \mu)$ is upper bounded for any $\lambda \geq 0$. Therefore,

$$\mathbf{g}(\mathbf{x}^*) \leq 0,\ \mathbf{h}(\mathbf{x}^*) = 0,$$

i.e., the primal feasibility (9) holds.

Next we show that the complementary slackness in (12) holds. We combine the primal feasibility of $\mathbf{x}^*$ and left half of (13)

$$f(\mathbf{x}^*) + \langle \lambda, \mathbf{g}(\mathbf{x}^*)\rangle \leq f(\mathbf{x}^*) + \langle \lambda^*, \mathbf{g}(\mathbf{x}^*)\rangle,\ \forall\, \lambda \geq 0.$$

By letting $\lambda \to 0$, we have

$$\langle \lambda^*, \mathbf{g}(\mathbf{x}^*)\rangle \geq 0.$$

On the other hand, in view of the fact that $\lambda^* \geq 0$ and $\mathbf{g}(\mathbf{x}^*) \leq 0$, we have

$$\langle \lambda^*, \mathbf{g}(\mathbf{x}^*)\rangle \leq 0.$$

All together, we have

$$\langle \lambda^*, \mathbf{g}(\mathbf{x}^*)\rangle = 0.$$

Thus, the complementary slackness (12) holds. $\square$

Exercise 8

Recall that the soft margin SVM takes the form of

$$\begin{aligned} \min_{\mathbf{w},b,\xi} \ & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i, \\ \text{s.t. } & y_i(\langle \mathbf{w}, \mathbf{x}_i\rangle + b) \geq 1 - \xi_i, i = 1,\dots,n, \\ & \xi_i \geq 0,\ i = 1,\dots,n, \end{aligned} \tag{14}$$

where $C > 0$. The corresponding dual problem is

$$\begin{aligned} \min_{\alpha} \ & \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j\rangle - \sum_{i=1}^{n}\alpha_i \\ \text{s.t. } & \sum_{i=1}^{n}\alpha_i y_i = 0, \\ & \alpha_i \in [0, C], i = 1,\dots,n. \end{aligned} \tag{15}$$

1. Show that the problems (14) and (15) always admit optimal solutions.
2. Suppose that the training data consists of two data instances $x_1 = 1$ and $x_2 = -1$, and the corresponding labels are $y_1 = 1$ and $y_2 = -1$.
   (a) Please solve problem (2) and find the primal and dual optimal solutions.
   (b) Please solve problem (14) with $C < \frac{1}{2}$ and find the primal and dual optimal solutions.
3. Suppose that the training data consists of four data instances: $\mathbf{x}_1 = (2, 3)$, $\mathbf{x}_2 = (1, 2)$, $\mathbf{x}_3 = (1, 3)$, and $\mathbf{x}_4 = (2, 2)$, and the corresponding labels are $y_1 = y_2 = 1$ and $y_3 = y_4 = -1$. Please solve the problem in (15) with $C = 10$.

Notice that, for the last two parts, you need to find all the primal and dual optimal solutions if they are not unique.

Solution: 1. First we show that the feasible set is nonempty. Given $\mathbf{w}_0 \in \mathbb{R}^d$, $b_0 \in \mathbb{R}$. Let $\epsilon = \min_i y_i(\langle \mathbf{w}_0, \mathbf{x}_i\rangle + b_0)$, then there exists $\xi \in \mathbb{R}^n$ such that

$$\min_i \xi_i = \max\{1 - \epsilon, 0\}.$$

Then

$$y_i(\langle \mathbf{w}_0, \mathbf{x}_i + b_0) \geq \epsilon \geq 1 - \xi_i,\ i = 1,\dots,n.$$

Thus $(\mathbf{w}_0, b_0, \xi)$ is a feasible solution of (14), i.e. the feasible set is nonempty.

Note that the objective function of (14) is convex quadratic and the feasible set of

(14) is polyhedral. The optimal value f^* is finite since

$$\begin{cases} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^n \xi_i \geq 0 & \Rightarrow f^* > -\infty, \\ \text{the feasible set is nonempty} & \Rightarrow f^* < +\infty. \end{cases}$$

Thus both problem (14) and (15) have optimal solutions and there is no duality gap.

2. (a) Problem (2) is

$$\begin{aligned} \min_{w,b}\ & \frac{1}{2}\|w\|^2, \\ \text{s.t. } & w + b \geq 1, \\ & w - b \geq 1. \end{aligned}$$

The primal solution is

$$(w^*, b^*) = (1, 0).$$

It follows from Exercise 1.6 that the primal solution is unique.
The dual problem of problem (2) is

$$\begin{aligned} \min_{\alpha}\ & -\frac{1}{2}(\alpha_1+\alpha_2)^2 - \alpha_1(1-\alpha_1-\alpha_2) \\ & -\alpha_2(1-\alpha_1-\alpha_2), \\ \text{s.t. }\ & \alpha_i \geq 0,\ i = 1, 2, \\ & \alpha_1 = \alpha_2. \end{aligned}$$

The dual solution is

$$(\alpha_1^*, \alpha_2^*) = (\frac{1}{2}, \frac{1}{2}).$$

(b) The dual problem of problem(14) is

$$\begin{aligned} \min_{\alpha}\ & \frac{1}{2}(\alpha_1^2 + 2\alpha_1\alpha_2 + \alpha_2^2) - (\alpha_1 + \alpha_2), \\ \text{s.t. }\ & \alpha_1 - \alpha_2 = 0, \\ & \alpha_i \in [0, C], i = 1, 2. \end{aligned}$$

As $C < \frac{1}{2}$, the dual solution is

$$(\alpha_1^*, \alpha_2^*) = (C, C).$$

Then

$$w^* = \alpha_1^* y_1 x_1 + \alpha_2^* y_2 x_2 = 2C,$$

which is unique. Since there is no duality gap, we have

$$q^* = 2C - 2C^2 = f^* = 2C^2 + C(\xi_1^* + \xi_2^*)$$

Thus the set of primal optimal solutions is

$$\{(w, b, \xi) : w = 2C,\ \xi_1 = t,\ \xi_2 = 2 - 4C - t,\ b = 1 - 2C - t,\ t \in [0, 2 - 4C]\}$$

3. The dual problem of problem(14) is

$$\begin{aligned}
\min_{\alpha}\ & \frac{1}{2}(13\alpha_1^2 + 5\alpha_2^2 + 10\alpha_3^2 + 8\alpha_4^2 + 16\alpha_1\alpha_2 - 22\alpha_1\alpha_3 \\
& - 20\alpha_1\alpha_4 - 14\alpha_2\alpha_3 - 12\alpha_2\alpha_4 + 16\alpha_3\alpha_4) - (\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4), \\
\text{s.t. }\ & \alpha_1 + \alpha_2 = \alpha_3 + \alpha_4 \\
& \alpha_i \in [0, 10],\ i = 1, \ldots, 4.
\end{aligned} \tag{16}$$

The dual optimal solution is

$$\alpha_1^* = \alpha_2^* = \alpha_3^* = \alpha_4^* = 10.$$

Then

$$\mathbf{w}^* = \sum_{i=1}^{4} \alpha_i^* y_i \mathbf{x}_i = \mathbf{0},$$

which is unique. Note that $q^* = 40$. Then consider the following inequalities.

$$\begin{cases}
10\sum_{i=1}^4 \xi_i & = q^* = 40 & \\
1 - b - \xi_i & \leq 0, & i = 1, 2 \\
1 + b - \xi_i & \leq 0, & i = 3, 4 \\
-\xi_i & \leq 0, & i = 1, 2, 3, 4
\end{cases}$$

The set of primal optimal solutions is

$$\{(\mathbf{w}, b, \xi) : \mathbf{w} = \mathbf{0},\ \xi_1 = \xi_2 = 1 - t,\ \xi_3 = \xi_4 = 1 + t,\ b = t,\ t \in [-1, 1]\}.$$

$\blacksquare$