Introduction to Machine Learning

Spring 2020

University of Science and Technology of China

Lecturer: Jie Wang — Homework 5

Posted: Apr. 18, 2020 — Due: May. 2, 2020

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. 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 explain it.

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

Solution: ■

Exercise 2: Visualization Lemma 10pts

Consider the primal problem as follows.

$$\begin{aligned} \min_{\mathbf{x}} \ & f(\mathbf{x}) \\ \text{s.t. } & \mathbf{g}(\mathbf{x}) \leq 0, \\ & \mathbf{h}(\mathbf{x}) = 0, \\ & \mathbf{x} \in X, \end{aligned} \tag{3}$$

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 (3) takes the form of

$$L(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{i=1}^{p} \mu_i h_i(\mathbf{x}). \tag{4}$$

Let

$$\mathbb{R}^{m+p+1} \supseteq S = \{(\mathbf{g}(\mathbf{x}), \mathbf{h}(\mathbf{x}), f(\mathbf{x})) : \mathbf{x} \in X\}. \tag{5}$$

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

Lemma 1. Visualization Lemma

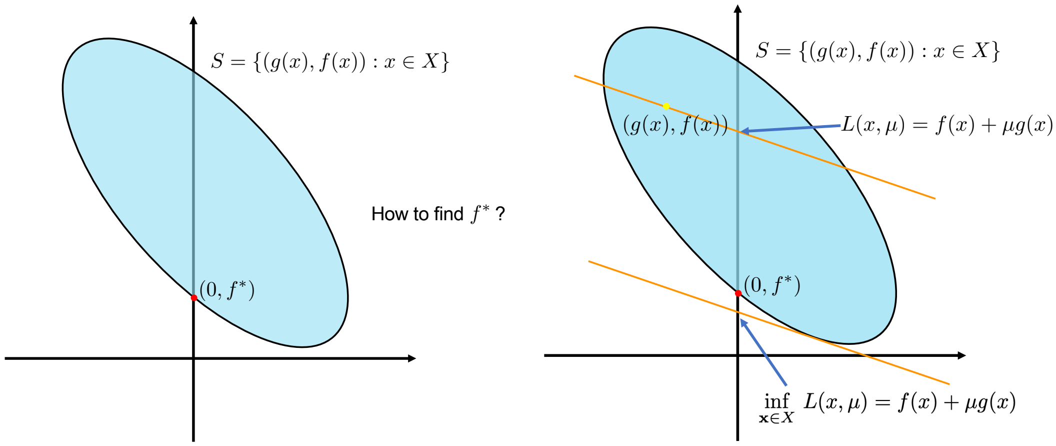
1. *The hyperplane with normal* $(\lambda, \mu, 1)$ *that passes through a vector* $(\mathbf{g}(\mathbf{x}), \mathbf{h}(\mathbf{x}), f(\mathbf{x}))$ *intercepts the vertical axis* $\{(\mathbf{0}, z) : z \in \mathbb{R}\}$ *at the level* $L(\mathbf{x}, \lambda, \mu)$.

2. *Among all hyperplanes with normal* $(\lambda, \mu, 1)$ *that contains in their positive halfspace the set* S *defined in (5), 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 interception of the vertical axis is* f^**, where*

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

Solution: ■



(a) (b)

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

Exercise 3: Geometric Multiplier 10pts

Let (λ^*, μ^*) be a geometric multiplier. Show that $\mathbf{x}^*$ is a global minimum of the primal problem (3) if and only if $\mathbf{x}^*$ is feasible and

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

Solution: ■

Exercise 4: Lagrange Dual Problem 10pts

Consider the primal problem (3) and the Lagrangian (4). 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 **dom** q is convex.
2. Show that $-q(\lambda, \mu)$ is a convex function.

Solution: ■

Exercise 5: Duality Gap 10pts

We denote the primal and dual optimal values by f^* and q^*. 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: ■

Exercise 6: Optimality Conditions 10pts

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

$$\mathbf{x}^* \in X,\, \mathbf{g}(\mathbf{x}^*) \leq 0, \mathbf{h}(\mathbf{x}^*) = 0, \qquad \text{(Primal Feasibility)}, \tag{6}$$

$$\lambda^* \geq 0, \qquad \text{(Dual Feasibility)}, \tag{7}$$

$$\mathbf{x}^* \in \underset{\mathbf{x}\in X}{\mathbf{argmin}}\, L(\mathbf{x}, \lambda^*, \mu^*), \qquad \text{(Lagrangian Optimality)}, \tag{8}$$

$$\lambda_i^* g_i(\mathbf{x}^*) = 0,\, i = 1, \ldots, m, \qquad \text{(Complementary Slackness)}. \tag{9}$$

Solution: ■

Exercise 7: Saddle Point Interpretation 10pts

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

$$L(\mathbf{x}^*, \lambda, \mu) \leq L(\mathbf{x}^*, \lambda^*, \mu^*) \leq L(\mathbf{x}, \lambda^*, \mu^*), \ \forall \, \mathbf{x} \in X, \ \lambda \geq 0. \tag{10}$$

Solution: ■

Exercise 8: Exercises of Dual Problems 40pts

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:

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

   and

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

2. Consider the problem

$$\begin{aligned} \min_{(x_1,x_2)} \ & 10x_1 + 3x_2 \\ \text{s.t.}\ & 5x_1 + x_2 \geq 4, \\ & x_1, x_2 = 0 \text{ or } 1. \end{aligned} \tag{11}$$

   (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 (11) 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. Consider the following problems.

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

$$\begin{aligned} &\min_{x,y} e^{-x} \\ &\text{s.t.}\, x^2/y \leq 1, \\ &\quad (x,y) \in \{(x,y) | y > 0\}. \end{aligned} \tag{12}$$

(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 $\mathbf{x}$.

$$\begin{aligned} &\min_{\mathbf{x},\mathbf{y}} -\mathbf{c}^\top \mathbf{x} + \sum_{i=1}^{m} y_i \log y_i \\ &\text{s.t.}\, \mathbf{P}\mathbf{x} = \mathbf{y}, \\ &\quad x_i \geq 0,\, i = 1, 2, \ldots, n, \\ &\quad \sum_{i=1}^{n} x_i = 1, \end{aligned} \tag{13}$$

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

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

$$\begin{aligned} &\min_{\mathbf{x},\mathbf{y}} \mathbf{c}^\top \mathbf{x} \\ &\text{s.t.}\, \mathbf{A}\mathbf{x} \preceq \mathbf{b}, \end{aligned} \tag{14}$$

where $\preceq$ denotes componentwise inequality.

Please find the dual problems of (12), (13), and (14) respectively.

Solution: ■

Exercise 9: Strong Duality in Linear Programming 10pts

Consider the Linear Programming in (14). 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 10: 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

$$\begin{aligned} \max_{x_1,\dots,x_n} & \sum_{i=1}^{n} v_i x_i \\ \text{s.t.} & \sum_{i=1}^{n} w_i x_i \le A, \\ & x_i \in \{0,1\},\ i = 1, 2, \dots, n. \end{aligned} \tag{15}$$

2. Use a graphical procedure to solve the dual problem where a Lagrange multiplier is assigned to the constraint $\sum_{i=1}^{n} w_i x_i \le A$.

3. Let f^* and q^* be the optimal values of the primal and dual problems, respectively. Show that

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

4. 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

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

so that the relative value of the duality gap tends to 0 as $k \to \infty$.

Solution: ■

Exercise 11: The Dual Problem of SVM 20pts

Recall that the soft margin SVM takes the form of

$$\begin{aligned} \min_{\mathbf{w},b,\xi} \ & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i, \\ \text{s.t. } & y_i(\langle \mathbf{w}, \mathbf{x}_i\rangle + b) \geq 1 - \xi_i, i = 1,\ldots,n, \\ & \xi_i \geq 0,\ i = 1,\ldots,n, \end{aligned} \tag{16}$$

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

$$\begin{aligned} \min_{\alpha} \ & \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j\rangle - \sum_{i=1}^{n}\alpha_i \\ \text{s.t. } & \sum_{i=1}^{n}\alpha_i y_i = 0, \\ & \alpha_i \in [0, C], i = 1,\ldots,n. \end{aligned} \tag{17}$$

1. Show that the problems (16) and (17) always admit optimal solutions.
2. Let $(\mathbf{w}^*, b^*)$ be the solution to (16) and α^* be the corresponding solution to (17).
   (a) When does α_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: ∎