**Introduction to Machine Learning**
Spring 2021
University of Science and Technology of China

Lecturer: Jie Wang — Homework 5
Posted: May. 12, 2021 — Due: May. 21, 2021
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**

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. The Euclidean distance between a linear classifier $f(\mathbf{x}; \mathbf{w}, b) = \langle \mathbf{w}, \mathbf{x} \rangle + b$ and a point $\mathbf{z}$ is

   $$d(\mathbf{z}, f) = \min_{\mathbf{x}} \{\|\mathbf{z} - \mathbf{x}\| : f(\mathbf{x}; \mathbf{w}, b) = 0\}.$$

   Please find the closed form of $d(\mathbf{z}, f)$.

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

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

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

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

6. 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 corresponding equality holds. In other words, there exist $i, j \in \{1, 2, \ldots, n\}$ such that

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

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

8. Can we remove the inequalities that hold strictly at the optimum to problem (2) without affecting the solution? Please justify your claim rigorously.

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

**Solution:** ■

**Exercise 2: Exercises of Dual Problems**

1. Please find the sets of all optimal solutions and all Lagrange multipliers, and sketch the dual function for the following two-dimensional convex programming problems:

$$\begin{aligned}\min_{x_1,x_2}\ & x_1\\ \text{s.t.}\ & |x_1|+|x_2|\le 1,\\ & (x_1,x_2)\in X=\mathbb{R}^2,\end{aligned}$$

   and

$$\begin{aligned}\min_{x_1,x_2}\ & x_1\\ \text{s.t.}\ & |x_1|+|x_2|\le 1,\\ & (x_1,x_2)\in X=\{(x_1,x_2):|x_1|\le 1,|x_2|\le 1\}.\end{aligned}$$

2. Consider the problem

$$\begin{aligned}\min_{(x_1,x_2)}\ & 10x_1+3x_2\\ \text{s.t.}\ & 5x_1+x_2\ge 4,\\ & x_1,x_2=0 \text{ or } 1.\end{aligned} \tag{3}$$

   (a) Sketch the set of constraint-cost pairs

$$\{(4-5x_1-x_2, 10x_1+3x_2): x_1,x_2=0 \text{ or } 1\}.$$

   (b) Sketch the dual function.

   (c) Solve the problem (3) and its dual, and relate the solutions to your sketch in part (a).

3. Please use duality to show that in three-dimensional space, the (minimum) distance from the origin to a line is equal to the maximum over all (minimum) distances of the origin from planes that contain the line.

4. Derive the dual of the projection problem

$$\begin{aligned}\min_{\mathbf{x}}\ & \|\mathbf{z}-\mathbf{x}\|^2\\ \text{s.t.}\ & \mathbf{A}\mathbf{x}=0,\end{aligned}$$

   where the matrix $\mathbf{A}\in\mathbb{R}^{m\times n}$ and the vector $\mathbf{z}\in\mathbb{R}^n$ are given. Show that the dual problem is also a problem of projection on a subspace.

5. Derive the dual of the following unconstrained geometric program problem

$$\min_{\mathbf{x}} \log\left(\sum_{i=1}^{m}\exp\left(\mathbf{a}_i^T\mathbf{x}+b_i\right)\right),$$

where the vector $\mathbf{a}_i \in \mathbb{R}^n$ and the scalar $b_i$ are given for $i = 1, 2, \cdots, m$. Show that the dual problem is an entropy maximization problem.

6. Consider the following problems.

   (a) Let $x, y \in \mathbb{R}$.

$$\begin{aligned} &\min_{x,y} e^{-x} \\ &\text{s.t.}\, x^2/y \leq 1, \\ &\qquad (x, y) \in \{(x, y) | y > 0\}. \end{aligned} \tag{4}$$

   (b) Let $\mathbf{c}, \mathbf{x} \in \mathbb{R}^n, \mathbf{y} \in \mathbb{R}^m$. Suppose $x_i$ is the $i^{th}$ component of $\mathbf{x}$.

$$\begin{aligned} &\min_{\mathbf{x},\mathbf{y}} \; -\mathbf{c}^\top \mathbf{x} + \sum_{i=1}^{m} y_i \log y_i \\ &\text{s.t.}\, \mathbf{P}\mathbf{x} = \mathbf{y}, \\ &\qquad x_i \geq 0,\, i = 1, 2, \ldots, n, \\ &\qquad \sum_{i=1}^{n} x_i = 1, \end{aligned} \tag{5}$$

   where $\mathbf{P} \in \mathbb{R}^{m \times n}$ has nonnegative elements, and its columns add up to one.

   (c) (**Linear Programming**) Let $\mathbf{c}, \mathbf{x} \in \mathbb{R}^n, \mathbf{b} \in \mathbb{R}^m$.

$$\begin{aligned} &\min_{\mathbf{x},\mathbf{y}} \mathbf{c}^\top \mathbf{x} \\ &\text{s.t.}\, \mathbf{A}\mathbf{x} \preceq \mathbf{b}, \end{aligned} \tag{6}$$

   where $\preceq$ denotes componentwise inequality.

   Please find the dual problems of (4), (5), and (6) respectively.

**Solution:** ■

**Exercise 3: Discussions on Geometric Multiplier**

Consider the primal problem

$$\begin{aligned} \min_{\mathbf{x}} \ & f(\mathbf{x}) \\ \text{s.t.} \ & g_i(\mathbf{x}) \leq 0, \ i = 1, \cdots, m, \\ & h_i(\mathbf{x}) = 0, i = 1, \cdots, p, \\ & \mathbf{x} \in X. \end{aligned} \tag{7}$$

Are the following claims on the geometric multiplier for the primal problem correct? Justify the claims rigorously if they are correct. Otherwise please give a counterexample for each.

1. The geometric multiplier for the primal problem (7) always exists.
2. If the geometric multiplier exists, then it is unique.
3. Let $(\lambda^*, \mu^*)$ be a geometric multiplier. Then, the problem $\mathbf{argmin}_{\mathbf{x} \in X} L(\mathbf{x}, \lambda^*, \mu^*)$ always admits at least one solution, where $L(\mathbf{x}, \lambda, \mu)$ is the Lagrangian for (7).
4. If $(\lambda^*, \mu^*)$ is a geometric multiplier and the problem $\mathbf{argmin}_{\mathbf{x} \in X} L(\mathbf{x}, \lambda^*, \mu^*)$ admits a solution $\mathbf{x}^*$, then $\mathbf{x}^*$ is feasible.
5. Let $(\lambda^*, \mu^*)$ be a geometric multiplier. Then, $\mathbf{x}^*$ is a global minimum of the primal problem if and only if $\mathbf{x}^*$ is feasible and $\mathbf{x}^* \in \mathbf{argmin}_{\mathbf{x} \in X} \ L(\mathbf{x}, \lambda^*, \mu^*)$.

## Exercise 4: Strong Duality in Linear Programming

Consider the Linear Programming in (6). We denote the primal and dual optimal values by $f^*$ and $q^*$. Let $b_i$ be the $i^{th}$ component of $\mathbf{b}$ and $\mathbf{a}_i^\top$ be the $i^{th}$ row of $\mathbf{A}$.

1. Suppose $f^*$ is finite and $\mathbf{x}^*$ is an optimal solution. Let $I \subset \{1, 2, \ldots, m\}$ be the set of active constraints at $\mathbf{x}^*$:

$$\mathbf{a}_i^\top \mathbf{x}^* = b_i,\ i \in I,$$
$$\mathbf{a}_i^\top \mathbf{x}^* < b_i,\ i \notin I.$$

   Show that there exists a point $\mathbf{z} \in \mathbb{R}^m$ such that

$$z_i \geq 0,\ i \in I,$$
$$z_i = 0,\ i \notin I,$$
$$\sum_{i \in I} z_i \mathbf{a}_i + \mathbf{c} = 0,$$

   where $z_i$ is the $i^{th}$ component of $\mathbf{z}$. Further show that $\mathbf{z}$ is a dual optimal solution with objective value $\mathbf{c}^\top \mathbf{x}^*$.

2. Suppose $f^* = \infty$ and the dual feasible set is nonempty. Show that $p^* = \infty$.

**Solution:** ■

**Exercise 5: Duality Gap of the Knapsack Problem** 20pts

Given objects $i = 1, 2, \dots, n$ with positive weights $w_i$ and values $v_i$, we want to assemble a subset of the objects so that the sum of the weights of the subset does not exceed a given $A > 0$, and the sum of the values of the subset is maximized.

1. Show that this problem can be written as

$$\begin{aligned} \max_{x_1,\dots,x_n} \ & \sum_{i=1}^{n} v_i x_i \\ \text{s.t.} \ & \sum_{i=1}^{n} w_i x_i \le A, \\ & x_i \in \{0,1\},\ i = 1,2,\dots,n. \end{aligned} \tag{8}$$

2. Find the dual problem of (8) where a Lagrange multiplier is assigned to the constraint $\sum_{i=1}^{n} w_i x_i \le A$.

3. Suppose that the objects are sorted in the order $\frac{v_1}{w_1} \le \frac{v_2}{w_2} \le \cdots \le \frac{v_n}{w_n}$. Please solve the dual problem above. (Hint: the objective function $q(\lambda)$ is piecewise linear.)

4. Consider a slack version of the primal problem:

$$\begin{aligned} \max_{x_1,\dots,x_n} \ & \sum_{i=1}^{n} v_i x_i \\ \text{s.t.} \ & \sum_{i=1}^{n} w_i x_i \le A, \\ & x_i \in [0,1],\ i = 1,2,\dots,n. \end{aligned} \tag{9}$$

   Let $f_R^*$ be the optimal value of problem (9) and and $q_R^*$ be the optimal value of the dual problem, respectively. Show that

$$f_R^* = q_R^*.$$

5. Let $f^*$ be the optimal value of problem (8) and and $q^*$ be the optimal value of the dual problem, respectively. Show that

$$0 \le q^* - f^* \le \max_{i=1,\dots,n} v_i.$$

   (Hint: find the relationship between $(f^*, q^*)$ and $(f_R^*, q_R^*)$.)

6. Consider the problem where $A$ is multiplied by a positive integer $k$ and each object is replaced by $k$ replicas of itself, while the object weights and values stay the same. Let $f^*(k)$ and $q^*(k)$ be the corresponding optimal primal and dual values. Show that

$$\frac{q^*(k) - f^*(k)}{f^*(k)} \leq \frac{1}{k} \frac{\max\limits_{i=1,\ldots,n} v_i}{f^*}.$$

Thus, the relative value of the duality gap tends to 0 as $k \to \infty$.

**Solution:** ■

**Exercise 6: The Dual Problem of SVM**

Suppose that the training set is $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$, 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. The soft margin SVM takes the form of

$$\begin{aligned} \min_{\mathbf{w},b,\boldsymbol{\xi}} \ & \frac{1}{2}\|\mathbf{w}\|_2^2 + C\sum_{i=1}^n \xi_i, \\ \text{s.t. } & y_i(\langle \mathbf{w}, \mathbf{x}_i\rangle + b) \geq 1 - \xi_i, i = 1, \ldots, n, \\ & \xi_i \geq 0, \ i = 1, \ldots, n, \end{aligned} \tag{10}$$

The corresponding dual problem is

$$\begin{aligned} \min_{\alpha} \ & \frac{1}{2}\sum_{i=1}^n\sum_{j=1}^n \alpha_i\alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j\rangle - \sum_{i=1}^n \alpha_i \\ \text{s.t. } & \sum_{i=1}^n \alpha_i y_i = 0, \\ & \alpha_i \in [0, C], i = 1, \ldots, n. \end{aligned} \tag{11}$$

1. Show that the problems (10) and (11) always admit optimal solutions.
2. Let $(\mathbf{w}^*, b^*)$ be the solution to (10) and $\alpha^*$ be the corresponding solution to (11).
   (a) When does $\alpha_i^*$ equals to $C$, $i = 1, \ldots, n$? Please give an example and find the corresponding solutions.
   (b) When dose $\mathbf{w}^*$ equal to 0? Please give an example and find the corresponding solutions.

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

**Solution:** ■

**Exercise 7: An Example of the Soft Margin SVM**

Recall that the soft margin SVM takes the form of

$$\begin{aligned}\min_{\mathbf{w},b,\xi} \ & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i, \\ \text{s.t.}\ & y_i(\langle \mathbf{w}, \mathbf{x}_i\rangle + b) \geq 1 - \xi_i, i = 1,\ldots,n, \\ & \xi_i \geq 0,\ i = 1,\ldots,n,\end{aligned} \tag{12}$$

where $C > 0$.

1. The function of the slack variables used in the optimization problem for soft margin hyperplanes takes the form $\sum_{i=1}^{n}\xi_i$. Instead, we could use $\sum_{i=1}^{n}\xi_i^p$, with $p > 1$. The soft margin SVM becomes

$$\begin{aligned}\min_{\mathbf{w},b,\boldsymbol{\xi}} \ & \frac{1}{2}\|\mathbf{w}\|_2^2 + C\sum_{i=1}^{n}\xi_i^p, \\ \text{s.t.}\ & y_i(\langle \mathbf{w}, \mathbf{x}_i\rangle + b) \geq 1 - \xi_i, i = 1,\ldots,n, \\ & \xi_i \geq 0,\ i = 1,\ldots,n,\end{aligned} \tag{13}$$

   Please find the dual problem of (13) and the corresponding optimal conditions.

As shown in Figure 1, the training set is $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{11}$, where $\mathbf{x}_i \in \mathbb{R}^2$ and $y_i \in \{+1, -1\}$. Suppose that we use the soft margin SVM to classify the data points and get the optimal parameters $\mathbf{w}^*$, $b^*$, and $\boldsymbol{\xi}^*$ by solving the problem (13).

2. Please write down the equations of the separating hyperplane ($H_0$) and the marginal hyperplanes ($H_1$ and $H_2$) in terms of $\mathbf{w}^*$ and $b^*$.

3. Please find the support vectors and the non-support vectors.

4. Please find the values (or ranges) of the optimal slack variables $\xi_i^*$ for $i = 1, 2, \ldots, 11$. (*Hint: The possible answers are* $\xi_i^* = 0$, $0 < \xi_i^* < 1$, $\xi_i^* = 1$, *and* $\xi_i^* > 1$). How do the slack variables change when the parameter $C$ increases and decreases?

**Solution:** ■

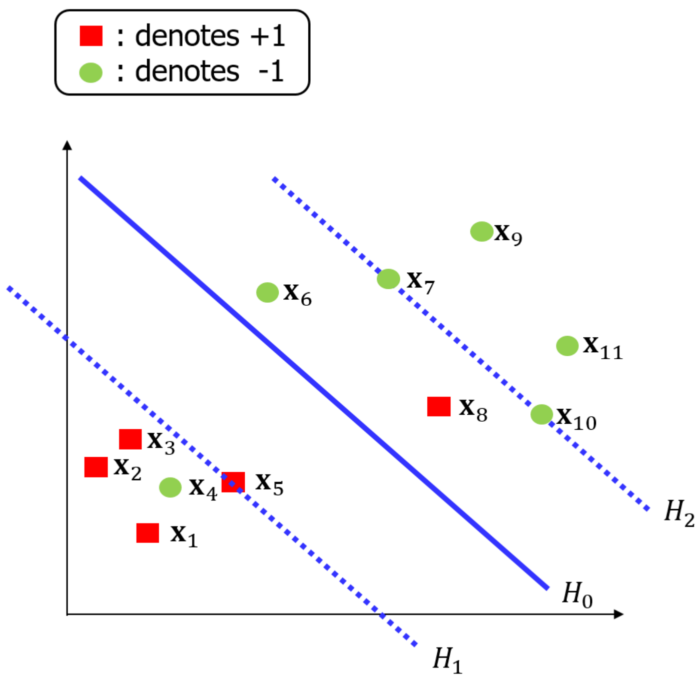


Figure 1: Classifying the data points using the soft margin SVM. $H_0$ is the separating hyperplane. $H_1$ and $H_2$ are the marginal hyperplanes.